

PDC Center for High Performance Computing

# PDC Newsletter

**Great Success for KTH & PDC in Innovations Towards Exascale**

- Rossen Apostolov, PDC, page 4

**CEEC: Centre of Excellence for Exascale CFD** - Niclas Jansson, PDC, page 5

**Plasma-PEPSC CoE** - Stefano Markidis, PDC, page 6

**BioExcel CoE Awarded Third Round of Funding for Four More Years**

- Rossen Apostolov, PDC, page 7

**PerMedCoE@PDC** - Alessandra Villa, PDC, page 8

**PDC Systems Course** - Xin Li, PDC, page 9

**AMD GPU Hackathon for Research Software Developers Hosted by**

**ENCCS in Stockholm** - Thor Wikfeldt, ENCCS, page 10

**GROMACS 2023** - Szilárd Páll, PDC, & Andrey Alekseenko, SciLifeLab/KTH, page 12

**Dardel Updates** - Gert Svensson, PDC, page 14

**EuroHPC Summit 2023** - Dirk Pleiter, KTH, page 15

**Co-design Efforts for Major Performance Improvements in GROMACS**

**GPU-Accelerated Parallelisation** - Szilárd Páll, PDC, & Alan Gray, NVIDIA, page 16

**Second Phase of ENCCS** - Thor Wikfeldt, ENCCS, page 17

**AI Frameworks on the New Dardel GPU Partition** - Xavier Aguilar, PDC, page 18

**Nascent NAISS** - Björn Alling, NAISS, LiU page 20





**Patrick Norman**  
Director, PDC

### 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 is hosted by EECS, KTH, and is a member of the National Academic Infrastructure for Supercomputing in Sweden (NAISS).

**Editor:** Patrick Norman

**Editorial staff:** Genet Edmondson

**Graphic Design:** Maria Malmqvist

**E-mail:** [support@pdc.kth.se](mailto:support@pdc.kth.se)

ISSN 1401-9671

### Cover

The cover image shows Niclas Jansson (PDC), Stefano Markidis (PDC) & Rossen Apostolov (PDC), who are respectively the Directors of the BioExcel, CEEC & Plasma-PEPSC centres of excellence, at the EuroHPC Summit that was held in Gothenburg in March 2023.

### 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](mailto:support@pdc.kth.se)

**www:** <http://www.pdc.kth.se>

**Phone:** +46 8 790 7800

## Editorial

Swedish high-performance computing (HPC) is in a transition phase at the moment, and centre stage is taken by the National Academic Infrastructure for Supercomputing in Sweden (NAISS). At the start of this year, NAISS took over responsibility for academic HPC, storage, and data services when the Swedish National Infrastructure for Computing came to a close. This change involves much more than a mere rebranding of the organisation and a relocation of the head office (see *Nascent NAISS*). A well-founded sense of optimism has spread throughout Swedish academic institutions and industry alike, and this national optimism was tangible at the EuroHPC Summit 2023 conference, which was organised by the European High Performance Computing Joint Undertaking (EuroHPC JU) and held in Gothenburg in March (see *EuroHPC Summit 2023*).

The Swedish membership of the EuroHPC JU programme is bringing exascale computing to the doorsteps of Swedish researchers, and there is no alternative but for all of us to embrace this paradigm change and leverage it to the best of our abilities. This is where the concept of the HPC ecosystem for research becomes significant: knowledge, competencies, and applications are coming to the fore as being equally as important as hardware for utilising the full potential of the coming exascale HPC research infrastructure as we strive to advance sciences and engineering.

On the European level, a call was launched last year for centres of excellence (CoEs) to be formed and address the software side of the exascale ecosystem. Out of the nine CoEs that were awarded funding, no less than three are led by PDC (see *Great Success for KTH & PDC in Innovations Towards Exascale*). This rather overwhelming success is quite obviously rooted in strong scientific research environments, and here a key role is being played by the Swedish e-Science Research Centre (SeRC, <https://e-science.se>), which is led by the KTH Royal Institute of Technology. (The Karolinska Institute, Stockholm University and Linköping University are additional members.) But somewhat less obviously, perhaps, it is important to also acknowledge the long-term and continuous support that PDC has received directly from KTH to build up the competencies and knowledge at PDC that is needed now more than ever before.

In the new NAISS era of Swedish HPC, this strong commitment from KTH will continue, and it will arguably be no less than a decisive factor in realising the necessary component of applications within the Swedish national ecosystem that enables forefront research on our systems. There is now no longer a question of whether accelerators

will reach into and affect work in your favourite e-science domain. However, that is not the same as saying we must all learn, or even care, about the often daunting technical details involved with modern parallel programming, some of which can be read about in this newsletter. PDC will take a leading national (and in some fields also international) role in the development of HPC software applications in sciences and engineering. We will do so with the ambition of providing user application interfaces that are less complex and that make it easier for less experienced users to access our increasingly complex hardware resources.

On the topic of NAISS hardware resources, PDC can proudly announce that the Dardel system has now reached its full intended capacity with the GPU and CPU partitions at positions 77 and 153, respectively, on the most recent TOP500 list (see [Dardel Updates](#)). Bringing this heterogeneous system into use has involved – and continues to involve – hard work for the systems and application experts at PDC, but it represents a step that must be taken in order for our community of researchers to remain competitive. In addition, the experiences gained in this process will be invaluable for whatever the future has in store for Swedish HPC.

*Patrick Norman, Director PDC*

## Inauguration of Dardel

23 August 2023, 13.00-16:45  
KTH main campus, Stockholm



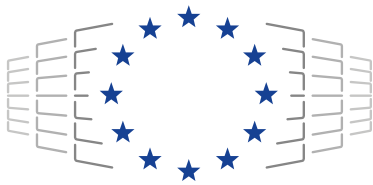
PDC and NAISS would like to invite everyone who is interested in using high-performance computing to expedite academic and business/ industrial research to the inauguration of the complete Dardel system, which is hosted at PDC.

Now that both phases of Dardel have been installed, the inauguration of the whole system is planned for the 23<sup>rd</sup> of August 2023 starting at 13.00 in lecture room F2, Lindstedtsvägen 22, at the KTH Royal Institute of Technology main campus in Stockholm. More information will be sent out soon, but you can sign up immediately at <https://www.kth.se/form/dardel-inauguration-2023> to reserve a place for attending in person.

## In This Issue

### Editorial

<i>Patrick Norman</i> .....	2
<b>Great Success for KTH &amp; PDC in Innovation Towards Exascale</b>	
<i>Rossen Apostolov</i> .....	4
<b>CEEC: CoE for Exascale CFD</b>	
<i>Niclas Jansson</i> .....	5
<b>Plasma-PEPSC CoE</b>	
<i>Stefano Markidis</i> .....	6
<b>BioExcel Awarded 3rd Round of Funding for 4 More Years</b>	
<i>Rossen Apostolov</i> .....	7
<b>PerMedCoE@PDC</b>	
<i>Alessandra Villa</i> .....	8
<b>PDC Systems Course</b>	
<i>Xin Li</i> .....	9
<b>AMD GPU Hackathon for RSDs Hosted by ENCCS</b>	
<i>Thor Wikfeldt</i> .....	10
<b>GROMACS 2023</b>	
<i>Szilárd Páll &amp; Andrey Alekseenko</i> ..	12
<b>Dardel Updates</b>	
<i>Gert Svensson</i> .....	14
<b>EuroHPC Summit 2023</b>	
<i>Dirk Pleiter</i> .....	15
<b>Co-design Efforts for Major Performance Improvements in GROMACS GPU-Accelerated Parallelisation</b>	
<i>Szilárd Páll &amp; Alan Gray</i> .....	16
<b>Second Phase of ENCCS</b>	
<i>Thor Wikfeldt</i> .....	17
<b>AI Frameworks on the New Dardel GPU Partition</b>	
<i>Xavier Aguilar</i> .....	18
<b>Nascent NAISS</b>	
<i>Björn Alling</i> .....	20
<b>PDC-Related Events</b> .....	20
<b>HPC Sources</b> .....	20



**EuroHPC**  
Joint Undertaking

## Great Success for KTH & PDC in Innovations Towards Exascale:

**3 of 9 EuroHPC JU CoEs Lead by PDC at KTH**  
**Rossen Apostolov, PDC**

As supercomputer technology progresses rapidly towards achieving performance in the exascale range, software applications need to be modified or, in some cases, completely redesigned and rewritten to make efficient use of exascale capabilities. That needs to be done ahead of, or at least in parallel with, the technological developments so that researchers will be able to start utilising exascale systems as soon as they are available. A lot of preliminary work is needed to prepare for the transition to exascale computing.

Europe has been preparing for the transition by establishing Centres of Excellence (CoEs) for High-Performance Computing (HPC). These began in 2015 under the auspices of the EC Horizon 2020 programme. In 2018, the European High Performance Computing Joint Undertaking (EuroHPC JU) was set up, and in 2022, it launched calls for CoEs to undertake research and innovation actions that would develop and adapt HPC applications for the exascale and post-exascale eras.

Nine European CoEs were awarded EuroHPC JU funding which began at the start of 2023; some are completely new centres, while others are continuations of already existing CoEs. Of those nine newly funded CoEs across Europe, KTH is coordinating three.

BioExcel, the Centre of Excellence for Computational Biomolecular Research, is directed by Rossen Apostolov at PDC. It was established in 2015 to provide life sciences researchers with high-quality, user-friendly software and to increase their HPC expertise and skills. This extension of the project is known as BioExcel-3.

CEEC, the Centre of Excellence for Exascale CFD, is directed by Niclas Jansson from PDC. CEEC focuses on adapting and developing computational fluid dynamics (CFD) applications for exascale.

Plasma-PEPSC, the Plasma Exascale-Performance Simulations Centre of Excellence, is directed by Stefano Markidis at PDC and will support exascale application development for plasma science in a range of research areas.

In addition, KTH and PDC are partners in two more CoEs: EXCELLERAT, the European Centre of Excellence for Engineering Applications, and PerMedCoE, the Centre of Excellence for Personalised Medicine. PDC has been heavily involved with EXCELLERAT since it started in 2018 and will continue to assist its mission of providing expertise (particularly to the aeronautics, automotive, energy and manufacturing sectors) on utilising HPC to benefit engineering research and development as we move towards exascale.

It is a really great success for KTH and PDC that the EuroHPC JU has awarded funding to continue BioExcel and EXCELLERAT and to start up CEEC and PLASMA-PEPSC. These CoEs are all high-profile projects for HPC research and innovation with a major impact on science and technology.

Recently, the directors of the three CoEs that are being coordinated by KTH all attended the EuroHPC Summit in Gothenburg, which ran from 20-23 March 2023. This year's summit focused on "European Supercomputing Excellence in the Exascale Era". Stefano Markidis gave a presentation on the "Scientific Expectations in the Exascale Era" for researchers from the astrophysics and space physics communities, and Rossen Apostolov was one of the organisers of the open workshop session about Centres of Excellence and National Competence Centres.







## **CEEC: Centre of Excellence for Exascale CFD**

**Niclas Jansson, PDC**

For many centuries, scientific discovery relied on performing experiments and the subsequent deduction of new theoretical models. The advent of powerful computers, coupled with new and ever more efficient numerical algorithms, makes it possible to simulate complex systems with increasing realism and to automatise even model discovery using AI technologies. Computational Fluid Dynamics (CFD) is one of the most prominent areas that clearly requires and even motivates exascale computing to be part of the engineering and academic workflows. Given the physical scaling and the availability of highly efficient simulation codes, CFD is one of the few application areas that has the potential of reaching exascale performance in the near future.

CEEC is a recently started centre of excellence (CoE) for high-performance computing (HPC) applications, co-funded by the European High Performance Computing Joint Undertaking (EuroHPC JU). The centre is currently funded from the start of 2023 to the end of 2026. One of CEEC's main purposes is to implement exascale-ready workflows for addressing relevant challenges for future exascale systems, including those procured by EuroHPC. This includes facilitating significant improvements in energy efficiency through the efficient exploitation of accelerated hardware architectures based on graphics processing units (GPUs) and utilising novel adaptive mixed-precision calculations. Emphasis will furthermore be given to new or improved algorithms, which will be necessary for exploiting upcoming exascale architectures.

The work of CEEC is driven by a collection of six different lighthouse cases of physical and engineering interest, ranging from aeronautical

to atmospheric flows, with the goal of reaching a technology readiness level (TRL) of four and even five for selected cases. These are the lighthouse cases chosen to illuminate future developments.

- Two aeronautic cases look at questions like why we have in-flight turbulence and how to make plane wings more efficient without risking structural safety.
- Topology optimisation of static mixers (which use the internal shape of a pipe to mix fluids without any moving parts) considers issues like what is the best pipe shape for mixing fluids.
- Localised erosion of an offshore wind turbine foundation investigates ways to minimise negative environmental impacts from the installation of wind turbines.
- Simulation of Atmospheric Boundary Layer (ABL) flows works on understanding and predicting ABL flows to optimise, for example, solar energy production and weather modelling.
- Merchant ship hull will improve simulations of water and air flow around ships.

All the CEEC software development for these cases is being done in five European HPC codes which span the entire spectrum of CFD applications, including compressible, incompressible and multiphase flows.

PDC is coordinating CEEC and leading the management work package, while the technical work is being performed in collaboration with several research groups and centres. The KTH Royal Institute of Technology brings in the merchant ship hull lighthouse case, as well as the spectral element code Neko. With expertise in application development and optimisation, KTH contributes to the software and performance engineering work package and leads the software engineering and software deployment tasks. Furthermore, KTH will work on exascale algorithms and is leading the efforts on fault-resilient algorithms, as well as adaptivity and error control. Finally, KTH leads the efforts to explore quantum computing as a forward-looking technology.

For more information about CEEC and the lighthouse cases, see <https://ceec-coe.eu>.



## Plasma-PEPSC CoE: Advancing Plasma Science with Exascale Computing

Stefano Markidis, PDC

Led by the KTH Royal Institute of Technology, with Dr. Markidis as the coordinator, the EuroHPC Plasma Exascale-Performance Simulation Centre of Excellence (Plasma-PEPSC CoE) unites ten European partners in a collaborative effort to lead plasma science into the era of exascale computing. With a budget of 7.9 million euros, Plasma-PEPSC aims to drive scientific breakthroughs in plasma science's most significant challenges (fusion energy, accelerator devices and space physics) through cutting-edge hardware and software advancements. Plasma-PEPSC started on January 2023 and will last for four years.

Plasma science has long been at the forefront of high-performance computing (HPC), driving innovation in both hardware and software domains. The overarching goal of Plasma-PEPSC is to take this development to the next level, enabling unprecedented simulations on current pre-exascale and future exascale platforms across Europe. Four flagship plasma codes with a large

user base – BIT, GENE, PIconGPU, and Vlasiator – serve as the focal points of the centre of excellence. By maximising their parallel performance and efficiency, we aim to achieve breakthroughs in controlling plasma-material interfaces, optimising magnetically confined fusion plasmas, designing next-generation plasma accelerators and predicting space plasma dynamics within the Earth's magnetosphere.

Plasma-PEPSC is built upon a foundation of algorithmic advances, encompassing load balancing, resilience, and data compression techniques. Additionally, we employ state-of-the-art programming models and libraries, such as MPI, accelerator and data movement APIs and runtimes, and in-situ data analysis tools. Through an integrated HPC software engineering approach, we deploy, verify, and validate extreme-scale kinetic plasma simulations that will set a community standard. To ensure seamless integration and the pursuit of excellence, Plasma-PEPSC adopts a continuous and integrated co-design methodology. We are actively engaged in providing and receiving direct input relating to the design and development of the European Processor Initiative (EPI) processor and the RISC-V accelerator and investigating quantum computing approaches for plasma simulations.

As we embark on this journey, we invite you to stay connected with the Plasma-PEPSC project at <https://plasma-pepsc.eu> as we will share our progress with the HPC community through updates and collaborations.



Above: PLASMA-PEPSC kick-off meeting at KTH and online, 24-25 January 2023

## BioExcel CoE Awarded Third Round of Funding for Four More Years

Rossen Apostolov, PDC

BioExcel ([www.bioexcel.eu](http://www.bioexcel.eu)), the leading European Centre of Excellence for Computational Biomolecular Research, has secured new funding from the Horizon Europe research funding programme to operate until the end of 2026. This iteration of the centre is known as BioExcel-3 and includes eight partner institutions which are coordinated by PDC.

BioExcel was established in 2015 and quickly positioned itself as a leading research and innovation hub for the biomolecular modelling and simulation communities. The centre's successes led to a second round of funding for BioExcel-2, which ran from 2019 to 2022. In the guise of BioExcel-3, the centre continues to focus on accelerating the development and usability of some of the most widely used European open-source software for computational biomolecular research: GROMACS for molecular dynamics simulations, HADDOCK for integrative docking, PMX for free energy calculation preparation, and BioBB (a workflow building blocks platform).

In contrast to other projects, BioExcel is strongly user-driven. Feedback by thousands of users of the core applications (for example, from user surveys) is analysed and incorporated into the software development maps. Moreover, the innovative training programme offered by BioExcel is constructed based on an analysis of gaps between user needs and existing training material. Tailored training activities are a fundamental aspect of the programme. The centre undertakes numerous support activities, such as offering public support forums (see <https://ask.bioexcel.eu> and [\[gromacs.bioexcel.eu\]\(http://gromacs.bioexcel.eu\)\), online documentation, webinars \(About 70 webinars are available at <https://bioexcel.eu/category/webinar>\), tutorials and in-depth support from experts in the field.](https://</a></p></div><div data-bbox=)

In light of the upcoming exascale computing era, BioExcel is working closely with hardware vendors to ensure optimal execution of the core biomolecular software on the latest supercomputing resources. In this third phase of the project, BioExcel – in close collaboration with the European High Performance Computing Joint Undertaking (EuroHPC JU), which is the main co-funding agency, and participating member states – is continuing its mission to advance science and technology in the life sciences, particularly by focusing on the following activities.

- Improving the usability of the core software and the reproducibility of results by increasing the adoption of standards and best practices, as well as developing biomolecular workflows for diverse computing environments
- Strengthening BioExcel's engagement with academic and industrial biomolecular research communities to guide software development and ensure it has a high scientific impact
- Providing state-of-the-art training and support to reduce the skills gaps for researchers and students and assist in the development of the next generation of biomolecular computational scientists
- Conducting wide and effective dissemination of results from our activities, promoting further use of the core software and encouraging further uptake of BioExcel's support services by the life sciences research community

BioExcel encourages active collaborations between academic and industrial researchers and will promote the exploitation of HPC resources for biomolecular research across EU member and associated states through a newly launched Ambassador Programme. (Contact Rossen Apostolov, [rossen@kth.se](mailto:rossen@kth.se), for further information.)

To learn more about the various resources and activities offered by BioExcel or to sign up to our newsletter, visit <https://bioexcel.eu>.





HPC/Exascale  
Centre of  
Excellence in  
Personalised  
Medicine

## PerMedCoE@PDC

Alessandra Villa, PDC

PerMedCoE (<https://permedcoe.eu>) is the high-performance computing (HPC)/exascale centre of excellence (CoE) for personalised medicine in Europe. This CoE is working to provide an entry point to HPC/exascale methodology to translate omics analysis into actionable models of cellular functions. (“Omics” is an informal term covering the biological sciences that end in “omics”, such as genomics and proteomics.) The core software packages for these efforts are COBREXA.jl, CellNOpt, MaBoSS and PhysiCell.

The KTH Royal Institute of Technology is one of the partners in the CoE, and PDC has been contributing to building PerMedCoE and the associated HPC research community, as well as mapping competencies needed by professionals in the personalised medicine field, designing target training materials, running use cases and optimising the performance and scalability of PerMedCoE workflows.

In particular, over the last few months, PDC has worked on the development of a competency hub in the field of computational personalised

medicine (which is hosted by EBI-EMBL) and also on developing training materials and occasionally on filling in the gaps between the fields of personalised medicine and HPC, and finally on evaluating the scalability and performance of PerMedCoE workflows.

The competency hub (<https://competency.ebi.ac.uk>) is a hub where researchers can browse competencies, career profiles and training resources to advance their careers in the life sciences. PDC has identified and defined a series of competencies that professionals in the field of computational personalised medicine require for different career profiles and collected that information in the hub. Some examples are a postdoctoral fellowship in molecular dynamics simulations and a position as a senior researcher and software architect.

Learning pathways are developed within the framework of the hub. The pathways are collections of training courses and materials to help researchers learn about specific topics. For example, PDC has just released a course on “Version control with Git” (see <https://cms.competency.ebi.ac.uk/learning-pathway/version-control-git>). Through this pathway, researchers can learn about version control, which is the practice of tracking and managing changes to software code. The aim of the course is to give an indication of the level of complexity of the computational techniques and knowledge required to keep track of what a developer did and when, and also to demonstrate how to work collaboratively on a document or contribute actively to an open-source project.

The image shows two screenshots of the PerMedCoE Competency Hub website. The left screenshot is the main page, featuring a header with the title 'Competency Hub' and a sub-header 'Browse competencies, career profiles and training resources to advance your career in the life sciences'. Below this is a section titled 'Who is the Competency Hub for?' with three columns: 'Students and professionals in computational biology', 'Professionals in computational biomolecular research', and 'Professionals in computational personalised medicine'. The right screenshot shows the 'Career profiles' section, which includes an 'Add your profile' button and a grid of profile cards for various roles: Junior Research Software Engineer, Senior Research Software Engineer, Junior Computational Chemist, Senior Computational Chemist, Research Associate in Biomolecular Modelling, PhD Student in Biomolecular Simulations, and Principal Investigator.

Above: PerMedCoE Competency Hub main page (left) and the page with examples of career profiles (overlaid on right)



Below: "Version control with Git" learning pathway

LEARNING PATHWAY

## Version control with Git

Overview

About this pathway This is an introduction to the concept of version control and to the terminology behind this.

**Introduction to version control**

Setting up Git and GitHub After this module you should be able to:

Version control using Git

- Navigate the concept behind version control
- Understand the terminology used version control
- Understand output messages of version control tools

Collaborating on a

Together with BioExcel, PDC co-organised the online “Introduction to HPC for Life Scientists” course hosted by Barcelona Supercomputing Center (BSC). There Szilárd Páll (PDC) and Berk Hess (KTH) gave a collection of inspiring lectures and exercises on “Mapping computation to HPC hardware and GPU accelerators and heterogeneous architectures” (see [https://permedcoe.eu/training\\_type/material](https://permedcoe.eu/training_type/material)). The workshop was a significant success, with remarks from the participant such as “The course has opened my eyes to the potential of HPC in my research field” and “Gaining hands-on experience with GROMACS and PhysiCell has given me the confidence to optimise my research projects for HPC environments”.

Finally, PDC is undertaking a performance and scalability analysis of different workflows for Personalised Medicine on the Dardel supercomputer system at PDC. These workflows are designed to solve different PerMed use cases (<https://permedcoe.eu/use-cases>) which tackle real-life medical problems, such as multiscale modelling of the COVID-19 virus and drug synergies for cancer treatment. In particular, Xavier Aguilar (PDC) has been looking at the types of problems that workloads face when they run on a high number of cores, for example, communication problems across nodes, work imbalances, or memory contention issues within a node. In addition, Xavier is exploring the challenges of deploying multiscale workloads in an optimal manner on an HPC infrastructure such as Dardel, since these are not the common monolithic parallel HPC codes.

## PDC Systems Course

Xin Li, PDC

With the biannual course “Introduction to PDC Systems”, we introduce PDC’s high-performance computing (HPC) infrastructure to new users. In this course, we start with an overview of the HPC infrastructure at PDC and then cover basic topics such as getting an account, logging in, running jobs, storing data, and compiling code. In addition, we also cover practical topics, including Bash Shell, SLURM job script, ThinLinc remote desktop, Singularity, Matlab and the Python virtual environment. This year the course also includes a tutorial on using the new graphics processing unit (GPU) partition of the Dardel system.

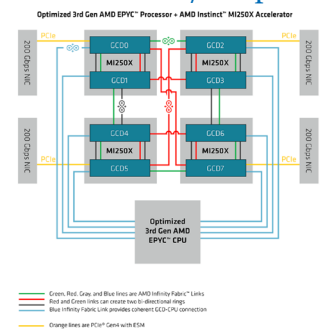
The first introductory course for 2023 was held online on the 13<sup>th</sup> and 14<sup>th</sup> of April. Thirty participants registered for the course. During the course, we walked through the slides (available at <https://pdc-support.github.io/pdc-intro>) and guided the participants with hands-on exercises for common procedures and useful commands on Dardel. Participants also practised compiling code and running jobs to deepen their understanding of how to use the Dardel system efficiently.

We aim to continue the introductory course in autumn 2023 and will do our best to help new users become comfortable and productive with working in the HPC environment at PDC. For information about the next course and other training events run by (or in collaboration with) PDC, you can sign up for the PDC general announcements mailing list <https://www.pdc.kth.se/contact/joining-pdc-mailing-lists-1.736925>, check on the PDC Events web page <https://www.pdc.kth.se/about/events> or follow PDC on Facebook at <https://www.facebook.com/kth.pdc>.

### Dardel GPU nodes

Dardel has 56 GPU nodes, each of which is equipped with

- One AMD EPYC™ processor with 64 cores
- 512 GB of shared fast HBM2E memory
- Four AMD Instinct™ MI250X GPUs (with an impressive performance of up to 95.7 TFLOPS in double precision when using special matrix operations)



Above: One of the course slides on using Dardel GPU nodes

# AMD GPU Hackathon for Research Software Developers Hosted by ENCCS in Stockholm

Thor Wikfledt, ENCCS

From 14-16 March 2023, the EuroCC National Competence Centre Sweden (ENCCS) hosted a hackathon event at the RISE Research Institutes of Sweden offices on the main KTH Royal Institute of Technology campus in Stockholm. The hackathon was aimed at research software developers interested in porting or optimising their code to run on AMD Instinct™ graphics processing units (GPUs) using GPU programming frameworks such as HIP, OpenMP, OpenACC, or SYCL.

The event utilised the Dardel supercomputer, Sweden's flagship supercomputer operated by PDC at KTH in Stockholm, which is accessible for academic research through NAISS. Participating teams were granted access to the Dardel system ahead of the hackathon.

The event had an impressive turnout, and eight teams were ultimately invited to join. The level of interest was indeed higher than what could be accommodated based on the number of expert mentors, but we hope to continue running GPU programming hackathons in coming years! To apply for the hackathon, teams submitted a detailed description of their project and the potential impact of their code on a specific organisation or the wider community. Prior to the hackathon, the teams that were accepted were expected to engage with recommended learning resources, profile their code, and meet virtually (online) with their assigned mentors.

The hackathon kicked off with an online day on the 7<sup>th</sup> of March, where teams were matched with mentors and treated to introductory seminars on GPU programming by seasoned experts from AMD and HPE. George Markomanolis, a principal member of technical staff at AMD, provided an in-depth walkthrough of compilation aspects and profiling tools like Omniperf, while John Levesque, a senior distinguished technologist at HPE,

*Below: Hackathon team working virtually with a mentor*



offered a comprehensive introduction to the HPE programming environment and profiling tools. Johan Hellsvik, an application expert at PDC, also gave an introduction to the PDC environment.

During the primary in-person segment of the hackathon from 14-16 March, participating teams received invaluable guidance from expert mentors representing HPE, AMD, PDC, and ENCCS. Each day commenced with a stand-up session, where teams shared their progress, challenges, and daily goals. Based on these discussions, mentors were assigned to provide appropriate support.

Although each of the teams and their mentors spent a significant portion of the day in separate rooms, concentrating on their projects, a vibrant sense of community was palpable during coffee breaks, on the hackathon chat channel, and especially during social events. In fact, teams and mentors gathered informally at a renowned Stockholm pub on the eve of the hackathon on the 13<sup>th</sup> of March, and an official dinner took place on the evening of the 14<sup>th</sup>. Additionally, participants enjoyed an exceptional guided tour of the PDC machine room, led by Luca Manzari and Gert Svensson, the system manager and deputy



*Above: Hackathon dinner on 14<sup>th</sup> of March*

*Below: Some hackathon participants with the Dardel system at PDC during the tour of the PDC computer hall*



director of PDC, respectively. The tour offered an in-depth look at the complexities of liquid cooling, backup power, and fire safety systems, along with intriguing stories and historical tidbits about PDC.

Overall, the hackathon proved to be a valuable opportunity for research software developers to enhance their GPU programming skills and collaborate with expert mentors in achieving their goals. The event was a resounding success, and we eagerly anticipate hosting similar events in the future! For a more detailed account of the hackathon experience from the point of view of the project teams, please read on for stories from the UppASD and IFS teams!

## **UppASD**

UppASD is a program for simulation of atomistic spin dynamics and spin-lattice dynamics. The program is written in Fortran 2003 with shared memory parallelisation over CPU cores by means of OpenMP. Team UppASD entered the hackathon with the expectation of getting expert advice on which programming model to work with for GPU offloading of compute intensive kernels, and getting started with the implementation. Being agnostic at first on the choice of HIP, OpenSYCL, and OpenMP, the team settled for working with OpenMP. The first step was to obtain a better understanding of the performance of the current OpenMP parallelised CPU code. Then work ensued to implement offloading directives for one of the dominant terms of the Hamiltonian, namely bilinear spin-exchange, as well as for the semi-implicit midpoint integrator. The good progress made during the hackathon forms a platform for porting all the main parts of UppASD to GPU code.

## **Integrated Forecasting System (IFS)**

Two teams from the European Centre for Medium-Range Weather Forecasts (ECMWF) offices in Bonn and Reading travelled to Stockholm. In their metaphorical suitcases were two components of the weather forecasting model IFS, which is currently being deployed and optimised on the LUMI system for the first Digital Twins of the European Commission's Destination Earth initiative.

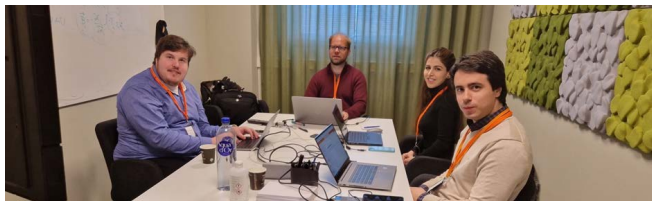
The objective of the first team was to make the existing initial offload implementation of the spectral transformation library, a key component and one of the most computationally expensive parts of the IFS, run faster on AMD MI250X GPUs. The second team worked on a proxy application for the physical parameterisations, with the goal of developing an optimal GPU adaptation recipe. For that, they could draw on a wide range of already available programming model implementations of the same algorithm, trying to make as many as possible work as fast as the reference results on NVIDIA A100 GPUs.

IFS is written mostly in Fortran, and early on, teething problems of the relatively juvenile software stack for AMD GPUs were encountered. But the Cray compiler could ultimately be convinced to generate working offload binaries for both applications. With the help of the fabulous mentors, profiling efforts were soon successful, and first optimisation approaches were identified. After three days, the spectral transforms team could enter a fourfold speed-up via a mixture of targeted optimisations. The physical parameterisations application managed to achieve on-par performance with NVIDIA A100 GPUs for a HIP implementation of the algorithm but failed to achieve improvements on other programming model implementations, in particular pragma-based approaches.

Nevertheless, it was a positive outcome for all, returning home with suitcases filled with faster code, a lot of newly acquired knowledge, memories of positive interactions with mentors and other participants, and an even longer list of ideas and tasks they want to tackle next.



Below: A team working with their mentor (top) and Hackathon participants and mentors at KTH



## GROMACS 2023: Readiness on the AMD GPU Heterogeneous Platform

Szilárd Páll, PDC, and Andrey Alekseenko, Scilifelab & Department of Applied Physics, KTH

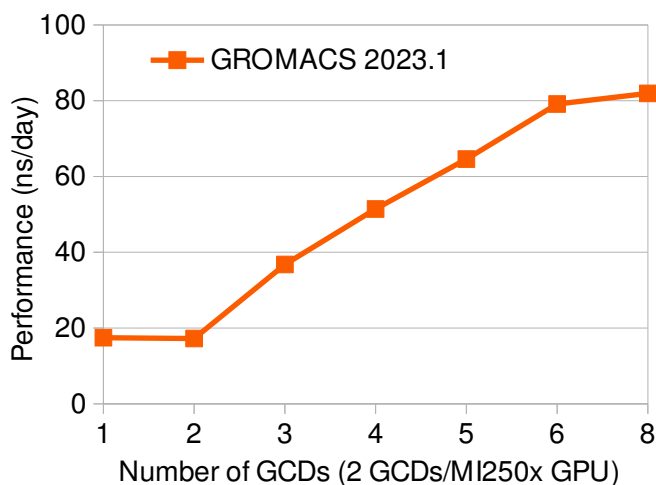
GROMACS, the widely used open-source molecular dynamics package, is one of the major high-performance computing (HPC) research codes which has long supported heterogeneous graphics processing unit (GPU) acceleration. GROMACS users have benefitted from utilising GPUs for nearly a decade.

Molecular dynamics research using GROMACS is one of the significant workloads on the Dardel system at PDC, as well as on the LUMI system in Finland. CPU performance – both absolute and strong scaling on CPU-only AMD platforms (including Dardel’s CPU partition and LUMI-C) – has been excellent. At the same time, the GROMACS codebase has been being prepared for the increasingly heterogeneous and diverse HPC platforms through algorithmic, parallelisation, and portability efforts, which were discussed in the 2022 no.1 edition of this newsletter (see <https://www.pdc.kth.se/publications/pdc-newsletter-articles/2022-no-1/preparing-gromacs-for-heterogeneous-exascale-hpc-systems-1.1174906>). Last year, the heterogeneous parallelisation and SYCL portability layer targeting AMD GPUs were enhanced in a number of key areas in preparation for providing broad support for these new

accelerator-based HPC resources in the official 2023 GROMACS release [1,2], which was released in February 2023.

The GROMACS team has implemented performance optimisations targeting the AMD CDNA2 architecture at the heart of the MI250X accelerators (which are used, for example, in the GPU partition of Dardel) to improve the performance of key compute kernels. This made significant performance gains possible (with improvements ranging from 1.2 to 2.0 times previous performance), but some peculiarities of the platform proved to be challenging to address in a portable way. For example, the peak single-precision floating point throughput requires the use of packed math operations on the AMD CDNA2 architecture [3], but the AMD compiler was not able to generate such instructions, while the manual code transformations that were required were difficult to integrate into the multi-platform SYCL kernels without leading to excessive complexity. Hence, such complex optimisations have been postponed. We have reported these issues to the vendor and hope such manual architecture-specific optimisations will not be necessary with future compilers.

The GROMACS code relies on its SYCL portability backend for AMD support using the hipSYCL application programming interface (API) [4], which adds an abstraction layer on top of the base AMD ROCm stack. Due to the high



Above: GROMACS 2023.1 intra-node strong scaling of the STMV benchmark system (1M atoms) on a Dardel-GPU node

iteration rates (typically less than 1 millisecond) and many small tasks characteristic to molecular dynamics and GROMACS, this workload puts a lot of pressure on the GPU runtime and overheads accumulate [5]. Therefore, the team has focused on identifying and eliminating such overheads. On the one hand, we have reported ROCm/HIP issues and got fixes for them. On the other hand, we have identified limitations of the SYCL abstraction layer and worked closely with the hipSYCL team to reduce related overheads. This motivated certain custom programming model extensions to reduce scheduling overheads which have made a significant impact (up to 25% reduction in wall-time) on the sub-millisecond iteration rate GROMACS simulations. At the same time, these improvements have also become important use-cases for the direction the SYCL standard should progress towards. The team are contributing to this through our SYCL Advisory Panel membership and its feedback process.

The GROMACS team has also worked on improving feature-completeness which, in the latest release, is on par with the most mature CUDA GPU backend when it comes to single-GPU, and multi-GPU/multi-node simulations, including using GPU-aware MPI. Therefore, key features of the GROMACS parallelisation, the GPU-resident mode and direct GPU communication, are supported in the latest release, allowing the best use of the accelerator-dense AMD GPU platform and its intra- and inter-node interconnects. The production code has gone through stabilisation and the official GROMACS 2023 release shows good performance, as shown in the graph on the previous page and in the table below, for a few representative benchmark systems, and is typically within 20-25% of the performance of AMD's own GROMACS fork.

Benchmark system	Performance (ns/day)
rnase_cubic (24k atoms)	630.0
adh_cubic (134k atoms)	215.0
STMV (1M atoms)	17.9

*Above: GROMACS 2023.1 single-GPU absolute performance on the Dardel GPU nodes for a range of common biomolecular system sizes*

Despite the small size of the team working on these features, major leaps have been made possible by early investments into portable algorithms (which was discussed in the PDC newsletter mentioned earlier; for additional details, see [6]) and portable standards-based APIs. In addition, close collaboration between GROMACS developers at PDC and other departments at the KTH Royal Institute of Technology was an important contributing factor, as was the contribution of portable algorithms and code by a wide range of hardware vendors, including Intel (through the KTH-Intel OneAPI Centre of Excellence) and NVIDIA (as part of the GROMACS-NVIDIA co-design collaboration).

## References

1. GROMACS 2023 release highlights <https://manual.gromacs.org/2023/release-notes/2023/major/highlights.html>
2. <https://bioexcel.eu/webinar-whats-new-in-gromacs-2023-2023-04-25>
3. AMD CDNA2 white paper. <https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>
4. hipSYCL project page. <https://github.com/illuhad/hipSYCL>
5. A. Alekseenko and S. Páll. 2023. Comparing the Performance of SYCL Runtimes for Molecular Dynamics Applications. In Proceedings of the 2023 International Workshop on OpenCL (IWOCL '23). Association for Computing Machinery, New York, NY, USA, Article 6, 1–2. <https://doi.org/10.1145/3585341.3585350>
6. S. Páll, A. Zhmurov, P. Bauer, M. Abraham, M. Lundborg, A. Gray, B. Hess & E. Lindahl. “Heterogeneous Parallelization and Acceleration of Molecular Dynamics Simulations in GROMACS.” The Journal of Chemical Physics 153, no. 13 (7 October 2020): 134110. <https://doi.org/10.1063/5.0018516>





# Dardel Updates

Gert Svensson, PDC

This year a number of updates have been made to Dardel. The interconnect has been upgraded from 100 Gbit/s to 200 Gbit/s and is now called Slingshot 11. This means that all the network adaptors have been changed to a new version. PDC has also upgraded the disk system with 50% more capacity, both in terms of the amount of data that can be stored and in relation to the metadata capacity. The water that was used for cooling in the first phase of the Dardel system (that is, the CPU partition) has been changed to a glycol-water mix to avoid biological growth in the system without using any dangerous pesticides.

Compared to when the first phase of Dardel was initially installed, the number of CPU nodes has been increased from 518 to 1270. PDC recently performed a new run of the High-Performance Linpack (HPL) Benchmark for the upgraded system to check the performance. The capacity of the CPU partition of Dardel has increased from 2.28 to 4.08 PFLOPS. In the latest TOP500 list from May 2023, the Dardel CPU partition has jumped way up to place 153 (see <https://top500.org/system/180013>) from place 345 in



Above: An HPE technician wearing protective gear while overseeing the process of changing the cooling water

Below: The used cooling water containing biocides is collected in big tanks and sent away to be disposed of.



the previous (November 2022) list. The Dardel graphics processing unit (GPU) partition has an unchanged capacity of 8.26 PFLOPS and is now in place 77 globally (see <https://top500.org/system/180126>) and comes in at number 25 in the European systems. The GPU partition is still the fifth most energy-efficient system in the world on the Green500 list (see <https://top500.org/lists/green500/2023/06>).

What remains to be done in the near future is to update the software for the Slingshot communication and the disk system, which will probably take place before the summer.

Overall, all the software in the system needs an update. PDC is discussing with HPE how to do this with minimal user impact. One idea is to install a new small test system at PDC with the latest software pre-installed. This would make it possible to upgrade to and test the latest version without any major impact on the current system. This would also simplify the installation and testing of future new releases.



# EuroHPC Summit 2023

Dirk Pleiter, CST, KTH

For many years the European supercomputing community, which had organised itself in PRACE, met regularly at the PRACEdays conferences. After an interim format called the European HPC Summit Week, the EuroHPC Summit was established with the EuroHPC Joint Undertaking as the organiser. The most recent EuroHPC Summit was held in March 2023 and these summit conferences will continue to take place on an annual basis in the country of the EU presidency. This year it was Sweden's turn to welcome Europe's supercomputing community in Gothenburg.

The plenary programme of the EuroHPC Summit 2023 had a strong focus on the state of the EuroHPC infrastructure. Most of the supercomputers resulting from the first round of EuroHPC investments are now operational and available for high-performance computing (HPC) users. Two systems, namely LUMI in Finland and Leonardo in Italy, made it to leading positions on the TOP500 list of the world's fastest supercomputers. The Research Centre Jülich presented the status of preparations for the next milestone, namely the deployment of the first EuroHPC exascale system. Another important pillar of EuroHPC is research and innovation efforts supporting the European Commission's (EC) strategy on digital sovereignty. During the plenary programme, the status of a major initiative aimed at developing European processors and compute accelerators based on the RISC-V instruction set architecture was



Above: EUMaster4HPC students during an award ceremony at the EuroHPC Summit, Gothenburg, 20-23 March 2023

© Julie de Bellaing

discussed. Another focus topic was the status of the implementation of the European quantum computing strategy.

The plenary programme on the opening day offered the opportunity to present a Swedish perspective on HPC and quantum computing. Here, the presentation of the new National Academic Infrastructure for Supercomputing in Sweden (NAISS) and the work of the Wallenberg Centre for Quantum Technology (WACQT) aligned well with the focus of the conference. The closing day provided the opportunity to put a new educational endeavour funded by EuroHPC in the spotlight, namely the education of Masters in HPC students through the EUMaster4HPC project. KTH is part of that project.

Twelve parallel sessions allowed this very diverse and broad community to present many new results and discuss upcoming challenges. Two examples were the sessions on upcoming exascale applications and efforts towards enabling digital twins on HPC infrastructures. This allowed KTH researchers to discuss the work within the new centre of excellence Plasma-PEPSC, where applications for simulating plasmas are prepared for upcoming exascale systems, as well as the BioDT project. The latter works with the biodiversity community on realising digital twins, which is a rather new way of using HPC.

With about 500 on-site attendees, which is a record number of participants, the event demonstrated the growing interest in supercomputing in Europe. Overall it was considered a great success. This sets high expectations for next year's EuroHPC Summit, which will take place in Antwerp. You can read more about this year's summit and projects here:

- EuroHPC Summit programme and slides <https://www.eurohpcsummit.eu>
- EUMaster4HPC project <https://eumaster4hpc.uni.lu>
- BioDT project <https://bioldt.eu>
- PLASMA-PEPSC project <https://plasma-pepsc.eu>

# Co-design Efforts for Major Performance Improvements in GROMACS GPU-Accelerated Parallelisation

Szilárd Páll, PDC, & Alan Gray, NVIDIA

The GROMACS open-source molecular dynamics (MD) code is among the early high-performance computing (HPC) scientific codes to make efficient use of graphics processing unit (GPU) accelerators with its GPU support having been available for more than decade. The heterogeneous parallelisation of the GROMACS code has gone through a gradual evolution, from offloading only the most compute-intensive parts of a simulation to increasingly moving most computation from the CPUs to the GPUs to make efficient use of modern accelerator-based systems.

Collaboration between academic and industry partners, including hardware vendors, has been an important driver in terms of hardware support for GROMACS. The GROMACS team's close collaboration with NVIDIA on GPU parallelisation has been of great benefit from the very early days of GPU support in relation to GROMACS.

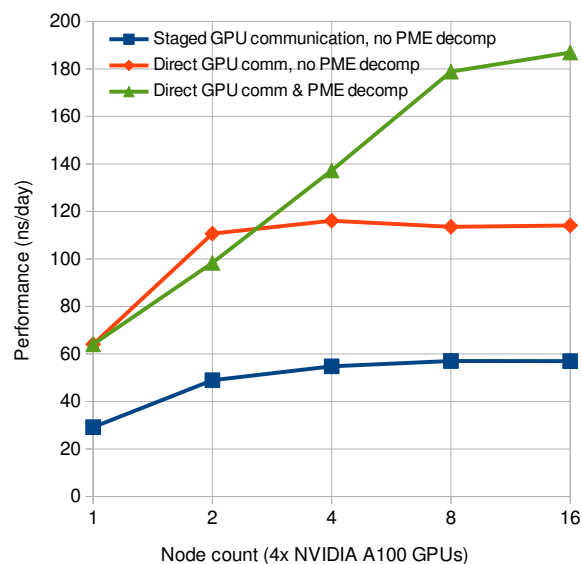
Funding through a Swedish Foundation for Strategic Research (SSF) Swedish Exascale Computing Initiative (SECI) project (from 2016 to 2022) enabled PDC-based researchers and members of the GROMACS team to focus on improving molecular dynamics algorithms for exascale HPC architectures. Thanks to this funding and shared objectives, we established a co-design collaboration between the GROMACS team and NVIDIA with a long-term goal to advance molecular dynamics algorithms and redesign the GROMACS parallelisation, adapting it to modern GPU-accelerated architectures. At the same time, this co-design effort was also successful at providing feedback to the NVIDIA software teams about the requirements of molecular dynamics algorithms and GROMACS implementation.

One of the major focuses of this co-design collaboration has been improving the strong

scaling of molecular simulations on multi-GPU machines. Portable implementations have been developed in GROMACS, including GPU-resident parallelisation with direct-GPU communication [1]. Research efforts have also focused on more efficient task scheduling on GPUs and explored new ways to express fine-grained parallelism in molecular simulation by employing task graph-based scheduling [2].

The direct GPU communication layer, which allows making efficient use of high-performance interconnects and helps avoid staging data movement through the CPU, was extended with GPU-aware MPI support in the 2022 GROMACS release. Building on this, recent efforts focused on removing a strong scaling bottleneck which limited the scalability of GROMACS on recent HPC architectures where most of the computational power is provided by GPUs.

The particle mesh Ewald (PME) algorithm is used most commonly to compute long-range electrostatics interactions in MD. This relies on fast Fourier transform (FFT) operations, which are hard to parallelise and scale, especially the typically small



Above: GROMACS 2023 STMV benchmark strong scaling on NVIDIA Selene [5] cluster, four A100 GPUs per node used (since this resembled standard HPC clusters)

Direct GPU communication (red) offers up to 2x improvement over the staged communication (blue), but cannot scale beyond two nodes due to the single PME GPU bottleneck. GPU PME decomposition (green), while slightly slower on two nodes, allows scaling further to 8-16 nodes and gives a 1.7x performance improvement over the single PME GPU case (red).

transforms required for biomolecular MD. While computing the short-range interactions could be scaled efficiently across multiple GPUs and multiple nodes with GROMACS, until recently the entire PME computation had to use a single (although dedicated) GPU which would quickly limit the scaling of such simulations (typically to 4 to 8 GPUs). Alternatively, the multi-node simulation could use CPUs for the entire PME computations, but this also has limited performance and scaling on modern architectures.

To lift this limitation in our co-design project, we have developed a PME decomposition algorithm better suited for GPUs (which relies on grid overlap reduction, instead of particle redistribution prior to spreading to the grid) and added support for distributed FFT computations using both the portable HeFFTe [3] library, as well as the recently introduced cuFFTmp library [4]. The former allows portability, while the latter is a more performant vendor library, which has been optimised with the GROMACS and MD use-case as one of its targets.

Thanks to the PME GPU decomposition, the recent 2023 release of GROMACS is able to distribute PME computation across multiple GPUs within or across compute nodes, thereby lifting a major scaling limitation and offering major performance improvements in simulations typically from 6 to 8 GPUs.

## References

1. S. Páll, A. Zhmurov, P. Bauer, M. Abraham, M. Lundborg, A. Gray, B. Hess & E. Lindahl, Heterogeneous Parallelization and Acceleration of Molecular Dynamics Simulations in GROMACS. *The Journal of Chemical Physics* 153, no. 13 (7 October 2020): 134110. <https://doi.org/10.1063/5.0018516>
2. <https://developer.nvidia.com/blog/a-guide-to-cuda-graphs-in-gromacs-2023>
3. heFFTe <https://icl.utk.edu/fft>
4. cuFFTmp <https://docs.nvidia.com/hpc-sdk/cufftmp/index.html>
5. NVIDIA Selene <https://www.top500.org/system/179842>



EuroCC National Competence Centre Sweden

## Second Phase of ENCCS

Thor Wikfeldt, ENCCS

ENCCS (<https://enccs.se>), the Swedish node of the EuroCC project, has entered its second funding period with a renewed focus on industry-driven solutions. Continuing on the path set out in the first phase of the project, the aim of ENCCS is to increase awareness and competence in high-performance computing (HPC) and assist companies, public authorities, and academia to access supercomputers to scale up their simulations, incorporate artificial intelligence (AI) and high-performance data analytics (HPDA) methods, and learn about emerging technologies like quantum and neuromorphic computing.

In this new phase, ENCCS services are built on three main pillars which have been designed to benefit a wide range of users. First, ENCCS will continue to provide training in multiple areas, primarily on key topics in HPC programming but also on AI, HPDA, quantum and neuromorphic computing, as well as domain-specific topics in quantum chemistry, materials science, computational fluid dynamics and so forth. The training program is designed to help users with different background experience to learn how to apply HPC tools and techniques to their research and development projects. ENCCS has an open-source lesson library with training material suitable for self-learning, which can be accessed at <https://enccs.se/lessons>. Second, ENCCS provides free-of-charge consultancy on software optimisation, parallelisation and porting to graphics processing units (GPUs) to enable users to maximise the performance of their code on HPC systems. Finally, ENCCS provides support to write applications and get access to EuroHPC JU supercomputers and provides guidance to new users on how to access and effectively use these powerful machines for free. For an overview of past successful projects, some highlights are presented at <https://enccs.se/industry-public-administration>.



ENCCS is hosted by the RISE Research Institutes of Sweden and Linköping University. This partnership provides ENCCS with access to state-of-the-art facilities and resources to deliver its services. The collaboration between RISE and Linköping University also provides a unique opportunity for ENCCS to engage with experts from different fields across academia, industry and the public sector, ensuring that the services offered by ENCCS are tailored to the needs of a wide range of users.

## Artificial Intelligence Frameworks on the New Dardel GPU Partition

**Xavier Aguilar, PDC**

The new Dardel graphics processing unit (GPU) nodes are finally online and open for all users. Each one of these nodes includes an AMD EPYC™ processor with 64 cores and four AMD Instinct™ MI250X GPUs. These AMD GPUs can perform up to 383 TFLOPS for half-precision floating-point format (FP16) and 95.7 TFLOPS for single-precision floating-point format (FP32) when using matrix operations, which are fundamental building blocks for machine learning. So these GPU nodes are a very suitable platform for running artificial intelligence (AI) workloads at PDC.

PDC has preinstalled several containers with the most popular AI frameworks, for example, PyTorch[1] and Tensorflow[2]. These containers

have multi-GPU support, and therefore, users have the possibility to train and run their AI models on Dardel using multiple GPUs in parallel. In addition, users are able to download these containers to their workstations, customise them as they please, and then upload them back to Dardel. Furthermore, if a containerised solution is not the best fit for you, you can instead have your own local installation of your preferred AI framework. If that is the case, just get in touch with PDC's Support team (via email to [support@pdc.kth.se](mailto:support@pdc.kth.se)), and we will help you to get your framework set up and running.

We have performed several experiments at PDC to measure the performance of one of these AI frameworks with the AMD GPU backend using three popular AI models. Moreover, we ran the same experiments on another supercomputer cluster using NVIDIA GPUs in order to compare both GPU backends. The first table below presents the hardware and software specifications for these two clusters. The second table shows an overview of the specifications for the GPU cards used in our experiments: AMD MI250Xs and NVIDIA A100s.

The experiments consisted of training three popular models for image classification (Inception-V3, Resnet-101, and Resnet-50) with synthetic data. The training performance was measured using the number of images per second achieved and averaged over ten epochs. (Note that the first epoch is omitted to avoid initialisation overheads.) We measured the performance when using fixed precision (FP32) and when using mixed

	AMD Cluster	NVIDIA Cluster
GPU	4 × MI250Xs	4 × A100s
CPU	1 × AMD EPYC™ 7A53 64-core	2 × 32-core Intel Xeon Gold 6338
Memory per node	512 GB	256 GB
Framework	Pytorch 1.13.0 & Rocm 5.4	Pytorch 1.13.0 & CUDA 11.7

Above: System specifications for the clusters used in our experiments

	AMD MI250X	NVIDIA A100
Peak FP32 performance	95.7 TFLOPS	19.5 TFLOPS / 156 TFLOPS
Peak FP16 performance	383 TFLOPS	78 TFLOPS / 312 TFLOPS
Memory	128 GB	40 GB
Peak Memory bandwidth	3276.8 GB/s	1555 GB/s

Above: GPU technical specifications

The peak performance for the NVIDIA A100 includes the plain float performance (number on the left) as well as the peak performance using the Tensor cores (number on the right).

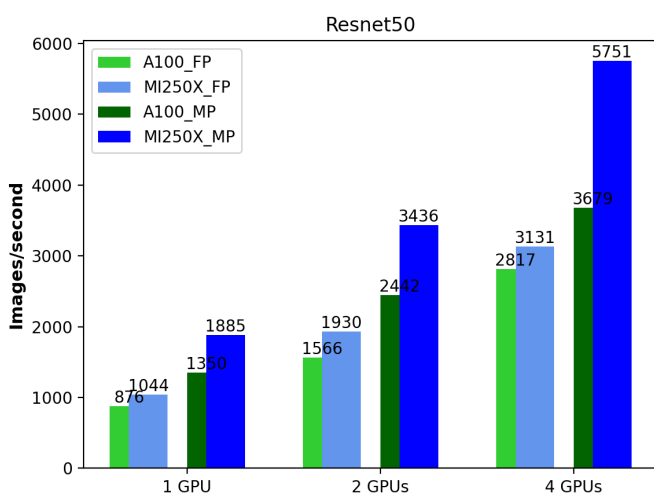
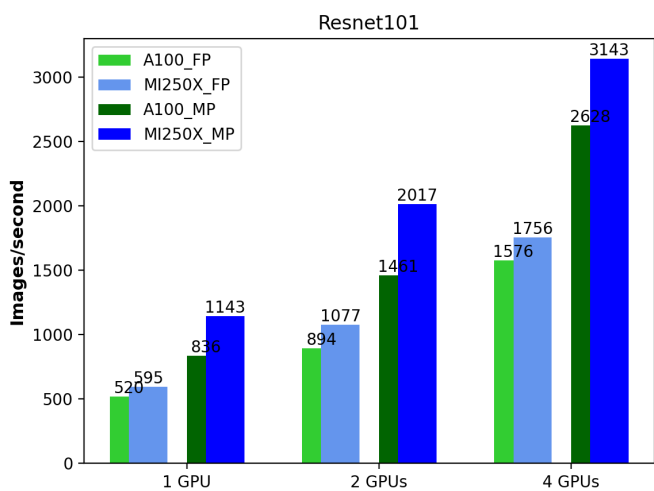
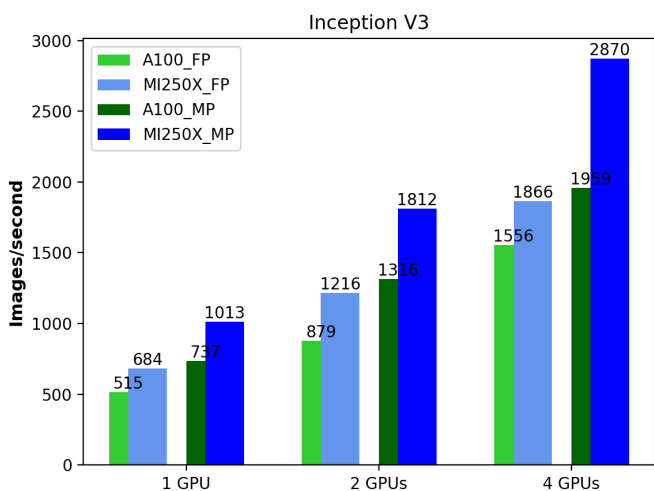
precision – in this mode, PyTorch converts some operations to FP16 to optimise computations. We used the same hyperparameters for the models on both systems. Our experiments were based on previous benchmarking work performed by

SURF[3]. The original code for these experiments can be found in [4].

The graphs to the left show the number of images per second for each of these models. The green bars are the NVIDIA A100s, and the blue bars are the AMD MI250Xs. The light shades of blue and green are for the fixed precision experiments (FP32), and the darker shades for the mixed precision ones.

We can see that for the Inception model, the MI250X always outperforms the A100, with the difference in throughput when using mixed precision and all four GPUs being especially remarkable. For the Resnet models, we can again see the MI250X outperforming the NVIDIA A100 in all cases, as a result of its higher specifications and higher theoretical peak performance, as well as a more mature AMD software stack. The figures also show that, for both backends, using mixed precision mode always helps to speed up computations quite considerably.

In summary, the results demonstrate that the new AMD MI250X cards at PDC have an impressive performance for AI model training. This new AMD hardware, together with AMD’s rapidly evolving GPU software stack and the fact that many AI frameworks are compatible with it, might turn AMD into an important player in the field of GPU computing for HPC and AI and might break the monopoly that NVIDIA has had for many years. Having more players in the GPU market, and especially in the GPU for machine learning market, opens up the number of choices for users and, in addition, will potentially hasten the progress of AI frameworks. Take advantage of this opportunity and start using Dardel and its AMD MI250Xs for your machine learning workloads!



Above: Number of images per second when training an Inception-v3, a Resnet101 and a Resnet50 model with synthetic images for ten epochs using A100s vs MI250Xs. The suffix “\_FP” in the legend stands for “fixed precision”, and the suffix “\_MP” for mixed precision.

## References

- <https://pytorch.org>
- <https://www.tensorflow.org>
- <https://communities.surf.nl/dataverwerking-en-analyse/artikel/performance-comparison-of-image-classification-models-on>
- [https://github.com/sara-nl/nvidia\\_amd\\_gpu\\_comparison](https://github.com/sara-nl/nvidia_amd_gpu_comparison)

# Nascent NAISS

**Björn Alling, Director, NAISS, Linköping University**

The work on building up the National Academic Infrastructure for Supercomputing in Sweden (NAISS) continues. Representatives from NAISS and Linköping University (LiU) recently attended a hearing about our application from Sweden to host a mid-range EuroHPC system. If it is successful, the EU will fund 35% and Sweden 65% of the system. A decision from EuroHPC is anticipated in mid-June. In parallel, work to secure the full Swedish co-funding is ongoing. Having such a system in Sweden would strongly benefit researchers here from the start of 2025.

NAISS is building up a branch structure to provide user support (up to mid-level) and training through experts working across Sweden. Agreement texts are being prepared and negotiations with more than ten other universities are ongoing. In this context, it is important to stress that NAISS has a responsibility to build a unified support organisation for the benefit of all our users – regardless of their field of research or where in Sweden they work. So determining an appropriate structure and approach for the branch organisation within NAISS is a complex puzzle that needs to be solved optimally. We hope that many of you who work within high-performance computing (HPC) and related fields want to be a part of NAISS and contribute to building for the future of HPC in Sweden.

The work on creating the contents and functionality of the NAISS website is well underway with the launch of the full version expected before midsummer. Currently, an early version with the most crucial information is available online at <https://www.naiss.se>. NAISS newsletters and the minutes from the steering committee meetings can be found at <https://www.naiss.se/newsletter> and <https://www.naiss.se/minutes/steering-group>, respectively.

The right ways for researchers to now acknowledge NAISS (and previously SNIC) resources are available at <https://www.naiss.se/policies/acknowledge>.

## PDC-Related Events

### Dardel Inauguration

**23 August, KTH main campus, Stockholm**

Details will be available soon at <https://www.pdc.kth.se/about/events>.

### Save the date: Introduction to GPUs Course

**15 & 22 September (afternoons, online) & 12-13 October (at PDC)**

Learn about GPU systems and how to adapt research code to run on GPUs!

### CodeRefinery Workshop

**19-21 and 26-28 September (6 half days, online)**

Details will be available soon at <https://www.coderefinery.org>.

## HPC Sources

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

### HPC in Sweden and Scandinavia

- ENCCS  
<http://enccs.se>
- NAISS  
<https://naiss.se>
- NeIC  
<http://neic.no>
- SeRC  
<https://e-science.se>
- SeSE  
<http://sese.nu>

### European HPC ecosystem

- EOSC  
<https://eosc-portal.eu>
- ETP4HPC  
<https://www.etp4hpc.eu>
- EuroHPC  
<https://eurohpc-ju.europa.eu>
- HPC in Europe  
<https://hpc-portal.eu>
- LUMI  
<https://www.lumi-supercomputer.eu>
- PRACE  
<https://www.prace-ri.eu>

### A selection of projects that PDC is involved with

- BioExcel CoE  
<https://bioexcel.eu>
- CEEC  
<https://www.ceec-coe.eu>
- EBRAINS  
<https://ebrains.eu>
- EUMaster4HPC  
<https://eumaster4hpc.uni.lu>
- EXCELLERAT  
<https://www.excellerat.eu>
- PerMedCoE  
<https://permedcoe.eu>
- Plasma-PEPSC CoE  
<https://biodt.eu>

### HPC news sources

- HPCwire  
<http://www.hpcwire.com>
- insideHPC  
<https://insidehpc.com>