CoE European Excellence in HPC Applications

NEWSLETTER

September 2019

WELCOME

Bernd Mohr

POP CoE
dissemination manager

"From Biomedicine to Material Science – most CoEs cover a specific application area. This is why we find that POP ideally complements the EU HPC CoE landscape, as it concentrates on a fundamentally desired property of HPC applications: **PERFORMANCE**!

After a very successful first phase of the POP project from October 2015 to March 2018, where we performed over 160 performance audits, performance plans, and proof-of-concept services for our customers, the project secured funding for a second 3-year phase starting December 2018. For our past and potential future customers nothing major changed: We still will provide free performance optimisation and productivity services for academic and industrial codes in all domains. We still offer our services free of charge to organisations inside the EU. Finally, we will continue with our successful training and tuning workshops programme including our performance analysis webinars.

For the second phase, teams from the Performance Tools team at UVSQ in France and IT4Innovations at VSB-TUO in the Czech Republic joined our group of performance experts. The other members are BSC, HLRS, JSC, NAG, RWTH Aachen, and Teratec.

To improve our service, the second phase of the project includes a few minor changes and additions. In a first step, we simplified our service offering by integrating Performance Audits and Plans into one new service called Performance Assessment. In addition, we will create a co-design data repository that includes statistics about common performance issues of HPC applications as well as micro-kernels extracted from real applications, each characterising fundamental performance behaviour. Hardware architects or system software designers from other EU projects will be able to get quantitative information of how to estimate the potential impact of an architectural or system software approach they may be developing. Finally, we aim to work even closer and more systematically with the application developers of the other CoEs."

CONTENTS

DON'T MISS!

First EoCoE-II Evaluation Performance Workshop
October 7 - 10, 2019;
Registration open until October 7

CompBioMed Conference 2019
September 25 – 27, 2019,
Registration open until September 15

E-CAM - AN INDUSTRY SUCCESS STORY

Mesoscale simulation of billion atom complex systems using thousands of GPGPU's, an industry success story

Abstract

E-CAM research software developer Jony Castagna working at the Science and Technology Facilities Council (STFC) in Daresbury, recounts his rewrite of DL_MESO allowing the simulation of billion atom systems on thousands of GPGPUs.

What is DL_MESO and why was it important to port it to massively parallel computing platforms?

DL_MESO is a software package for mesoscale simulations developed by M. Seaton at the Hartree Centre [1, 2]. It is basically made of two software components: a Lattice Boltzmann method solver, which uses the Lattice Boltzmann equation discretize on a lattice (2D or 3D) to simulate the fluid dynamic effects of complex multiphase systems; and a Dissipative Particle Dynamics (DPD) solver based on particle method where a soft potential, together with a coupled dissipation and stochastic forces, allows the use of Molecular Dynamics but with a larger time step.

The need to port DL_MESO to massively parallel computing platforms arose because often real systems are made of millions of beads (each bead representing a group of molecules) and small clusters are usually not sufficient to obtain results in brief time. Moreover, with the advent of hybrid architectures, updating the code is becoming an important software engineering step to allow scientist to continue their work on such systems.

How well were you able to improve the scaling performance of DL_MESO with multiple GPGPU's, and as a consequence, how large a system can you now treat?

The current multi-GPU version of DL_MESO scales with an 85% efficiency up to 2048 GPUs equivalent to about 10 petaflops of performance double precision (see Fig. 1 reproduced from E-CAM Deliverable 7.6 [3]). This allows the simulation of very large systems like a phase mixture with 1.8 billion particles (Fig. 2). The performance has been obtained using the PRACE resource Piz Daint supercomputer from CSCS.

What are the sorts of practical problems that motivated these developments, and what is the interest from industry (in particular IBM and Unilever) ?

DPD has the intrinsic capability to conserve hydrodynamic behavior, which means it reproduces fluid dynamic effects when a large number of beads is used. The use of massively parallel computing allows the simulation of complex phenomena like shear banding in surfactants and ternary systems present in many personal care, nutrition, and hygiene products. DL_MESO has been used intensively by IBM Research UK and Unilever and there is a long history of collaboration with Hartree Centre still going on.

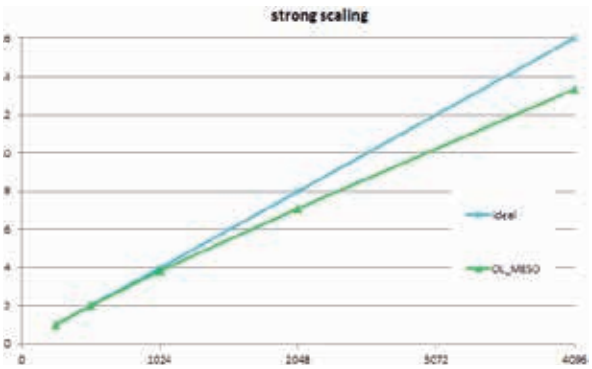


Figure 1. Strong scaling efficiency of DL_MESO versus the number of GPGPU for a simulation of a complex mixed phase system consisting of 1.8 billion atoms.

Where can the reader find documentation about the software developments that you have been doing in DL_MESO?

Mainly on the E-CAM modules dedicated to DL_MESO that have been reported on Deliverables 4.4 [4] and 7.6 [3], and also on the E-CAM software repository here .

Read the full story at https://www.e-cam2020.eu/gpu-re-write-of-dl-meso/
Get in touch: info@e-cam2020.eu

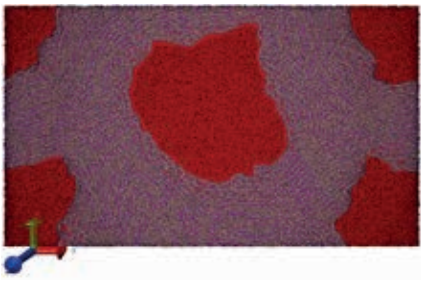


Figure 2. Snapshot of the simulated system.

References

[1] http://www.cse.clrc.ac.uk/ccg/software/DL_MESO/ [2] M. A. Seaton, R. L. Anderson, S.Metz, and W. Smith, "DL_MESO: highly scalable mesoscale simulations,"Molecular Simulation, vol. 39, no. 10, pp. 796–821, Sep. 2013. [3] Alan O'Cais, & Jony Castagna. (2019). E-CAM Software Porting and Benchmarking Data III (Version 1.0). Available in Zenodo: https://doi.org/10.5281/zenodo.2656216 [4] Silvia Chiacchiera, Jony Castagna, & Christian Krekeler. (2019). D4.4: Meso- and multi-scale modelling E-CAM modules III (Version 1.0). Available in Zenodo: https://doi.org/10.5281/zenodo.2555012

PREPARING CLOUD PHYSICS FOR EXASCALE

Accelerating cloud physics and atmospheric models using GPUs, KNLs and FPGAs

The Met Office NERC Cloud model (MONC) is an atmospheric model used throughout the weather and climate community to study clouds and turbulent flows. This is often coupled with the CASIM microphysics model, which provides the capability to investigate interactions at the millimetre scale and study the formation and development of moisture. One of the main targets of these models is the problem of fog, which is very hard to model due to the high resolution required – for context the main UK weather forecast resolves to 1km, whereas the fog problem requires 1metre or less.

A major driver here is Heathrow airport. Due to its location it is prone to fog, which requires the spacing between aircraft to be increased. As an airport at 98% capacity, this increased spacing causes delays and has a financial impact, but crucially it is unable to accurately predict when the fog will clear. In such a situation you don't just need accuracy, but also for the forecast to complete in a timely manner (it's useless if the results of a day's fog modelling at Heathrow are available 3 months later!).

Exascale has a significant role to play in accelerating these models and providing new and important capabilities that not only benefit scientists but also wider industry. Whilst MONC and CASIM are prime models for simulating fog, they are both heavily computationally intensive and it is important to find ways to accelerate them. We have focused on three technologies: GPUs, KNLs and FPGAs. These are not mutually exclusive as they potentially suit different parts of the code and we are exploring the potential role that each has to play.

GPU acceleration

Initially we focused on using OpenACC for GPU acceleration. In MONC we were able to isolate specific computationally intensive kernels whereas in CASIM we were forced to offload the entirety of the microphysics model due to the tightly coupled nature of the code. While there were several caveats and lessons learnt, experiments on Piz Daint (P100 GPUs) demonstrated a significant benefit to using GPUs. This was most significant for CASIM, which has the most computationally intensive aspects, where using GPUs reduced the runtime by six times compared to the Haswell CPU.

Knights Landing

We then focused on Knights Landing (KNLs). This technology has been deprecated by Intel but it is still useful to consider because of the high degree of vectorisation required to get good performance, a trend likely to continue in future generation Xeon CPUs. KNLs have much more memory readily available to them and we found that we were able to run much larger systems. This is important on the KNL, because smaller systems performed comparative to a node of ARCHER (Ivy Bridge), but when we go to much larger systems the KNL starts to out-perform Ivy Bridge more significantly and reduce the runtime by around 40%. However when the same experiment was run on a node of the Met Office's XC40 (Broadwell CPU) we found that this later generation CPU outperforms the KNL quite considerably. Whilst it doesn't look good for using KNLs to accelerate this model, it is still an important result.

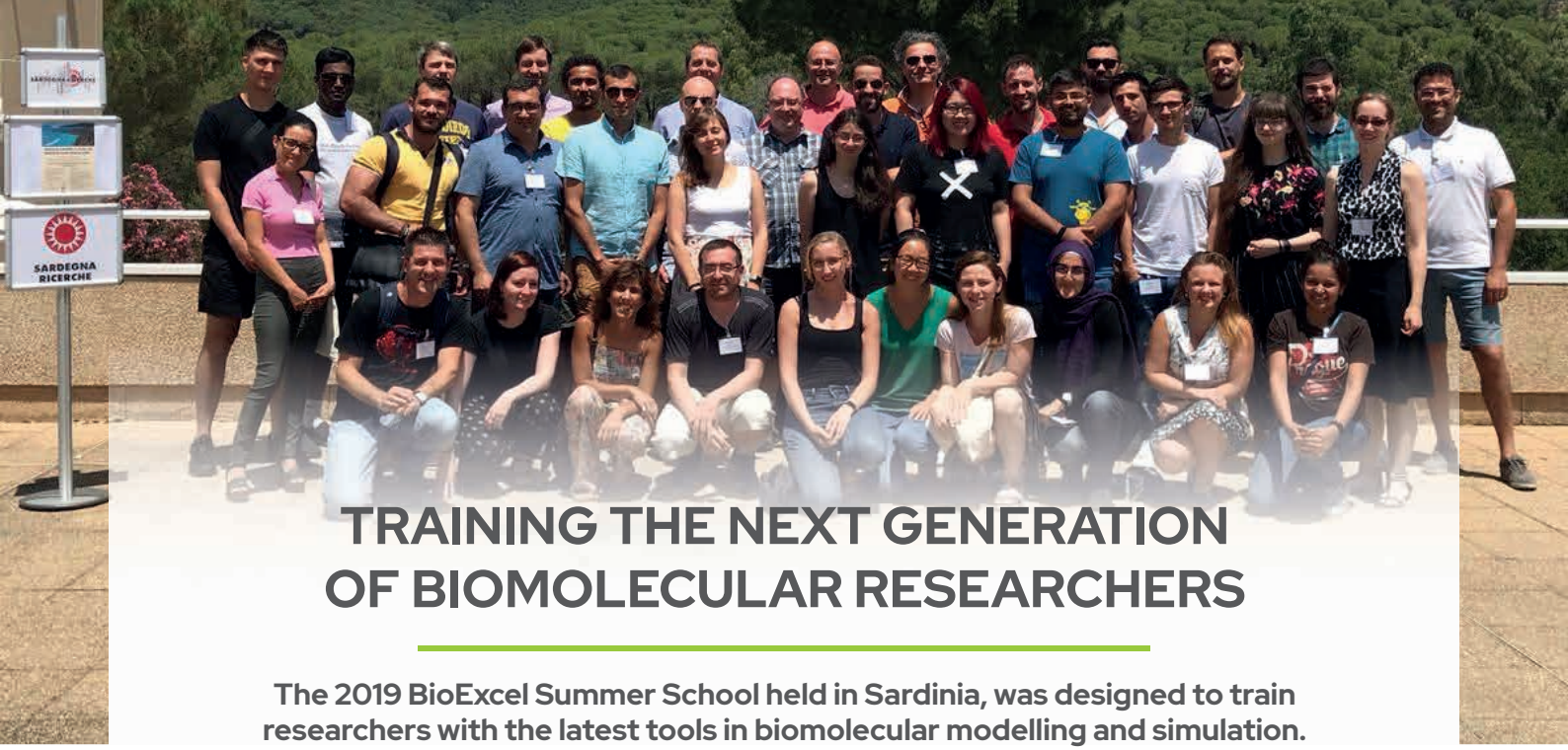
So currently we have GPUs in the lead for accelerating the model, but there is a significant downside here – the energy consumption.

GPUs require a very significant amount of energy and this makes them difficult (and expensive) to deploy in the field. The latest piece of work we have done, as part of EXCELLERAT (the European Centre of Excellence for Engineering Applications), is to port the MONC computational kernels over to FPGAs, using High Level Synthesis, and run them on our Kintex UltraScale which is mounted via a PCIe card.

This is still work in progress, but we can demonstrate some interesting performance characteristics in comparison to the CPU, with each of our kernels roughly providing the floating point performance of a single Broadwell core at the moment. Undoubtely there is more performance headroom we can exploit at the kernel level, but what is currently hitting us with this FPGA version is the cost of data transfer (DMA to and from the PCIe card) and we have a few ideas how to ameliorate this.

The result of this work, driven by real world use-cases, is insight into the applicability of three major technologies available for accelerating HPC codes for atmospheric modelling.

This post has appeared at the EPCC website before and is based on the abstract of a talk at the PASC mini-symposium 'Modelling Cloud Physics: Preparing for Exascale' (Zurich, 13 June 2019).



TRAINING THE NEXT GENERATION OF BIOMOLECULAR RESEARCHERS

The 2019 BioExcel Summer School held in Sardinia, was designed to train researchers with the latest tools in biomolecular modelling and simulation.

The 2019 BioExcel Summer School was held for the second consecutive year between 30th June – 5th July in Sardinia, aimed at early career researchers who are using or planning to use biomolecular modeling and simulation in their everyday research. As the centre primarily works with core developers of widely used tools, the summer school focused on providing hands-on training to support the biomolecular research community. BioExcel travel grants were also offered to selected participants to reduce financial barriers. Lectures were given on topics such as molecular dynamics simulations and biomolecular docking by experts in the field. Participants also had the opportunity to apply their theoretical knowledge through computer practicals on use cases integrating BioExcel flagship software - GROMACS, HADDOCK, PMX and CP2K. Networking events facilitated new discussions which helped bridge the gap between early career and experienced researchers.

Anna Verdino, PhD student at the University of Salerno says, "Thanks to this school, I have been able to meet people, exchange ideas and create new opportunities of collaboration that I think will be extremely useful for me in the future."

With the constant evolution of HPC and HTC techniques, BioExcel's mission is to enable researchers to exploit the power of computing e-infrastructure by providing support and training to beginner and advanced users.

HPC COE COUNCIL ESTABLISHED

The HPC CoE Council provides a joint voice of the CoEs in the European HPC ecosystem to strengthen the impact of HPC applications.

Elisa Molinari

Applications in High Performance Computing (HPC) help tackle some of the world's biggest societal, scientific, and economic challenges and therefore take a pivotal place in the European HPC Ecosystem alongside the development of Exascale technologies and provision of access to extreme-scale infrastructure. Many important applications are being supported through EU Centres of Excellence (CoEs) in HPC. On May 17, during the EuroHPC Summit Week (EHPCSW) in Poznán and with the support of the FocusCoE Collaborative Support Action, the CoEs established the HPC CoE Council (HPC3). The aim of the HPC3 is to provide a common forum for the CoEs to promote the development and usage of high-end HPC applications in all scientific fields. With the strong impulsion of EuroHPC and the upcoming arrival of (pre-)exascale machines, it is of primordial importance to strengthen the development of key applications that will enable science to be done on these machines.

Edouard Audit

To animate the HPC3, a council office was elected with Edouard Audit (top left) of the French Alternative Energies and Atomic Energy Commission (CEA),

Erwin Laure

who represents the Energy oriented CoE (EoCoE), as Chairperson, Elisa Molinari (right, CNR and Uni Modena, MaX CoE for Materials design at the Exascale) as Vice Chairperson and Erwin Laure (bottom left, KTH Royal Institute of Technology, BioExcel – CoE for Computational Biomolecular Research, see left photo) as General Secretary.

Currently, the HPC3 encompasses all active CoEs. In the future, the HPC3 will be open to newly created CoEs and to other entities strongly involved in application development and support.

EVENT HIGHLIGHTS

Summer School on Advanced Materials and Molecular Modelling with Quantum ESPRESSO

The aim of the school is to introduce students, postdocs, and other researchers to materials and molecular modelling with Quantum ESPRESSO. The school will cover basic concepts as well as recent advances and developments, with emphasis on density-functional-theory based methods and high-performance computing.

Ljubljana, Slovenia

First EoCoE-II Evaluation Performance Workshop

Performance engineering is about developing a thorough understanding of the interactions between software and hardware. In the first part of the workshop, participants deal with the core, socket, and node level, where the code gets executed that does the actual computational work. Once the architectural requirements of a code are understood and correlated with performance measurements, the potential benefit of optimizations can often be predicted.

Erlangen, Germany

CompBioMed Conference 2019

The conference will address all aspects of the rapidly burgeoning domain of computational biomedicine, from genome through organ to whole human and population levels, embracing data driven, mechanistic modelling and simulation, machine learning and combinations thereof.

London, UK

Euro-Par Conference 2019	26-30 August	Göttingen, Germany
Workshop on Machine learning for Weather and Climate Models	2 September	Oxford, UK
13th Parallel Tools Workshop	2 September	Dresden, Germany
High Performance Computing - Bulgaria 2019	2-6 September	Borovets, Bulgaria
MaX WEBINAR: Bridging the gap between Industry and Materials design in eXascale	4 September	Webinar
Implementing I/O Best Practices to Improve System Performance with Ellexus	9 September	Webinar
Parco Symposium: Tools and Infrastructure for Reproducibility in Data- Intensive Applications	10 September	Prague, Czech Republik
Picking flowers: Hands-on FLEUR	9-13 September	Juelich, Germany
Summer School on Advanced Materials and Molecular Modelling with Quantum ESPRESSO	16-20 September	Ljubljana, Slovenia
PRACE Autumn school 2019 – Big data and HPC	17-20 September	Ljubljana, Slovenia
MaX @Graphene Week 2019	23-27 September	Helsinki, Finland
European Research and Innovation Days	24-26 September	Brussels, Belgium
CompBioMed Conference 2019	25-27 September	London, UK
Graphics Processing Unit (GPU) Programming Hackathon 2019	30 Sep. - 4 Oct.	Lugano, Switzerland
First EoCoE-II Evaluation Performance Workshop	7-11 October	Erlangen, Germany
Advanced Gromacs, HADDOCK + PMX Workshop	9-11 October	Espoo, Finland
ESDW: Mesoscopic simulation models and High-Performance Computing	15-19 October	Espoo, Finland
Hands-on Introduction to HPC for Life Scientists	30 Oct. - 1 Nov.	Birmingham, UK
E-CAM ESDW: Inverse Molecular Design & Inference: building a Molecular Foundry	4-8 November	Dublin, Ireland
The International Conference for HPC Networking, Storage, and Analysis 2019 (SC19)	17-22 November	Denver, United States

CODES

GROMACS

The GROMACS project started in 1995 as one of the first-ever parallel simulation codes, with strong focus on simulation efficiency and generality. It is able to use diverse compute architectures as well as GPUs and runs on a wide variety of platforms. GROMACS-generated trajectories of protein conformations are part of many bioinformatics pipelines; its free energy calculations are a core part of drug development pipelines. The open source package uses neutral territory domain decomposition, dual-list dynamic-pruning for short-ranged interactions and multi-level parallelization to enable scaling both to tens of thousands of nodes on supercomputers and efficient high-throughput computing with accelerators.

Version 2019.3 released in June 2019

AngioSupport

AngioSupport is a toolchain comprising of a segmentation tool and a 1D wave propagation model. The patient geometry is captured from a bi-planar angiogram by segmentation (Coronary Angiographic Analysis Systems, Pie Medical Imaging). An existing 1D wave propagation model of the human vascular system was simplified and extended with the coronary system, as developed at the Eindhoven University of Technology. To simulate the pressure and flow propagation, by feeding the model with patient specific clinical measures such as patient length, weight, heart rate and aortic blood pressure, the pre-op FFR is calculated throughout the patients system. Subsequently the clinician can select standard stent sizes and deploy them in the area they deem affected by disease. Alternatively, they can simulate the CABG option by selecting the location of the anastomosis on the coronary tree.

PUBLICATIONS

Goldbach, L., Vermeulen, B. J., Caner, S., Liu, M., Tysoe, C., van Gijzel, L., ... & Bonvin, A. M. (2019). Folding Then Binding vs. Folding Through Binding in Macrocyclic Peptide Inhibitors of Human Pancreatic α-Amylase. ACS chemical biology, 14, 1751-1759, DOI: 10.1021/acschembio.9b00290

Aranda, J., Terrazas, M., Gómez, H., Villegas, N., & Orozco, M. (2019). An artificial DNAzyme RNA ligase shows a reaction mechanism resembling that of cellular polymerases. Nature Catalysis, 2(6), 544–552.

Terrazas, M., Genna, V., Portella, G., Villegas, N., Sánchez, D., Arnan, C., & Aviñó, A. (2019). The origins and the biological consequences of the Pur/Pyr DNA· RNA asymmetry. Chem, 5(6), 1364-1366, DOI: https://doi.org/10.1016/j.chempr.2019.04.002

COEs OVERVIEW

BioExcel 2 - Biomolecular Research

BioExcel is operating towards advancement and support of the HPC software ecosystem in the life science domain. Research and expertise covers structural and functional studies of the main building blocks of living organisms (proteins, DNA, membranes etc.) and techniques for modelling their interactions ranging from quantum to coarse-grained models up to the level of a single cell.

ChEESE - Solid Earth

In CheEESE, leading European HPC centers, academia, hardware developers, as well as SMEs, industry and public governance bodies such as civil protection are working together to prepare European flagship codes for upcoming pre-Exascale and Exascale supercomputing systems to tackle global challenges in the domain of solid earth.

CompBioMed - Biomedicine

CompBioMed is a user-driven Centre of Excellence in Computational Biomedicine, to nurture and promote the uptake and exploitation of high performance computing within the biomedical modelling community, supporting users in academia, industry and clinical practice.

E-CAM - Method and algorithm deployment

The overall objective of E-CAM is to create, develop and sustain a European infrastructure for computational science applied to simulation and modelling of materials and of biological processes of industrial and societal interest.

EoCoE-II - Energy

EoCoE will use the prodigious potential offered by the ever-growing computing infrastructure to foster and accelerate the European transition to a reliable and low carbon energy supply via targeted support to four carbon-free energy pillars: Meteorology, Materials, Water and Fusion, each with a heavy reliance on numerical modeling.

ESiWACE2 - Weather & Climate

ESiWACE2 aims to link, organise and enhance Europe's excellence in weather and climate modelling to enable leading European weather and climate models to leverage the performance of pre-exascale systems as soon as possible and prepare the weather and climate community to be able to make use of exascale systems when they become available.

COEs OVERVIEW

EXCELLERAT - Engineering

EXCELLERAT's goal is to facilitate the development of important codes for high-tech engineering, including maximizing their scalability to ever larger computing architectures and supporting the technology transfer that will enable their uptake within the industrial environment.

HiDALGO - Global Challenges

HiDALGO enables the assessment of Global Challenges problem statements by enabling highly accurate simulations, data analytics, artificial intelligence and data visualisation, but also by providing knowledge on how to integrate the various workflows and the corresponding data.

MaX - Materials Design

MaX aims to disenthrall the EU leadership in materials modelling, simulations, discovery and design by creating an ecosystem of capabilities, software applications and data workflows and analysis on HPC-oriented material simulations, designed for present and future HPC architectures.

POP-2 - Performance optimization & productivity

POP-CoE assesses the performance of computing applications, identifying issues affecting code performance as well as the best ways to address them.

For more information on the European HPC Centres of Excellence visit us online:

www.focus-coe.eu

@FocusCoE

FocusCoE has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement N°

With the help of

FOCUS CoE