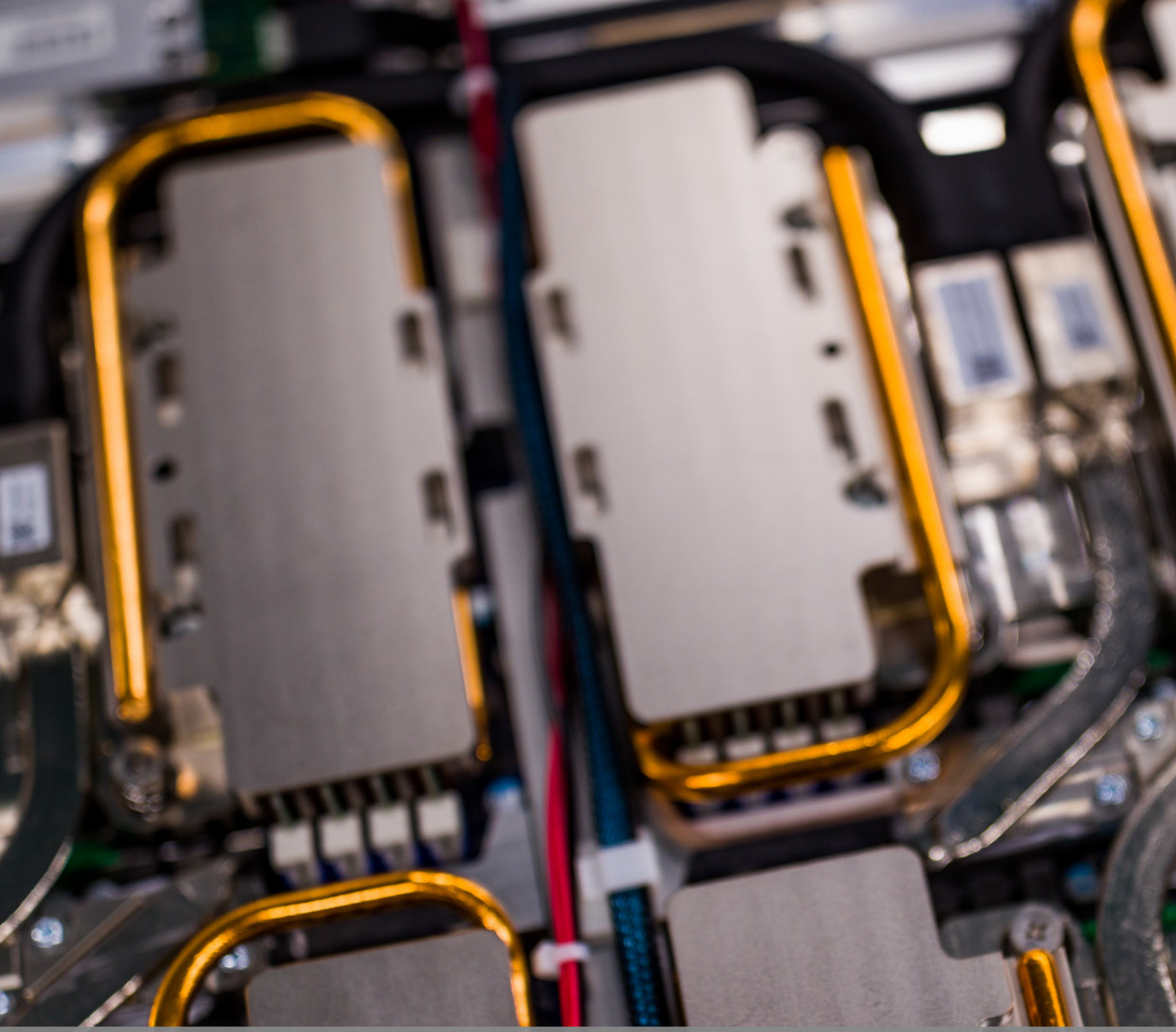


ASHPC23

Austrian-Slovenian HPC Meeting
Maribor: June 13 - 15, 2023

Austrian-Slovenian HPC Meeting 2023 – ASHPC23

Maribor, June 13 – June 15, 2023

https://ashpc.eu

Welcome to ASHPC23

A warm welcome to all participants of ASHPC23, which is being held from June 13–15, 2023 in the beautiful city of Maribor, Slovenia. It is the ninth meeting in this series, and the third organized jointly by the Austrian and Slovenian High-Performance Computing Community.

HPC is experiencing unprecedented prominence, as basically all scientific disciplines are increasingly relying on complex computational models and data analytics to drive their scientific discovery processes. This increase in the user community requiring access to HPC resources calls for significant investments to be made to support the scientific community in their endeavor to solve challenges faced by society. With the prominence of AI, even individuals that usually have no close contact with high-performance computation are suddenly aware of its importance and impact on our lives (even though only via a fraction of the discoveries that have been driven by HPC). This public awareness further amplifies expectations. At the same time, HPC's hunger for power and thus its impact on the environment is critically reflected upon, calling for further improvements in energy utilization and efficient use of HPC resources. All of this is happening in an economic climate where demand outpaces supply, both at the level of material as well as human resources.

In this dynamic environment, ASHPC23 offers a great opportunity for our community to get together, to share our achievements, the challenges we are facing and our visions for the future. It allows us to discuss how to progress together, to learn from each other's approaches and to build a better understanding of the future of HPC in our region.

ASHPC23 features a dense program of invited keynote speeches, presentations and posters. The keynotes, particularly, approach HPC from a range of different angles: Daniel Egger from IBM Zurich will provide an outlook on the integration of quantum computing into HPC environments. Abdulrahman Azab from the University of Oslo reviews challenges in Exascale computing, while Tomi Ilijas from Arctur d.o.o. takes a look at the innovation opportunities that HPC offers to small and medium enterprises, and Iztok Lebar Bajec from the University of Ljubljana presents approaches to machine translation relying on HPC. Last, but not least, Mario Kovac from the University of Zagreb presents the European Processor Initiative, highlighting the issue of digital sovereignty. This is complemented by numerous talks and poster presentations covering aspects of security, energy efficiency, quantum computing, updates from various HPC infrastructures as well as a plethora of applications utilizing HPC and each bringing their own challenges to an HPC environment. On behalf of the program and organizing committees I would like to take this opportunity to thank everybody involved in the preparation and organization of this fantastic event, specifically our local hosts and the entire team that helped in selecting and putting together an exciting program based on the contributions from the HPC community. We are very grateful to EVIDEN for their sponsorship.

ASHPC23 is organized by EuroCC Slovenia and EuroCC Austria – National Competence Centre for Supercomputing, Big Data and Artificial Intelligence, Austria, in cooperation with the Institute of Information Science in Maribor (IZUM), Slovenia, the Slovenian consortium for high-performance computing (SLING), the Vienna Scientific Cluster (VSC), Austria, and the Research Area Scientific Computing in Innsbruck, Austria.

We wish everybody a wonderful time in Maribor full of exciting and inspiring discussions!

The Programme Committee:

The Organizing Committee:

Schedule

Monday, June 12, 2023

Start	Session Moderator	Title
	EuroCC 2 Central European NCCs Workgroup Meeting (NCC/EuroCC staff only)	
15:00	Raman Ganguly (in-person) Markus Stöhr (online) Damjan Harisch (online)	EuroCC 2 NCCs Approaching Industry
16:00	COFFEE	
16:15	Pavel Tomšič (in-person) Claudia Blaas-Schenner (online)	EuroCC 2 NCCs Training
17:15	COFFEE	
17:30	Julia Wimmer (in-person) Tina Črnigoj Marc (online)	EuroCC 2 NCCs Communication and Dissemination
19:00	DINNER at IZUM	

Tuesday, June 13, 2023

Start		Title
08:30	ASHPC23 REGISTRATION	
09:30	Aleš Bošnjak & Andreas Rauber & Pavel Tomšič	WELCOME to ASHPC23
09:45	Daniel Egger	Integrating quantum computers in high-performance computing environments (KEYNOTE TALK)
10:30	Ratko Pilipović	Optimal surface quantum error codes for the NISQ era
10:45	Janez Povh	Evaluation of the D-Wave Hybrid solver on instances of Max-Cut problems
11:00	Mátyás Koniorczyk	On physical and physics-motivated QUBO–Ising heuristics: applications and perspectives
11:15	COFFEE	
11:45	Aleš Zamuda	Generative AI Using HPC in Text Summarization and Energy Plants
12:00	Lorenz Romaner	Density Functional Theory Simulations of Crystallographic Defects for Physical Metallurgy Applications
12:15	Lubomír Říha	A scalable way to cinematic-style visualization of unstructured CFD data
12:30	LUNCH	
14:00	Abdulrahman Azab	Functional and non-functional technology challenges in exascale supercomputing (KEYNOTE TALK)
14:45	Matjaž Pančur	Containers and MicroVMs: From Cloud to HPC?
15:00	Ioannis Vardas	Effects of Mapping Strategies on Average Duration and Throughput of Colocated HPC Applications
15:15	Barbara Krašovec	Security of HPC systems
15:30	COFFEE	
16:00	Caterina Caravita	Leonardo and the pre-exascale era
16:15	Ivan Vialov	Italian Pre-Exascale HPC System LEONARDO and its Austrian Share
16:30	Aleš Zemljak	EuroHPC Vega, 2 years & counting …
16:45	Jan Zabloudil	Status and developments at the Vienna Scientific Cluster
17:00	BREAK	
17:15	Ernst Haunschmid	Quantum Austria and Multi Site Computer Austria (MUSICA)
17:30	Martin Thaler	The UIBK HPC Ecosystem - LEO5
17:45	Michael Fink	Software Deployment at the UIBK LEO5 Cluster
18:00	Martin Belavić	Supek - Croatian Scientific and Educational Cloud HPC Resource
19:00	DINNER	

Wednesday, June 14, 2023

Start		Title
08:30	ASHPC23 REGISTRATION	
09:30	Tomi Ilijaš	HPC as an opportunity for innovative SMEs (KEYNOTE TALK)
10:15	Draško Tomić	Multi-head Additive Manufacturing with Optimal HPC Thermal Stabilization - AMOTS
10:30	Matej Borovinšek	Preliminary Analysis of Innovative Aerostructures Spectrum Powered by HPC
10:45	Jürgen Zechner	Development of High-Performance- and Real-Time-Computing Software for the Simulation of Electromagnetic Fields
11:00	COFFEE	
11:30	David Felhős	Enabling HPC-scale Engineering Simulations for SMEs
11:45	Raman Ganguly	EuroCC Supercomuting Accelerator – Grow with advanced technology
12:00	Tomas Karasek	NCC Czech Republic, entrance to the world of HPC
12:15	Marek Magryś	PLGrid Infrastructure in the Exascale Era
12:30	Gabrijela Zaharijas	SMASH – Marie Skłodowska-Curie COFUND project for postdoctoral researchers using HPC VEGA to develop machine learning applications for science and humanities
12:45	LUNCH	
14:00	Iztok Bajec	Lost in Vega(s): learning how to translate between English and Slovene on Vega (KEYNOTE TALK)
14:45	Soner Steiner	Time series in HPC: An exploratory study with distributed time series data for energy use cases
15:00	Alja Prah	Ensuring the quality of HPC software development
15:15	POSTER LIGHTNING TALKS	
15:30	POSTER SESSION + COFFEE	
	Pavel Tomšič	SCtrain: Supercomputing knowledge partnership
	Pavel Tomšič	The Slovenian National Competence Center for HPC – EuroCC@SLING
	Martin Belavić	Croatian Competence Centre for HPC - HR HPC CC
	Andreas Rauber	Research Data Infrastructure Landscape at TU Wien
	Blaž Vrhovšek	Data Storage and Analysis Challenges on HPC Systems in Genomic Medicine
	Katrin Muck	Python, Conda, and Virtual Environments on the Vienna Scientific Cluster
	Diego Dalla Costa	Computer Aided Engineering in Augmented Reality: Flow Visualizations for Hydro Power Applications
	Leon Bogdanović	Real-time human pose estimation using YOLOv7 on HPC
	Stefano Elefante	Cryo-EM Software Packages: A Sys-admins Point of View
	Marko Jukič	Inverse Molecular Docking Approaches on HPC Systems

	Andraž Filipčič	Open boundary molecular dynamics with GPU parallel computation
	Davor Sluga	Rand_gpu – A C/C++ library for generating pseudo-random numbers on the GPU
	Aleš Zamuda	Computing in DAPHNE: Integrated Data Analysis Pipelines for Large-Scale Data Management, HPC and Machine Learning
	Lubomír Říha	SPACE - Scalable Parallel Astrophysical Codes for Exascale - Center of Excelence
	James McKevitt	Accelerating Astrophysical Code Using CUDA Fortran
	Ivona Vasileska	SIMPIC: A simplified particle-in-cell plasma physics code optimized for multiple GPUs
16:45	Francesco Carlo Mantegazza	GPU-accelerated matrix-free solvers for the efficient solution of cardiac electrophysiology in `life`^{x}
17:00	Sascha Hunold	Massively Scaling Molecular Screening Workloads on EuroHPC Supercomputers
17:15	Vid Ravnik	Benchmarking the CmDock Molecular Docking Software on HPC Systems
17:30	BREAK	
17:45	Constanze Roedig	OpenScienceLabs for HPC on the European Open Science Cloud
18:00	Marko Ferme	Slovenian open access infrastructure – storing and sharing big data
18:15	Dejan Lesjak	Federated storage status and plans in SLING centres
19:30	DINNER	

Thursday, June 15, 2023

Start		Title
08:30	ASHPC23 REGISTRATION	
09:30	Mario Kovač	European Processor Initiative (KEYNOTE TALK)
10:15	Siegfried Höfinger	Energy Efficiency in HPC - European, Global and National Perspective
10:30	Domen Tabernik	A GPU Cluster for Small Teams to Efficiently Conduct Deep Learning Research and Development
10:45	Alois Schlögl	Running Windows-applications on a Linux HPC cluster using WINE
11:00	COFFEE	
11:30	Sascha Hunold	MPI is Good, Control is Better: Checking Performance Guidelines of Collectives
11:45	Philippe Swartvagher	Rank Reordering within MPI Communicators to Exploit Deep Hierarchal Architectures of Supercomputers
12:00	Alena Melic	Joint propagation characteristics of acoustic and electromagnetic waves in shallow ocean
12:15	CLOSING of ASHPC23	
12:30	LUNCH	
14:00	END of ASHPC23	

Contents

Keynote Talk:

Integrating quantum computers in high-performance computing environments

Daniel J. Egger

IBM Quantum, IBM Research Europe – Zurich, Switzerland

In the past years quantum computers have made significant progress in their performance as measured by their scale, speed, and quality. For instance, in 2022 IBM Quantum announced a 433-qubit quantum processor and a factor of 10 increase in circuit execution speed. Furthermore, coherence times and gate fidelities have also improved as exemplified by median T_1 times which now exceed 0.3 ms on new devices. While quantum computers are expected to solve hard computational tasks, with which classical computers struggle, they are not going to replace them. Instead, they will work in tandem in an integrated high-performance computing framework. Here, we will discuss a framework that integrates classical and quantum computers using an open-source middleware built on Qiskit; a leading quantum computing software kit. We will see how certain challenging computational tasks can be sent to several quantum computers and how the high-performance classical compute environment can support the execution of quantum circuits and the processing of the results.

For instance, quantum computers may benefit applications such as machine learning, optimization, the simulation of natural sciences and finance. Furthermore, converting application-level quantum circuits to hardware executable circuits features many optimization steps that can effectively be carried out in a high-performance classical computing environment. In summary, the close integration of the quantum and classical information processing paradigms is expected to significantly increase computational performance.

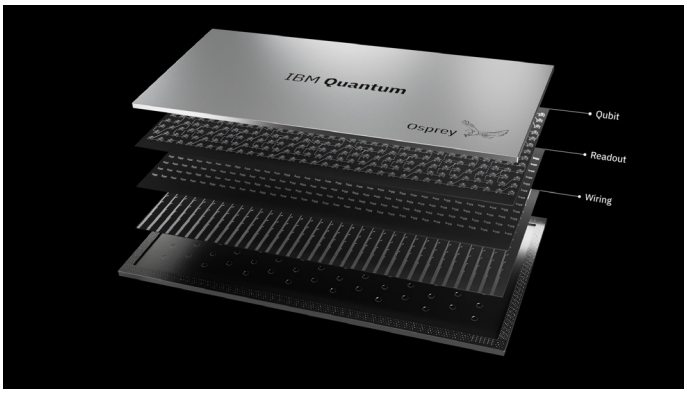


Fig. 1: Osprey processor with 433 Qubits.

Optimal surface quantum error codes for the NISQ era

Samira Sayedsalehi[a], *Ratko Pilipović*[b], Alberto A. Del Barrio[c], and Nader Bagherzadeh[a]

[a] *Department of Electrical Engineering and Computer Science, University of California, USA*
[b] *Faculty of Computer and Information Science, University of Ljubljana, Slovenia*
[c] *Department of Computer Architecture and Automation, Complutense University of Madrid, Spain*

Quantum computing possesses the capacity to bring about a significant transformation in diverse domains, including cryptography, machine learning and chemistry. This technology draws upon the principles of quantum mechanics enabling the development of robust computational systems that outperform classical computers in carrying out specific calculations. Nonetheless, QC encounters certain physical constraints, such as qubit's short coherence time and sensitivity to noise, which hinder it from realizing its immense capabilities.

Recently, Noisy Intermediate-Scale Quantum (NISQ) computers have emerged as a stepping stone towards extensive quantum computing. However, quantum machines face notable obstacles due to their vulnerability to various types of noise that affect computations. To address these concerns, researchers have developed quantum error correction techniques that mitigate the effects of errors and improve the reliability of computations on NISQ devices.

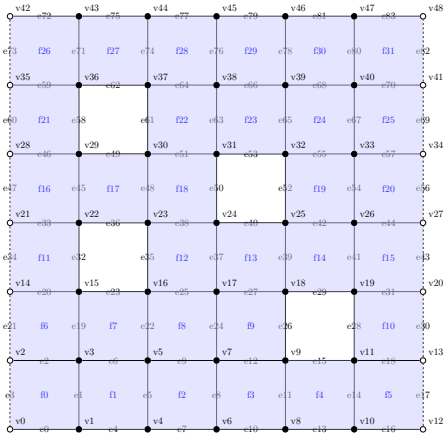


Fig. 1: Obtained surface code using Simulated Annealing.

Effective quantum error codes, including surface and topological codes [2], have been developed to address errors in NISQ computers. These codes rely on interactions between adjacent physical qubits for logical encoding, offering efficient means of error mitigation with a higher error threshold. However, drawbacks like low encoding density and the difficulties in obtaining a universal gate set limit to the practical use of these codes. Thus, researchers continue to explore quantum error correction strategies specifically for NISQ devices - including surface code optimization as an important area of investigation.

Our study seeks to improve the performance of NISQ devices by implementing a surface QEC code. This will help mitigate qubit errors and facilitate the development of large-scale quantum computers. Using Simulated Annealing [3], we aim to optimize both the code rate (the ratio between encoded logical and physical qubits with minimal redundancy) and the error rate. Fig. 1 shows one of the surface codes we have developed. The edges represent physical qubits, while the vertices and faces represent X and Z stabilizers, which help to detect bit and phase flips of physical qubits, respectively. The depicted surface code encodes five logical qubits with a logical error rate of 0.0564.

Acknowledgements: This research was supported by Slovenian Research Agency under Grant BI-US/22-24-114 (Bilateral Collaboration Project).

References

[1] M. H. Devoret and R. J. Schoelkopf, "Superconducting circuits for Quantum Information: An outlook," Science, vol. 339, no. 6124, pp. 1169–1174, 2013.

[2] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, "Surface codes: Towards practical large-scale quantum computation," Physical Review A, vol. 86, no. 3, p. 032324, 2012.

[3] K. A. Dowsland and J. Thompson, "Simulated annealing," Handbook of natural computing, pp. 1623–1655, 2012.

Evaluation of the D-Wave Hybrid solver on instances of Max-Cut problems

Janez Povh[a,f], **Vid Eržen**[c], **Jure Brence**[c,d], **Timotej Hrga**[a,c], **Blaž Škrlj**[b], **and Jaka Vodeb**[c,e]

[a] *Faculty of Mechanical Engineering, University of Ljubljana, Slovenia*
[b] *Institute of Mathematics, Physics and Mechanics – IMFM, Slovenia*
[c] *Faculty of Mathematics and Physics, University of Ljubljana, Slovenia*
[d] *Department for Knowledge Technologies, Jožef Stefan Institute, Slovenia*
[e] *Department of Complex Matter, Jožef Stefan Institute, Slovenia*
[f] *Rudolfovo Institute, Slovenia*

We are well into the second quantum revolution, where the knowledge gained from the advent of quantum theory is now being used to build quantum devices. One of the biggest potential benefits of such an undertaking is a new and better way of computing, which has been termed "quantum computing". There are two main avenues being developed currently; digital quantum computing, which provides universality in terms of which algorithms can be deployed on such devices; and quantum annealing (QA), which has only one specific type of algorithm available, but is therefore much more developed in terms of device size and performance. If we only take into account the number of quantum bits or qubits available on the two types of devices, which is by no means a complete performance metric, the biggest device using digital quantum computing is IBM's Osprey with 433 qubits, while the D-Wave Advantage hosts over 5000 qubits. The main driver behind building quantum computers is achieving quantum supremacy, where a quantum computer employing a quantum algorithm significantly outperforms any classical algorithms. This was done by Google in 2019 for a random circuit sampling problem, which holds no practical value. The more realistic goal is quantum advantage, where a quantum computer outperforms all known classical methods in solving a given problem.

In this work, we focus on solving the Max-Cut optimization problem using the D-Wave Advantage quantum annealer and comparing it to state-of-the-art simulated annealing samplers [1] as well as an exact Branch and Bound solver BiqBin [2]. According to a recent review, quantum annealers have so far never been able to demonstrate quantum advantage in practical use cases, which provides motivation for this work. Here we perform a benchmark test using a standard library of Max-Cut instances [3] in order to assess the performance of the Hybrid solver, provided by the company D-Wave, as well as their quantum processing unit (QPU), and compare them to the performance of the BiqBin solver. We also compare the Hybrid solver to the Simulated Annealing Bifurcation Machine developed recently by Toshiba. We find that the D-Wave Hybrid solver improves the best-known solution for several benchmark instances and we provide a detailed analysis of the computing time required in order to solve instances of Max-Cut using D-Wave's Hybrid solver.

References

[1] Goto, H., Tatsumura, K. and Dixon, A.R., Science advances **5**, 2372 (2019).

[2] Gusmeroli, N., Hrga, T., Lužar, B., Povh, J., Siebenhofer, M., and Wiegele, A., ACM Transactions on Mathematical Software (TOMS) **48**, 1 (2022).

[3] Chu, P.C. and Beasley, J. E., Journal of heuristics **4**, 63 (1998).

On physical and physics-motivated QUBO–Ising heuristics: applications and perspectives

Mátyás Koniorczyk[a], **Péter Naszvadi**[a], **Krzysztof Domino**[b], **and Bartłomiej Gardas**[b]

[a] *Wigner Research Centre for Physics, Hungary*
[b] *Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, Poland*

Quadratic unconstrained binary optimization (QUBO) problems, including MAX-CUT as a special case, are important hard problems of mathematical optimization. Recently they have attracted much attention because they are the very problems that quantum annealers can address directly: QUBO problems are equivalent to the Ising model, a central and extensively studied problem in physics.

Even though the usefulness and efficiency of quantum annealers are the subjects of many debates, their development is steadily going on. And the equivalence of QUBO and Ising models bridges between two different, well-developed approaches and insights to the problem: that of mathematicians and physicists. The former is based mainly on semidefinite relaxations or advanced cuts, whereas the latter offers a physical insight and includes DMRG/tensor network algorithms, classical and quantum annealing, or simulated bifurcation.

In the present talk we introduce the first railway application of quantum annealers that had been proposed partly by some of us. The railway dispatching / conflict management problem is addressed, which is equivalent to a job-shop scheduling with blocking and no-wait constraints. It is a computationally hard problem and the calculations have to conclude in a very limited time. We present our proof of concept demonstration, described also in [1] based purely on quantum annealing, but calculated alternatively also on classical computers with tensor network algorithms [2] and with a brute-force method on GPUs. We also present our ongoing work where we show that hybrid quantum-classical heuristics are valid and practically useful options in dealing with such a problem. We describe modeling strategies, benefits and limitations of quantum annealing approaches.

To illustrate one of the limitations of quantum annealers in detail, we discuss deeper the problem that like most heuristics they do not necessarily find the optimum. We present a probabilistic discriminator based on ideas from statistical physics [3] that can indicate whether the optimum was found. We also consider the standard (Fortet) linearization of a QUBO, and, as a mathematical approach, discuss the possibility of using MILP duality for the purpose of deciding whether an optimum was found. The latter approach is a general one, not specific to quantum annealing.

Finally we discuss the perspectives of hybrid (HPC+quantum) solvers, and the potential benefits from using physics-motivated algorithms for solving QUBOs in general.

References

[1] Domino, K., Koniorczyk, M., Krawiec K., Jałowiecki K., Deffner S., Gardas B., Entropy **25**, 191 (2023).

[2] Rams M., Mosheni M., Eppens D., Jałowiecki K., Gardas B., Phys. Rev. E **104**, 025308 (2021).

[3] Domino K., Koniorczyk M., Puchała Z., Quantum Information Processing **21**, 288 (2022).

Generative AI Using HPC in Text Summarization and Energy Plants

Aleš Zamuda

Faculty of Electrical Engineering and Computer Science, University of Maribor, Slovenia

This contribution presents some specific workloads for Generative AI modalities and discusses their HPC deployment — the Summarizer for generating text summaries [1], EcoMod for generating procedural 3D images of natural ecosystems with trees [2], and load balancing of energy power plants (PSADE@NPdynєjDE) for generating schedules using data analytics and machine learning [3]; and then discusses their MPI integration and HPC deployment. Namely, Generative AI for modalities such as text generation using Transformers (like ChatGPT), image generation using Stable Diffusion (like Midjouney and DALL-E), and video speech generation (like Synthesia) has provided recent interesting applications served by HPC deployments. Therefore, two models for Generative AI, from Summarizer and TPP-PSADE@NPdynєjDE, extended to support HPC deployment using MPI, are described below and some results are presented.

In text summarization, once the text model has been built, different lengths of summaries can be created with Summarizer using CaBiSDETS algorithm in parallel [1]. But moreover, before the generation of summaries commences and just before a complete parameterization of the summary generation process, the input text can be preprocessed in parallel by computing the cosine sentence similarity pairs in parallel using MPI in an integrated pipeline. During summary optimization, the fitness function evaluations can be run in parallel.

Some demonstrative results for Summarizer computational architecture are provided in Fig. 1 based on a recent lecturing material on Optimization Algorithms and Autonomous Systems at Las Palmas de Gran Canaria from March 2023. For this experiment, the configuration for SLURM was as follows: `srun --ntasks-per-node=16 --mpi=pmix ./summarizer.sif ./summarizer`; and for Summarizer, the parameters were: `--GMAX 1000 --NP 319 --summarylength 500 --epsilonLengthSummary 20`. The SLURM `--nodes` parameter was demonstratively increased from 1 to 5 and the following scaling and timings were observed as seen on Fig. 1.

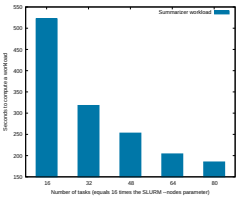


Fig. 1: Summarizer scaling.

When generating energy plans, like load balancing schedules for energy power plants, data analytics and machine learning algorithm PSADE@NPdynєjDE [3] can be used. Here, a part of computation for the whole problem is split and offloaded to an offline pre-computation phase. In this initial phase, an HPC can be used to generate a lookup matrix, included as a surrogate matrix during the next, online optimization phase. An example set of long runs (1 million fitness evaluations) to generate a surrogate matrix with 0.01 MW granularity step is demonstrated in Fig. 2, for power demands from 110 MW to 975 MW and 0.01 MW apart from 10^6 simulations for each of 86500 power demand scenarios.

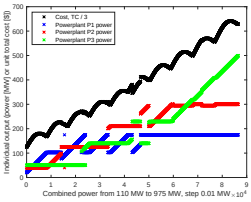


Fig. 2: Power plant plans.

In future work, additional integrated pipelines deployments using DAPHNE runtime would be interesting, e.g. using micro-benchmarks for multi-document text-summarization [1], 3D forest scenery [2], energy scheduling [3], and trajectory optimization [4].

Acknowledgement: This work is supported by ARRS programme P2-0041 and EU project no. 957407.

References

[1] Zamuda, A., and Lloret, E., Journal of Computational Science **42**, 101101 (2020).

[2] Zamuda, A., and Brest, J., Information Sciences **278**, 1–21 (2014).

[3] Glotić, A., and Zamuda, A., Applied Energy **141**, 42–56 (2015).

[4] Zamuda, A., and Hernández Sosa, J. D., Expert Systems with Applications **119**, 155–170 (2019).

Density Functional Theory Simulations of Crystallographic Defects for Physical Metallurgy Applications

Lorenz Romaner **and David Holec**

Department of Materials Science, Montanuniversität Leoben, Austria

Density functional theory (DFT) calculations are a powerful tool to explore the properties of crystal phases and crystallographic defects which determine functional and mechanical properties of materials. With the ongoing increase in computational power of HPC infrastructures, more complex systems can be treated such as random alloys or extended defects including in particular dislocations or grain boundaries. However, due to the unfavorable $O(N^3)$ scaling, where N denotes the amount of electrons in the system, direct simulations are still often extremely costly and alternative approaches are hence required to obtain computationally tractable schemes. Of great interest is the possibility to combine DFT with other methods such as molecular mechanics, mean-field methods based on thermokinetic principles or with machine learning.

In this contribution we will present materials simulations addressing solid state segregation which determines the chemical composition of grain boundaries in metallic alloys as shown in Fig. 1. Intergranular embrittlement phenomena are of strong relevance for many technological materials including steels, nickel-based alloys, coinage metals or refractory metals. The decisive process is grain boundary segregation where solutes diffuse and enrich at grain boundaries thereby modifying their cohesive properties which can favor or impede crack propagation.

We present a multi-scale simulation framework for grain boundary segregation employing atomistic, thermokinetic and data-driven computational methods and show how chemistry and propensity against crack propagation can be modeled from basic knowledge of composition and processing parameters [1,2]. Validation examples with several experimental methods including atom probe tomography, high resolution transmission electron microscopy and micromechanical testing are presented. Furthermore, we discuss the potential of statistical learning methods to replace density functional theory simulations in future, in particular to address advanced phenomena such as concentration dependence and co-segregation.

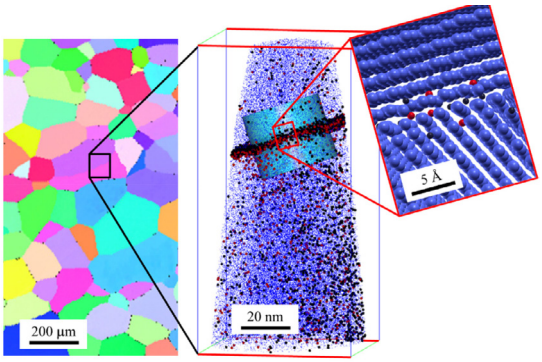


Fig. 1: Illustration of a polycrystalline microstructure (right), zoom into the grain boundary as obtained from atom probe tomography (middle) and zoom into the atomistic structure as investigated by DFT investigations (right).

Furthermore, a direct combination of DFT with molecular statics in a so-called QM/MM method is presented. By doing so, one is able to treat large systems, such as extended defects, in matrix materials for which interatomic potentials exist, with a locally enhanced chemistry. The latter is the limiting factor of standard MM, but can be consistently treated using DFT. On the other hand, extended defects such as grain boundaries or dislocations are computationally extremely expensive (in fact, apart from a few high-symmetry cases, prohibitive) at the DFT level. We will demonstrate its working implementation on an example of segregation to γ/γ and γ/α_2 grain boundaries in a TiAl intermetallic alloy.

References

[1] Leitner, K., Scheiber, D., Severin, J., Sophie, P., Clemens, H., Povoden-Karadeniz, E., Romaner, L. Materials and Design **142**, 36 (2018).

[2] Scheiber, D., Povoden-Karadeniz, E., Kozeschnik, E., Romaner, L. Acta Materialia **224**, 117482 (2022).

A scalable way to cinematic-style visualization of unstructured CFD data

Markéta Hrabánková, Ondřej Meca, Tomáš Brzobohatý, *Lubomír Říha*, Milan Jaroš, and Petr Strakoš

IT4Innovations, VŠB - Technical University of Ostrava, Czech Republic

This work presents a novel high-performance post-processing workflow for cinematic visualization of high-fidelity computational fluid dynamics (CFD) datasets using the volume path-tracing technique. The workflow consists of ***(1)*** conversion of simulation results from an unstructured mesh by our library MESIO[1] to a regular grid and storage of data in the sparse volumetric database in the OpenVDB format [2] and ***(2)*** interactive or offline remote rendering of the OpenVDB database by the professional path-tracer Blender Cycles. The workflow can convert hundreds of time steps on a mesh with 1 billion cells (tens of TB of data) to a regular grid with 11 billion voxels in just a few minutes using thousands of CPU cores.

The key achievements of this work are ***(i)*** reducing the post-processing time since using standard tools for the visualization of scientific data takes several hours,***(ii)*** the quality of the visualization, and ***(iii)*** interactive exploration of simulation results. Our primary focus was to process CFD simulations with many time steps; therefore, we could parallelize even sequential action. The highlight of the optimization we have achieved per workflow steps is as follows: ***(a)*** Loading of unstructured mesh with simulation results and voxelization of unstructured mesh into a regular grid with filtering. We outperform the leading open-source software for post-processing and visualization Paraview by 6.3x for meshes with a low growth rate. We also achieve near-linear scalability for meshes with high growth rates, for which ParaView does not scale at all. ***(b)*** Store converted data into OpenVBD. We propose overcoming this library's sequential nature by storing hundreds of time steps in parallel. ***(c)*** Import OpenVDB into Blender. We proposed and implemented parallel conversion of OpenVDB data into Blender rendering structures. The parallelization is done over time steps, and it is linear. ***(d)*** Interactive or offline rendering of the results. We proposed a rendering method based on CUDA unified memory that can load hundreds or thousands of time steps into a single multi-GPU server and use the combined performance of all GPUs to render a single frame. We have also extended this approach to multi-node multi-GPU rendering to deliver interactive rendering for high-resolution volumetric data.

Therefore we can render 4.4GB of NanoVDB volumetric data per time step at a rate of over five frames per second. For all performance measurements we use the Karolina cluster at IT4Innovations, the CPU partition consists of 720 compute nodes (2xAMD 7H12, 64 cores, 256 GB), and the GPU partition consists of 72 accelerated nodes (2xAMD 7763, 64 cores, 1024 GB; 8xNVIDIA A100, 40 GB).

To sum up, we can load and process 1024 time steps of the unstructured one billion cell CFD simulation (approximately 34TB of data) for fast evaluation quality (22 million voxels) in 5 minutes (0.28 sec per time step) and for the final production rendering quality (11,4 billion voxels) in 37 minutes (2.16 sec per time step) on 64 compute nodes.

Fig. 1: 1B cell unstruct. mesh, velocity mag. on vortex structures

References

[1] Meca, O., Riha, L., Jansik, B., Brzobohaty, T., Toward highly parallel loading of unstructured meshes, Advances in Engineering Software, vol. 166, p. 103100, (2022).

[2 Museth, K. VDB: High-Resolution Sparse Volumes with Dynamic Topology. ACM Trans. Graph. 32, 3, Article 27, (2013)

Keynote Talk:

Functional and non-functional technology challenges in exascale supercomputing

Abdulrahman Azab

Scientific Computing Services, IT Services, University of Oslo, Norway
Management Office, Nordic e-Infrastructure Collaboration (NeIC), Norway

Data is becoming the main driving force behind industrial, research, and social progress. By 2025, an estimated 463 exabytes of data will be created daily worldwide [1]. The challenge of analysing huge amount of data is forcing significant advancements in computer processing capacity. HPC is the discipline in computer science in which supercomputers are used to solve complex scientific problems. Much of the current research and development in HPC is focused on exascale computing. Exascale computing means working towards a system with a floating point performance of at least 1 ExaFLOPS (i.e., 10^{18} or a million million million floating point operations per second (FLOPS)). The first petascale (10^{15} FLOPS) computer entered operation in 2008 [2]. In June 2020 the Japanese supercomputer Fugaku achieved 1.42 ExaFLOPS in HPL-AI benchmark. In June 2022, the world's first public exascale computer, Frontier, became the world's fastest supercomputer.

Exascale/pre-exascale computing should "not" be thought of just as a huge number of floating point units, so that HPC centers can compete on who has the largest supercomputer in "size". If an exascale "sized" supercomputer is not able to run an exascale "application", does this actually make it an exascale supercomputer? Or it can be simply described as a set of smaller supercomputers located in one room?

Fig. 1: Pre-exascale supercomputer LUMI (© Fade Creative)

There are several technical challenges in the way to actually reach exascale computing. One is how to actually develop exascale applications i.e., with billion-way parallelism – 1 billion floating point units performing 1 billion calculations per second. Another challenge is the "memory wall" challenge. If exascale/pre-exascale systems are supposed to be the "fastest", how can we manage the time and energy required to move data from memory into the compute units, and from the compute units out to storage, not to be larger than the time and energy required to perform a floating-point operation on that data. In addition, there are several non-functional challenges like porting and optimization of scientific software. Would Linux containers solve the problem? What about MPI and GPU applications?

The presentation will address some functional and non-functional challenges, including a few system-specific ones, and how they are currently handled in Frontier and the EuroHPC petascale systems.

References

[1] Desjardins, Jeff. "How much data is generated each day?" World Economic Forum. April 17, 2019. https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/

[2] National Research Council (U.S.) (2008). The potential impact of high-end capability computing on four illustrative fields of science and engineering. The National Academies. p. 11. ISBN 978-0-309-12485-0.

Containers and MicroVMs: From Cloud to HPC?

Matjaž Pančur

Faculty of Computer and Information Science, University of Ljubljana, Slovenia

Containers have become a core technology in the world of HPC, providing an easy to use, flexible and powerful way to package and run our applications. As HPC environments are shifting from CPU only towards using different types of compute accelerators, like GPUs, FPGAs, etc.. Generally speaking, as HPC clusters continue to grow in complexity and scale, containers are likely to play an increasingly important role in enabling efficient and flexible computing.

While we can easily say that using containers in HPC systems is now mainstream, you can not say that for virtual machines. Virtual machines do have an important role in HPC cluster *support services*, but they are not usually used on fast compute partitions of HPC clusters or when running HPC applications/computations.

Traditional virtual machines provide a great deal of flexibility and isolation, but they are also resource-intensive and could be slow to start up and shut down. The overhead of a hypervisor depends on several factors, such as the hardware configuration, the workload running on the virtual machines, and the type of hypervisor being used. For hardware-assisted virtualization technologies like Intel VT-x and AMD-V, the overhead of a traditional hypervisor can be on the order of 5-10%.

MicroVMs, or *micro virtual machines*, are lightweight virtual machines, that usually run a single application or process. Overall, microVMs are a relatively new technology that is rapidly gaining popularity in the cloud computing and virtualization space. They are designed to be much smaller and faster than traditional VMs. Instead of running a full operating system, a microVM typically runs a lightweight kernel and only the bare minimum of software needed to support a single application. While microVMs hypervisors offer several advantages over traditional hypervisors in terms of memory efficiency, scalability, and most importantly, security, there are also some challenges associated with their use in an HPC environment.

The best known example of a microVM hypervisor is Firecracker[1], which was rewritten from scratch in Rust by AWS for their FaaS (Function-as-a-Service) platform. Firecracker is designed to boot in less than 125 milliseconds and use less than five megabytes of memory. This makes it well suited for running serverless workloads, which typically require fast and efficient scaling to meet demand. In addition to its use in the AWS FaaS platform, Firecracker is also used by other companies and organisations for a variety of use cases, including running containers. But FaaS as a primary use case also comes with some inherent limitations that make it less suitable for typical cloud environments. Firecracker, for example, is optimised for FaaS, so it's not really suited for running HPC applications. It only supports 6 emulated devices and not even a PCI device model, so it can not passthrough network and other devices. This means that no RDMA is possible with a HW NIC inside the VM or using accelerators like GPUs. PCI Passthrough support is also not planned as this is not so usable for FaaS.

This functional gap was filled by Cloud Hypervisor [2]. This is an open source hypervisor backed by Intel, Microsoft, AMD, ARM, Tencent, Alibaba and other industry heavyweights. It is much better suited for running HPC workloads and was originally developed for cloud-native workloads. It uses the same base project [3] as Firecracker.

References

[1] Firecracker, https://github.com/firecracker-microvm/firecracker, https://firecracker-microvm.github.io

[2] Cloud Hypervisor, https://github.com/cloud-hypervisor/cloud-hypervisor, https://cloudhypervisor.org

[3] rust-vmm, https://github.com/rust-vmm

Effects of Mapping Strategies on Average Duration and Throughput of Colocated HPC Applications

***Ioannis Vardas*, Sascha Hunold, Philippe Swartvagher, and Jesper Larsson Träff**

Faculty of Informatics, TU Wien, Austria

Due to resource contention like memory bandwidth, as well as limited scalability, applications are not always able to efficiently utilize the high core count of modern, deeply hierarchical HPC systems, like VSC-5. In such cases, colocating multiple applications to share compute nodes can increase the job system throughput. However, colocation can adversely impact individual application performance due to resource contention. Past research suggests that good mappings that take the system's topology into account can improve the performance of parallel applications that do not share resources [1].

We implement eight application-oblivious and topology-aware mapping strategies that produce process-to-core mappings and support colocation. We evaluate them using eight MPI applications on VSC-5. Our results show that colocation combined with specific mappings outperforms the exclusive allocation that is widely adopted by HPC clusters, both in average application duration and makespan.

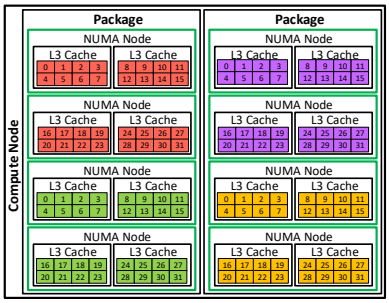


Fig. 1: Four applications sharing a node with `bbb` mapping.

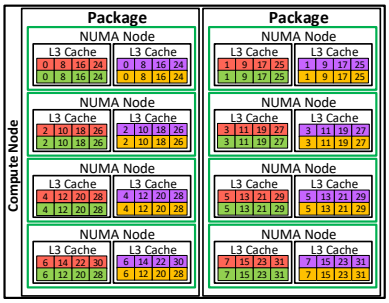


Fig. 2: Four applications sharing a node with `bcc` mapping.

Our mapping strategies consider three levels of resource hierarchy, the node, the package or socket and the NUMA node. The processes at each level are allocated either in a block or a cyclic way. Block allocates all cores within a resource before allocating from the next resource in the same level of hierarchy whereas cyclic performs the opposite. We denote the allocation (block or cyclic), for every level of the resource hierarchy with a single character, b or c. Figs. 1 and 2 show two mappings, `bbb` and `bcc`, of four colocated applications in a single VSC-5 node. We use different colors to depict the processes that belong to different applications. The `bbb` mapping (Fig. 1) increases the communication locality, whereas `bcc` (Fig. 2) increases resource utilization. Mappings also affect the number of shared resources among different applications.

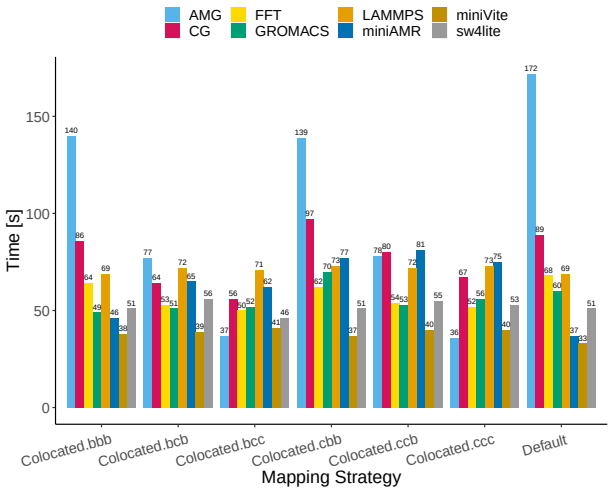


Fig. 3: Performance of eight colocated applications with six different mappings on eight VSC-5 nodes.

Fig. 3 shows the performance of eight colocated applications, each with 128 processes, that share eight nodes in VSC-5 and use six different mappings. We compare them against the default policy that allocates one node to each application where it runs in isolation. We notice that miniAMR and miniVite benefit more from mappings that increase communication locality, such as `bbb` or the default. Overall, `colocated.bcc` outperforms the default by 1.5× and 2.4× in terms of average duration and makespan respectively.

References

[1] von Kirchbach, K., Lehr, M., Hunold, S., Schulz, C., Träff, J.L.: Efficient Process-to-Node Mapping Algorithms for Stencil Computations. In: Proceedings of IEEE CLUSTER, 1 (2020)

Security of HPC systems

Barbara Krasovec

Centre for Network Infrastructure, Jozef Stefan Institute, Slovenia
EGI Computer Security and Incident Response Team

Security concerns in HPC systems differ from those in other sectors such as finance, healthcare, or defense. But a compromise or an attack can impact a variety of assets: computer systems, network systems, data storage, personnel, software repositories, services, large data repositories, visualization environments, and analysis environments. While HPC systems are not among the most attractive targets, they are particularly vulnerable because they are not well protected, because security practices are inefficiently implemented and because they allow multiple users to share data, storage,and compute resources in a complex environment. The security of high-performance computers is currently neglected due to the pursuit of performance. However, we have to remember a major security incident in 2020, impacting dozens of HPC centers due to a combination of stolen credentials, an unpatched kernel vulnerability and lack of security monitoring, that caused major damage and knocked several large HPC centres offline, forcing them to reinstall.

Technical concerns for HPC are further complicated by the complex and ever-changing threat landscape, the diverse and complex hardware, software and applications and an increasing amount of data being processed very quickly [3]. The private and public sectors have joined forces in a working group called High-Performance Computing Security, which is managed by the National Institute of Standards and Technology and the National Science Foundation, and works on defining the HPC security architecture [4].

The WLCG community is one of the best represented in the field of high-performance and distributed computing security. Its members are already bound by a set of security policies [1] and must contact the security team if there is a security incident that could impact users, services, or operations. The same goes for members of the EGI community, where the EGI CSIRT is coordinating incident response, where incidents are usually large-scale and involve multiple stakeholders. Security is also part of the EuroHPC JU Multi-Annual Strategic Plan for the period 2023 to 2027.

Currently, multiple security frameworks and standards define different policies, guidelines and best practices to manage an organisation's security risks, such as CIS controls, ISO standards, NIST cybersecurity framework [2], SOC 2 and others.

Yet, essential security requirements in the HPC are complex, including host hardening, remote central logging, automated system deployment and configuration management, restricted user access and security monitoring. Workloads and HPC architectures are evolving quickly, especially on multi-purpose systems where user activity is unpredictable and configurations change quickly. The question therefore becomes: what to monitor and how to detect anomalies? The system complexity makes it easy to hide malicious activities. We know of attacks in HPC that have gone undetected for months. How to adapt to prevent or detect such incidents will be the topic of this talk.

References

[1] WLCG Security, https://wlcg.web.cern.chusing-wlcgcomputer-security

[2] NIST security Framework, https://www.nist.gov/cyberframework

[3] R. Bulusu, P. Jain, P. Pawar, M. Afzal and S. Wandhekar, "Addressing security aspects for HPC infrastructure," 2018 International Conference on Information and Computer Technologies (ICICT), DeKalb, IL, USA, 2018, pp. 27-30, doi: 10.1109/INFOCT.2018.8356835.

[4] High-Performance Computing (HPC) Security, draft NIST SP 800-223. https://www.nist.gov/news-events/news/2023/02/high-performance-computing-hpc-security-draft-nist-sp-800-223-available

Leonardo and the pre-exascale era

Caterina Caravita

Cineca - HPC Department, Italy

Leonardo is one of three pre-exascale supercomputers announced by EuroHPC Joint Undertaking, it is hosted and managed by Cineca, in Bologna (Italy), and it is currently ranked at the 4th position in the Top500 list.

It counts 4992 computing nodes and more than 100 PB of storage, and it is expected to reach a computational performance around 250 PFlop/s. It provides a general purpose Data Centric module and a Booster module designed to satisfy the most computational-demanding requirements with nodes equipped with GPUs. Its pre-exascale performances will be able to contribute, for instance, at the mitigation and management of risks due to extreme situations, natural events, earthquakes, tsunamis, volcanic events, flash floods, for the fight against pandemic and epidemic situations.

The supercomputer is devoted to the academic and industrial research across Europe, and the resources can be requested through national and international calls for proposals.

I will present the main features and potentialities of the European supercomputer Leonardo, with an overview on how to get resources and how to work on the cluster.

Fig. 1: The HPC-system Leonardo from the outside ...

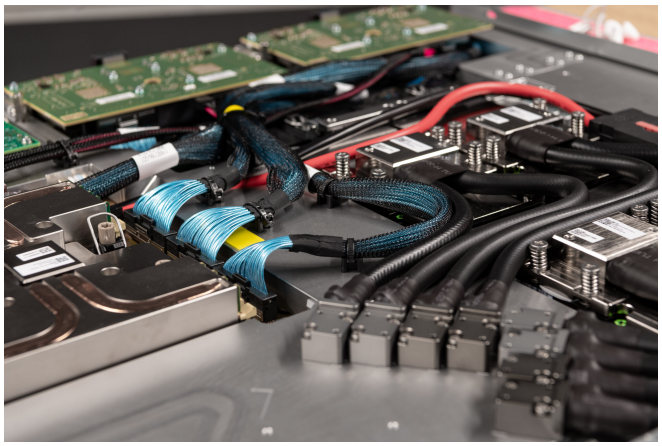


Fig. 2: ... and from the inside. [1]

References

[1] https://leonardo-supercomputer.cineca.eu/

Italian Pre-Exascale HPC System LEONARDO and Its Austrian Share

Ivan Vialov **and Claudia Blaas-Schenner**

VSC Research Center, TU Wien, Austria

The Italian pre-exascale HPC system LEONARDO [1] begins the regular operation of its "Booster" module, which draws its enormous computing power from NVIDIA Ampere GPUs. Currently it is the fourth most powerful supercomputer in the world and second most powerful in Europe. Booster encompasses 3456 compute nodes each equipped with 4 QC-selected A100 GPUs and a single 32-cores Intel Ice Lake CPU to drive those. The GPUs are interconnected with 200 GB/s bidirectional NVLink. All this results in 238.70 PFlop/s according to the Top500 list.

Austria gets access to this supercomputer through its share in the consortium, which also includes Greece, Hungary, Slovenia and Slovakia, that invested in the machine. This Austrian share amounts to some 500k GPU-hours, which is comparable to the resources of the current VSC-5 GPU installation. In order to distribute this compute time VSC has put forward the "AURELEO: Austrian Users at LEONARDO" call [2].

CINECA (the hosting entity of LEONARDO) formulated fairly strict requirements to the granularity of the allocation and scalability of the code being run on Booster. VSC and the Steering Committee have selected 11 out of 20 applications with 50k GPU-hours/year each allowing embarassingly parallel GPU applications. Not surprisingly the prevalent topics were machine learning and artifical intelligence with applications to physics varying from quantum mechanics to weather prediction and general AI topics such as language models and their interactions. However, the classical computational physics topics, such as molecular dynamics, quantum chemistry or Galerkin methods for PDEs were represented as well.

The Austrian contribution to LEONARDO, as well as that of many other countries of the consortium, consists in financing the high-level support team (HLST). The job of the LEONARDO HLST is to support all users of the supercomputer in regards to more complicated issues, such as porting, debugging, optimizing of software to Booster, and interaction with developers with respect to LEONARDO-specific questions.

Fig. 1: Actually installed LEONARDO.

In this talk we present the general information about Booster, some early benchmarks and discuss the AURELEO call.

References

[1] https://leonardo-supercomputer.cineca.eu/

[2] https://vsc.ac.at/access/aureleo-austrian-users-at-leonardo-supercomputer

EuroHPC Vega, 2 years & counting ...

Aleš Zemljak[a], **Andrej Filipčič**[b], **and Dejan Valh**[a]

[a] *IZUM - Institute of Information Science, Slovenia*
[b] *Jožef Stefan Institute, Slovenia*

HPC Vega, the 1st in the batch of 8 selected EuroHPC supercomputers recently entered its 3rd year of operation. Named after the greatest Slovenian mathematician Jurij Vega, it remains Slovenia's most powerful supercomputer. Reaching a sustained performance of 6.9 PFlop/s, it continues to serve Slovenia's scientific community, as well as researchers across the European Union.

In fact, HPC Vega not only remained as popular with users as in its inaugural year, but in its 2nd year of operation, hosted even more projects and users. At the beginning of 2023, we have already hosted an accumulated sum of almost 200 projects and 470 users, since starting in April 2021. Beside new projects and users applying for being allocated resources, there are several projects, users and groups returning to request extensions or new resources for running new projects. Several also reported successful completion of their activities on HPC Vega, including reporting significant advances in their respective fields.

While projects using our supercomputer range through various scientific domains, there are some projects that may be easier recognizable to the general public. First, the "Slovenian Genome Project", which focuses on systematically learning about the genomic variability of Slovenians, trying to faster and more reliably diagnose rare genetic diseases, easier detect new genes and develop a disease risk assessment, for diseases that represent an important health burden in our modern-day society. Also, the project "Development of Slovenian in a Digital Environment", which aims to meet the needs of computer products and services in the field of language technologies for the Slovenian language for research organizations, enterprises and the general public. It is creating open source tools for the Slovenian language in the digital environment. And lastly, let's not forget the company "In silico", a commercial user from Croatia, harnessing the power of HPC Vega to ensure its costumers the best ship propellers and overall vessel's hydrodynamics possible, to safely and comfortably navigate the seas and their wavy nature.

Fig. 1: HPC Vega CPU blade - 3 CPU nodes

	EuroHPC JU	SI
2022-01	112.800	4.287.925
2022-02	614.614	2.379.253
2022-03	1.564.340	1.998.779
2022-04	2.244.704	2.436.402
2022-05	1.612.510	2.197.189
2022-06	816.797	3.305.099
2022-07	843.310	3.465.315
2022-08	948.659	3.559.294
2022-09	1.916.137	1.093.007
2022-10	1.294.753	711.916
2022-11	1.226.910	660.284
2022-12	706.442	661.056
Average	1.158.498	2.229.627
SUM	13.901.976	26.755.519
% of total	34,19%	65,81%

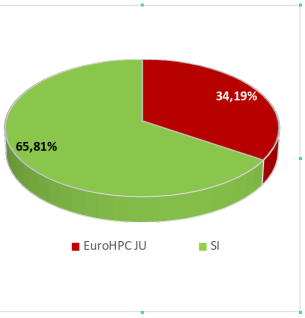


Fig. 2: HPC Vega GPU Partition usage in 2022 - month-by-month

The distribution of usage among the various user categories reveals interesting insights into our users needs and requirements. While the CPU partition was, in the greatest part, taken advantage of by pledged and opportunistic job projects, like ATLAS and to a lesser extent Belle 2, about 27% of resources were used by Slovenian, EuroHPC or Commercial projects. The most attention in 2022 though was on the GPU partition, which exhibited a strong increase of user interest, both from Slovenian and EuroHPC projects. It is therefore less surprising that on the GPU partition side, the ratio between Slovenian and EuroHPC usage, almost precisely resembles the ownership shares ratio, i.e. SI : EuroHPC 65% : 35%.

That said, perhaps somewhat surprisingly, but very interesting to witness, from the start of 2023, all of this has changed dramatically. Since then, the major utilizers of HPC Vega are EuroHPC Regular Access awardees, as well as a handful of Slovenian research projects, pushing the opportunistic jobs share down to about only 20%, essentially throwing all predictions based on analytics of past usage out the proverbial window.

Status and developments at the Vienna Scientific Cluster

Jan Zabloudil[a,b], **Moritz Siegel**[a], ***Sanaz Sattari***[a], **Ivan Vialov**[a], **Markus Hickel**[a], **Katrin Muck**[a], **Josef Beiglböck**[a], **Siegfried Höfinger**[a], **Elias Wimmer**[a], ***Florian Goldenberg***[a], **and Markus Stöhr**[a,b]

[a] *VSC Research Center, TU Wien, Austria*
[b] *BOKU-IT, University of Natural Resources and Life Sciences, Austria*

VSC-5 represents the latest HPC cluster generation operated at the Vienna Scientific Cluster (VSC). It is currently the most powerful academic supercomputer in Austria and with a peak performance of 4.30 PFlop/s ranked 301/335 in the June/November 2022 Top500 list. The system features a parallel CPU partition with 710 compute nodes and a GPU partition with 60 Nodes. Overall the system is equipped with 1540 AMD EPYC 7713 CPUs (98560 cores) and 120 NVIDIA A100 GPUs. It is a direct (hot) water cooled system and we will discuss the experiences made in terms of stability, energy efficiency, and maintenance demands.

Furthermore, in this past year we have spent considerable efforts into how to install and maintain application software and make it accessible to the user community. We will report on the lessons learned when working with the SPACK software management system and on the direction in which we are heading.

Since both clusters currently in operation (VSC-4 and VSC-5) share a common file-system, access to the same data is now possible from all compute and login nodes. Therefore we have added the option to submit jobs to both clusters independent of the login node used. We will discuss on the design of SLURM to allow for this feature and the consequences for cluster users.

Finally, this talk will close with an outlook to future developments at the VSC.

Fig. 1: Views of VSC-5.

Quantum Austria and Multi Site Computer Austria (MUSICA)

Ernst Haunschmid

TU Wien, Austria

On behalf of the Federal Ministry of Education, Science and Research (BMBWF) and funded by the European Union under the Development and Resilience Plan (2020-2026), the Austrian Research Promotion Agency (FFG) and the Austrian Science Fund (FWF) are implementing the Quantum Austria Funding Initiative from 2021-2026. FFG and FWF will use a selection of their respective funding instruments for personnel and infrastructure for this purpose.

Part of the funding is provided for research infrastructure in the highly innovative field of next-generation high-performance computing, quantum computing as well as their interconnection. The scientific issues of the projects in the field of quantum research and quantum technology may cover, for example, the following topics: - Specific preparation and control of quantum states; - New algorithms and mathematical theoretical concepts that make use of the superposition and entanglement of quantum states; - Developments and applications in quantum communication; quantum sensor technology, quantum metrology, quantum simulation, quantum computing, and quantum information; - Development of ideas based on quantum phenomena in the neighbouring fields of physics, mathematics, chemistry, and in biological systems.

The project proposal "MUSICA" (Multi-Site Computer Austria) with a funding amount of 20 million euros was approved in April 2022 and was started in October 2022. This cooperative project under the consortium leadership of the Vienna Scientific Cluster aims to build up a powerful additional research infrastructure in the field of high-performance computing in Austria in the coming years. This new HPC infrastructure will be distributed over three locations in Austria (Vienna, Innsbruck, Linz), whereby it should be a single system from the user's point of view. In addition to the general strengthening of traditional HPC resources, the focus is particularly on supporting Artificial Intelligence with dedicated accelerators and integrating cloud native applications on top of an on-premises cloud infrastructure.

In addition to MUSICA there is QACI (Quantum Accelerated Computing Infrastructure), another project funded within Quantum Austria and located at the University of Innsbruck; QACI is funded with 9 million euros and includes the procurement of a quantum computer and the interface to MUSICA. A fully integrated, hybrid infrastructure for quantum and high-performance computing is to be created, which, for example, enables the execution of hybrid quantum algorithms.

The UIBK HPC Ecosystem - LEO5

Martin Thaler

IT-Services (ZID), Universität Innsbruck, Austria

High Performance Computing (HPC) has a long tradition in Innsbruck, for almost 20 years there has been a close cooperation between the Research Area Scientific Computing [1] and ZID (IT Services). Our current local systems [2] are LEO5 (3968 cores, 24.8 TB RAM, 54x Nvidia Ampere PCIe 4.0 GPUs, HDR 100 Gbit/s IB network, 2023), LEO4 (1452 cores, 8 TB RAM, 4x Nvidia Volta SXM2 GPUs, EDR 100 Gbit/s IB network, 2018) and LEO3E (900 cores, 4 TB RAM, FDR 56 Gbit/s IB network, 2015). LEO3 was decommissioned in May 2022 after 10.5 years. It was a high performing, efficient and reliable workhorse over its entire lifetime.

HPC cluster LEO5: This cluster is another step in our strategy of renewing LEO cluster systems about every four years. As a starting point, we conducted a user survey to better understand the core needs of our scientists. Based on the results of the user survey and our system-monitoring data (Check_MK and Splunk), we designed a cost-effective and extensible architecture for LEO5 (HDR 100 Gbit/s, CPU-based with many enterprise GPUs for DP/SP/ML&DL performance).

Server room: All components have redundant power connections, as we have to perform a RCD test every 6 months. The entire system is protected by UPS and connected to a diesel generator. With the help of smart PDUs and the integrated service processors (BMC/IMM2) we measure the power consumption of the individual hosts including all network and storage components. In addition, we are currently installing power meters to cover the entire server room (incl. air-conditioning cabinets and the rooftop chillers).

Hardware and performance: We almost quadrupled the HPL CPU performance (from 55 TFLOPS to 235 TFLOPS), and the theoretical peak performance (DP/SP/ML&DL) for the GPUs increased by a factor of about up to 10/12/20. We have increased the high mem nodes by tripling the number of cores (320 in total) and quintupling the RAM (10 TB in total). On the Infiniband side, we reduced the blocking factor from 1:2.6 to 1:1, and a 40-port IB-HDR switch from Nvidia (Mellanox) acts as a managed director switch (37 ports are available for breakout cables, 3 dedicated ports for file servers and login nodes). Compared to LEO4, we have taken a significant step forward in scratch capacity and performance. We have increased the high-performance storage tier to 100 TB and the total scratch capacity to 1.8 PB, plus IOPS and bandwidth performance have been more than tripled. We have a total of 86 SSDs with 1.6 TB each (9x RAID5, 4 hot spares) and 216 HDDs with 12 TB each (4x DDP with 54 drives each, a total of 192 TB hot spare capacity).

Operational side: We chose Slurm as the batch scheduler for LEO5. For performance reasons, we have installed the dedicated Nvidia IB drivers and HPC-X software (ucx, xpmem, knem), which we have tightly coupled with the batch scheduler. At the moment, we are deciding how to evaluate Slurm's job statistics for the next few years (self-developed dashboard in Splunk or external software such as Open XDMoD [3]). To extend the Storage Scale monitoring GUI, we are currently installing the Scale Bridge for Grafana [4]. We are also evaluating to export the LEO5 scratch with CNFS (clustered NFS) to the other LEOs. All of those aforementioned actions will intend to benefit our customers and make their work more convenient and comfortable.

References

[1] University of Innsbruck - Research Area Scientific Computing - https://www.uibk.ac.at/scientific-computing/index.html.en

[2] University of Innsbruck - ZID (IT Services) - https://www.uibk.ac.at/zid/systeme/hpc-systeme

[3] Open XDMoD - https://open.xdmod.org/10.0/index.html

[4] IBM Spectrum Scale Bridge for Grafana - https://github.com/IBM/ibm-spectrum-scale-bridge-for-grafana

Software Deployment at the UIBK LEO5 Cluster

Michael Fink

IT-Services (ZID), Universität Innsbruck, Austria

The usefulness of HPC systems crucially depends on the availability of end-user facing software. Due to the immense increase in software diversity and complexity, and users' expectations of software availability from day 1, traditional methods of manual or semi-automatic installation of a large software portfolio become effectively unfeasible.

For our rollout of large portions of the software portfolio on the newly acquired Leo5 cluster, we leverage the collective know-how of the HPC and scientific software communities, substantiated by fully automated software installations using the Spack [1] and Anaconda [2] software management systems. Combined, the current installation comprises more than 600 distinct software packages and libraries, many of them in several versions, variants, and compiled for different toolchains.

In this presentation, we describe our Spack and Anaconda setups. Highlights will include local modifications, our long-term deployment strategy involving the coexistence of multiple Spack/Anaconda instances, challenges and surprises, examples of middleware integration, and our experience with the usability of the systems.

References

[1] Todd Gamblin, Matthew P. LeGendre, Michael R. Collette, Gregory L. Lee, Adam Moody, Bronis R. de Supinski, and W. Scott Futral. The Spack Package Manager: Bringing Order to HPC Software Chaos. In Supercomputing 2015 (SC'15), Austin, Texas, November 15-20 2015. LLNL-CONF-669890.

[2] Anaconda Software Distribution, https://docs.anaconda.com

Supek - Croatian Scientific and Educational Cloud HPC Resource

Emir Imamagić, Jurica Špoljar, and *Martin Belavić*

University Computing Centre, University of Zagreb, Croatia

The main goal of the Croatian Scientific and Educational Cloud (HR-ZOO) project [1] was to build a national research and innovation e-infrastructure for the Croatian scientific and academic community. The HR-ZOO infrastructure [2] provides a long-term advanced computing, storage and network resources as well as scientific software and specialized support for researchers that are necessary for multidisciplinary science and education.

The advanced computing component consists of a 1.25 PFlop/s high performance computing (HPC) resource, a cloud computing resource with over 11,000 processor cores, and a virtual data centre infrastructure for the deployment of critical services. The storage component provides 10 PB of storage accessible via Network File System (NFS) or Object Storage S3 protocols. The resources are deployed in five data centers connected by the new 100 Gbit/s network infrastructure.

In this talk, we will focus on the HPC component of the HR-ZOO infrastructure - a resource named Supek. Supek is based on HPE Cray EX2500 and HPE ClusterStor E1000 systems. Supek consists of three partitions - the CPU partition with 52 worker nodes providing a total of 6656 AMD Epyc 7763 CPU cores, the GPU partition with 20 worker nodes providing a total of 1280 AMD Epyc 7763 CPU cores and 80 NVIDIA A100 and the big memory partition with 2 worker nodes providing 128 AMD Epyc 7763 CPU cores and 4 TB of RAM each. HPE ClusterStor E1000 storage is all-flash based and provides 580 TB of storage with a measured performance of: 240 GB/s sequential write, 450 GB/s sequential read and 3.2 MIOPS random read/write. Supek's interconnect technology is 200 Gbit/s Slingshot and the batch system is PBS Professional.

We will present the challenges we had in migrating users from the existing typical Beowulf architecture cluster Isabella based on Intel CPUs, BeeGFS shared storage, Infiniband interconnect and Son of Grid Engine batching system. Furthermore, we will present the challenges we had in preparing scientific applications that we provide to our users. Special focus will be on the advantages and optimization in resource allocation that we achieved by using the PBS Professional batching system.

Finally, we will present how HPC users will benefit from other components of the HR-ZOO infrastructure – namely cloud computing resources that will be used to offload applications that do not fit into the HPC domain, and storage resources for offloading data from the Supek's high-performance storage.

Fig. 1: HR-ZOO logo.

References

[1] https://www.srce.unizg.hr/hr-zoo/

[2] Dobrenić Dobriša, Imamagić Emir, Marić Ivan, Project HR-ZOO - The Final Stages of Implementation, Computational Chemistry Days (2022).

Keynote Talk:

HPC as an opportunity for innovative SMEs

Tomi Ilijaš

Arctur d.o.o., Slovenia

High-performance computing (HPC) is a game-changer for businesses looking to stay ahead of the competition. By leveraging cutting-edge technologies such as cloud supercomputing, artificial intelligence (AI), machine learning (ML), and big data analysis, companies can develop products and services with higher added value, paving the way for novel industrial applications. Embracing HPC can lead to significant business benefits, including reduced costs for product and service development, considerable savings in human resources costs, faster development processes, and decreased time to market.

European Small and Medium-sized Enterprises (SMEs) are the backbone of Europe's economy and as such, important contributors for the economic development. In Europe, the HPC landscape is rapidly evolving. European Commission [1] and the European High Performance Computing Joint Undertaking (EuroHPC JU) [2] are encouraging European industry, especially SMEs, to adopt state-of-the-art technologies and shift towards innovation by offering calls for proposals, co-financed R&D projects and offering funding opportunities such as Open Calls. Moreover, the EuroHPC JU has also launched the EuroHPC supercomputers, a pan-European network of petascale and pre-exascale supercomputers located across the EU. These supercomputers provide significant computational power to researchers and industries, enabling them to solve complex problems and innovate in various fields.

To help SMEs fully understand the world of HPC and its use in business, Centres of Excellence (CoEs) [3] and National Competence Centres (NCCs) [4] were established as main contact points and work as knowledge and action hubs which bring together the European world-class experts. National Competence Centres were established under the auspices of EuroCC project [4] and are the central points of contact for HPC in more than 30 European countries. In Slovenia, NCC HPC SLING [5] is operating as an entry point for academia, public administration, students and industry. The NCC is not only promoting the use of HPC for industry, but also offering consulting, training and expertise support for industry, access to the HPC infrastructure and HPC4SME tool, which can significantly help SMEs to foster the HPC uptake. HPC4SME Automated Assessment Tool (HPC4SME AAT) [6] was developed and is widely used among SMEs to discover if and how the organisation can benefit from supercomputing technology in R&D and other business processes.

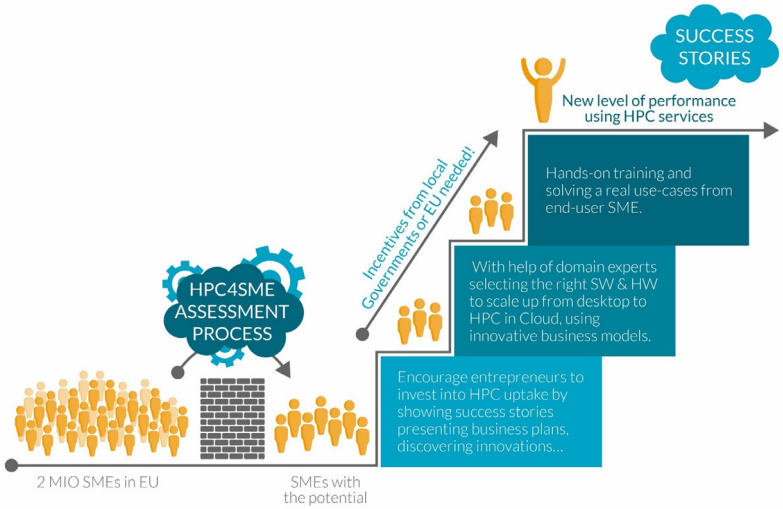


Fig. 1: Explaining the HPC4SME AAT process – reaching the SMEs with potential and help them shift towards Industry 4.0.

One of the NCC HPC SLING partners is Arctur [7], a Hi-Tech SME and the main private-owned supplier of HPC services and solutions in CEE. The company has extensive experience in development and deployment of complex IT solutions (e.g. AI, Blockchain, HPC) especially for SMEs in various sectors: from manufacturing to tourism, logistics and health. One of the major Arctur's endeavours is to raise SMEs' awareness of HPC and demonstrate its benefits through inspiring success stories. The FF4EuroHPC [8] project is an excellent example of the benefits of HPC for industry. This project focuses on developing solutions that use HPC to solve complex industry problems. The solutions are then tested in real business environments, and the resulting business benefits gained during the experiments are presented to the wider HPC community. This project demonstrates the significant impact that HPC can have on businesses and highlights the importance of innovation in a rapidly evolving business landscape. During the Industry Session, some the FF4EuroHPC Success Stories will be introduced in detail.

References

[1] European Commission: https://commission.europa.eu/index_en

[2] EuroHPC JU: https://eurohpc-ju.europa.eu/index_en

[3] Centers of Excellence – European Excellence in HPC Applications: https://www.hpccoe.eu/eu-hpc-centres-of-excellence2/

[4] EuroCC 2: https://www.eurocc-access.eu/

[5] NCC HPC SLING: https://www.sling.si/

[6] HPC4SME AAT: https://hpc4sme.eu/

[7] Arctur d.o.o.: www.arctur.si

[8] FF4EuroHPC: https://www.ff4eurohpc.eu/

Multi-head Additive Manufacturing with Optimal HPC Thermal Stabilization - AMOTS

Draško Tomić

Ruder Bošković Institute, Croatia

The use of supercomputers in the sphere of small and medium enterprises is gaining more and more importance. Supercomputers enable them to quickly develop new products that, by their functionality and quality, significantly surpass those that are developed in the classic way, through prototypes. However, supercomputing resources are relatively expensive, and their use often exceeds the financial capabilities of entrepreneurs.

With the aim of solving this challenge, a program was launched within the EuroHPC project that enables entrepreneurs to develop their creative and innovative ideas on supercomputers. As part of that program, Mikrotvornica d.o.o., Ruđer Bošković Institute, and Red Fluid, developed a model of an advanced multi-additive 3D printer [1] using OpenFoam CFD simulation software.

Experimental measurements on the printer prototype confirmed the value and usability of the CFD model developed in this way and enabled Mikrotvornica to enter the market with a highly competitive product. All OpenFoam 3D printer simulations were performed on the Yotta HPC cluster. A number of simulations were performed, and each one lasted about 4 days on average, using 4 computer nodes and a total of 96 cores. On more than four nodes, we observed a slight decrease in scalability. The reason probably lies in the high communication between computer nodes that use Mellanox ConnectX-4 2x25 Gb/s Ethernet with sub-microsecond latency as an interconnection. On the other hand, the use of fast and low-latency Ethernet enables the use of advanced HPC Cloud services, thus significantly easing the use of the HPC cluster.

Fig. 1: Finite element mesh of the 3D printer used in OpenFoam CFD simulations.

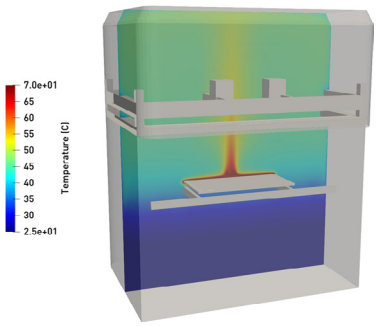


Fig. 2: Heat distribution within the CFD model of the 3D printer.

During the AMOTS project, we also evaluated the advantages and disadvantages of using our own HPC cluster and rented HPC resources. Our analysis shows that there is about a half-cost advantage to using leased HPC resources. This especially applies to those resources that are supported by cloud services, as is the case with the Yotta cluster. Not least, the Mikrotvornica staff gained working knowledge of using HPC resources and OpenFoam. So they are now in a position to continue using HPC as a tool for developing their products. Consequently, they are in a position to take a competitive position in today's challenging 3D printer market segment. And they will be able to achieve this by delivering printers faster, with greater functionality, and cheaper than those of their main competitors.

References

[1] Tomic D., Davidovic D., Subic T, Mesaric J, and Skala K., Improving 3D printing of garments by using HPC Cloud, 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 2020, pp. 233-236, doi: 10.23919/MIPRO48935.2020.9245130

Preliminary Analysis of Innovative Aerostructures Spectrum Powered by HPC

Matej Borovinšek[a], **Miran Ulbin**[a], **Zoran Ren**[a], **Jure Pocrnjič**[b], **and Sašo Knez**[b]

[a] *Faculty of Mechanical Engineering, University of Maribor, Slovenia*
[b] *AFormX d.o.o., Slovenia*

The aerospace industry has been experiencing a revolution in recent years thanks to advancements in electrification technology and initiatives such as the Green Deal and Clean Sky. To maximize the benefits of electric propulsion, innovative aircraft configurations require innovative load-carrying structures. Structural engineers need tools that can help them explore a broader range of different structural configurations early in the design process to determine the best option before focusing on optimization. The application of High-Performance Computing (HPC) in structural analysis can lead to more optimized aerostructures, reducing the risk of the design process and the weight and cost of the final product in manufacturing.

In response to these challenges, an experiment was conducted to develop a tool that uses HPC to do rapid iterations of low-fidelity Finite Element Method (FEM) structural analysis of several different configurations to find the best one. It was developed based on open-source projects CalculiX [1] and PrePoMax [2] to reduce its cost for the final user. The tool (Fig. 1) allows for the exploration of a wider range of structural configurations early in the design process, enabling structural engineers to develop a better understanding of which configuration would be most effective and efficient. The rapid iteration process also allows the engineers to maximize their efficiency and helps develop their intuition for which structural design would work best.

The developed tool is an important step towards more efficient and cost-effective structural design in the aerospace industry. By enabling engineers to explore a wider range of options, the tool helps to reduce the risk of costly errors in the design process. The benefits of using HPC in the design process were shown to be asymmetric, with the benefits being much greater when used earlier in the process.

The collaboration between the University of Maribor and AFormX was a successful example of how academic and industry partnerships can drive innovation and improve the competitiveness of the aerospace industry.

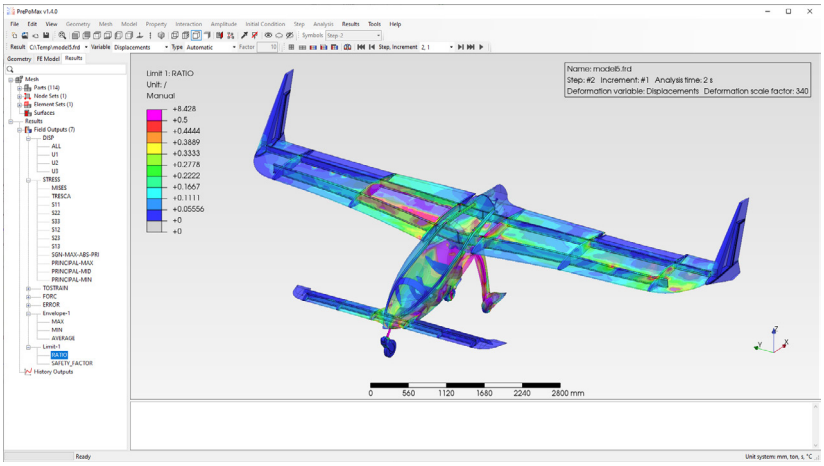


Fig. 1: Results envelope of multiple CalculiX simulations.

References

[1] CalculiX, http://www.calculix.de (2023).

[2] PrePoMax, https://prepomax.fs.um.si (2023).

Development of High-Performance- and Real-Time-Computing Software for the Simulation of Electromagnetic Fields

***Jürgen Zechner*, Thomas Rüberg, and Lars Kielhorn**

TAILSIT GmbH, Austria

For the simulation of electromagnetic (EM) fields we couple Finite Element Method (FEM) with the Boundary Element Method (BEM). The latter is generally advantageous for the simulation of radiation or wave propagation. With our approach only the workpiece needs to be discretized with a mesh such that it's behavior can be modeled with the FEM. The EM fields in the surrounding volume are completely covered by the BEM without the need of an air mesh or artificial boundary conditions at the border of the region of interest. This approach allows us to simulate movements of workpieces without cumbersome and often failing re-meshing of the air. Typical applications are the simulation of actuators like magnetic valves or magnetic latches [1].

In this presentation we discuss this approach from two different perspectives:
The level of detail in the development and simulation of electromagnetic devices is constantly increasing. This affects the complexity for the BEM software implementation. We make use of an octree-based Fast Multipole Method (FMM) in order to scale linearly with the model size. Moreover, the algorithm ideally scales with the number of CPU cores. We briefly describe the HPC-aware implementation of the algorithm as well as the current state and upcoming steps in our development. Additionally, we present some results of calculations done on the VSC-4 supercomputer [2] (see Fig. 1).

Another application of our simulation approach is the real-time computation in the field of electromagnetic navigation which can be used to display the current position of surgical instruments that are not in the surgeon's field of view. The navigation is based on the generation of a well-defined electromagnetic field that is detected by a sensor on the surgical instruments. An algorithm can calculate the spatial position of the surgical instruments from this. The situation is displayed on a screen or VR glasses together with images of the patient. The main disadvantage of this navigation approach is that ferromagnetic obstacles or components themselves generate a distortion field, which can lead to a substantial loss of accuracy. We briefly discuss our algorithm which calculates these distortion fields in real-time, thus restoring the accuracy of the navigation, and the potential of an implementation on GPUs.

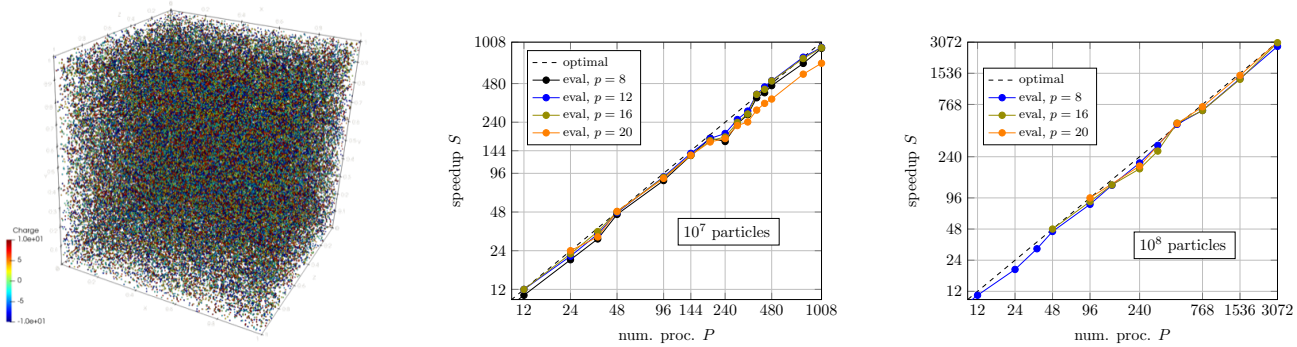


Fig. 1: Randomly charged point cloud in a unit cube (left). Speedup for the evaluation of the electric potential for 10^7 (middle) and 10^8 (right) particles with different FMM expansion degrees p.

References

[1] T. Rüberg, L. Kielhorn, and J. Zechner, Electromagnetic Devices with Moving Parts — Simulation with FEM/BEM Coupling, Mathematics **9(15)**, 1804 (2021).

[2] J. Zechner, L. Kielhorn, T. Rüberg, and S. Steiner, PARTS — Electromagnetic Simulations with Fast Boundary Element Methods using HPC, PRACE Whitepaper, https://doi.org./10.5281/zenodo.7071801 (2022).

Enabling HPC-scale Engineering Simulations for SMEs

Dávid Felhős **and László Ady**

Karman Mechanics Ltd., Hungary

There are many obstacles on the road towards application of engineering simulations when an SME would like to establish this technology in its everyday practice. One can classify these obstacles in two major groups: costs and skills. On the side of costs, we have hardware and license costs, and the costs related to engineering hours. The obstacles from the standpoint of skills concern the need for practical simulation expertise, the ability to conduct continuous development, and a deep understanding of the underlying theory.

We state that our software delivers a solution to overcome all these problems at a high technological level.

We apply open-source simulation tools just like OpenFOAM, OpenRadioss, Yade, Code-Aster, Calculix as well as ParaView and we develop simulation wizards, to automate the processes. It enables the users to get their results in many cases without facing the technological issues of numerical simulations and HPC clusters. On the other hand, we enable the possibility for advanced users to create their own optimization or arbitrary development codes on the web-based application that we call simply HPC Solver [1].

During development, we focused on the different requirements expected under the different conditions of use. We have investigated the aspects playing a role for SMEs during the integration of a solver system. Our experiences have led us to create a modular system. The HPC Solver architecture is built in a way that it can be safely and relatively simply implemented on various supercomputer architectures. The solution includes the preparation of tasks, maintenance of solvers, tracking and optimization of runs. Data and IT security were kept priority at all stages of the development [2,3].

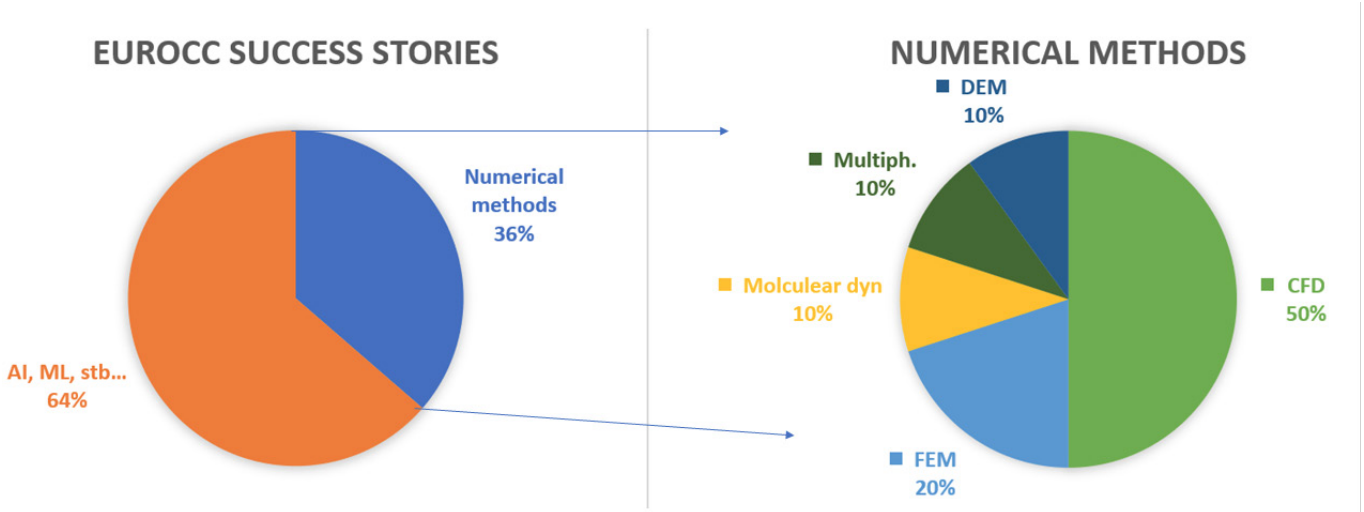


Fig. 1: Partition of engineering applications in HPC success stories.

References

[1] Gróza M., Felhős D., Open-Source Software for Engineering Simulation - A Sneak Peek, Benchmark Magazine (July 2022).

[2] Albini A., Tokody D., Rajnai Z, Theoretical study of cloud technologies , Interdisciplinary Description of Complex Systems **17** (2019).

[3] Ady L., Schuster Gy, Beware of human, 2020 IEEE 3rd International Conference and Workshop in Óbuda on Electrical and Power Engineering (CANDO-EPE) (2020).

EuroCC Supercomuting Accelerator – Grow with advanced technology

Raman Ganguly[a] **and Thomas Mayerhofer**[b]

[a] *EuroCC Austria, University of Vienna, Austria*
[b] *EuroCC Austria, INiTS, Austria*

This presentation focuses on EuorCC's second phase, where participating countries collaborate to offer robust support to industrial customers. The goal is to deliver tailored, comprehensive, and well-structured assistance for each company. Under Austria's leadership, a cooperation model takes shape: a unified service portfolio that encompasses 33 nations and targets businesses across Europe. This service package aims to be inclusive and adaptive, meeting clients' distinct needs regardless of their origin or technological advancement. The customized service package addresses each client's requirements, whether at the beginning of their high-performance computing (HPC) transformation journey or already immersed in the process. Clients from any participating country – Germany, France, or Austria – can access the expertise and infrastructure of all EuroCC members using a dedicated application link and a streamlined process.

Offered services include:

1. Tech Feasibility Check: EuroCC experts set up a video meeting to discuss potential technical solutions and ensure supercomputing is a suitable technology for the business.
2. Get HPC Ready: Learn about high-performance computing and prepare to use HPC effectively. A technical overview of available hardware, common software stacks, and workflows is provided, with the possibility of creating custom training or workshops based on specific needs.
3. Business Plan: Creation of a customized financial plan to support the success of a proof of concept.
4. Financing Advice: Assistance obtaining state funds for implementing AI, Big Data, or HPC projects.
5. PoC Programming Support: Collaboration with experts to analyze code and receive suggestions on improving performance for HPC, including parallelization with OpenMP, MPI, or offloading to GPUs. National HPC infrastructure serves as the test bed for the proof of concept.
6. Project Support: Extensive support for AI, Big Data, or HPC projects, such as finding partners, appropriate resources, and ongoing assistance throughout the project.
7. Access to HPC Infrastructure: EuroCC experts facilitate access to national HPC infrastructure for testing code or running projects. If more resources are required than are available in a particular country, access to powerful machines across Europe can be granted, with training and onboarding included.

The initiative aims to break down national barriers and serve the European market with a comprehensive and indispensable service portfolio. Initially, the cooperative effort will focus on assisting startups and small to medium-sized enterprises.

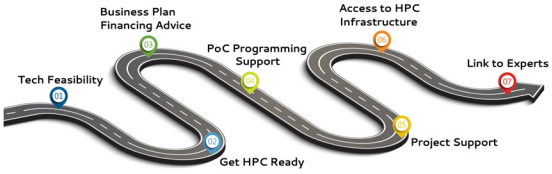


Fig. 1: The way of the offered services.

NCC Czech Republic, entrance to the world of HPC

Tomas Karasek

IT4Innovations, VSB – Technical University of Ostrava, Czech Republic

The National Center of Competence (NCC) Czech Republic was established in 2020 as one of the 31 NCCs across Europe under the auspices of the EuroCC project. Since 2023, its activities have continued under the EuroCC 2 project.

Our **mission** is to provide access to the knowledge and support for the use of High-Performance Computing (HPC) and associated technologies such as High-Performance Data Analytics (HPDA) and Artificial Intelligence (AI) for all stakeholders from the academia, industry and public institutions. Our **vision** is to be a single point of contact for industry, academia and public institutions from the Czech Republic to get access to knowledge and competencies in the HPC and associated technologies and through the network of NCCs make HPC competencies in the Czech Republic accessible Europe wide. We aim to increase the uptake of HPC technologies by stakeholders in the Czech Republic, increase their awareness and preparedness and improve their digital skills in HPC and associated technologies.

In this talk, activities of the NCC, Czech Republic, consisting of training activities, workshop organisation, collaboration with the industry, academia and public sector, and collaboration with other NCCs will be presented. The main focus will be on the success stories resulting from the collaboration with Small and Medium Enterprises (SME) and public administration. Three success stories will be presented to demonstrate the use of HPC technologies in the Energy, Healthcare and Manufacturing sectors.

The objective of the project with Orgrez a.s. was to determine whether Computational Fluid Dynamics (CFD) simulations could be used for both the fast and efficient description of the Selective Catalytic Reduction (SCR) process and, therefore, as a tool to mediate the design of a computational application for the design of this technology.

The main goal of our cooperation with University Hospital Ostrava was to deploy and test a tool providing remote automatic tissue segmentation from patient image data obtained from computed tomography (CT) or magnetic resonance imaging (MRI) on HPC infrastructure at IT4Innovations. The methods used to enable tissue segmentation were based on deep learning (DL). The objectives were (i) to provide a service based on state-of-the-art algorithms for automatic segmentation of the desired tissues as an AI-based annotation service and (ii) to collect the data after the automatic segmentation and validation by medical doctors and to provide HPC-based training of new models or enhancement of the existing models through fine-tuning.

The main objective of the collaboration with the NCC and the Armatury group was the proof-of-concept, which would compare available CFD Simulation approaches for solving moving bodies within the fluid domain. All the methods can be used to study stationary and time-dependent problems where the geometry changes shape due to the motion of solid boundaries and the deformation of solid domains. Each method has advantages and disadvantages, and their direct comparison of the specific case provides valuable information for deciding which approach should be used in the future.

PLGrid Infrastructure in the Exascale Era

Marek Magryś **and Marta Maj**

Academic Computer Centre Cyfronet AGH, Poland

The Polish Grid Infrastructure [1] was built and expanded as part of the PL-Grid project (2009-2012) and its next editions to provide the Polish scientific community with an IT platform based on computing clusters serving e-Science in various fields. The infrastructure supports scientific research by integrating experimental data and results of advanced computer simulations conducted by geographically dispersed teams. PLGrid Infrastructure enables Polish scientists to conduct scientific research based on large-scale simulations and computations using supercomputers and clusters of commodity servers in a distributed computing environment.

The Polish Grid Infrastructure is managed by the PLGrid Consortium, established in January 2007, which consists of the following institutions: Academic Computer Centre Cyfronet AGH, Krakow (coordinator), Interdisciplinary Centre for Mathematical and Computational Modelling, Warsaw University, Warsaw, Poznan Supercomputing and Networking Center, Poznan, Academic Computer Centre, Gdansk, Wroclaw Centre for Networking and Supercomputing, Wroclaw and National Centre for Nuclear Research.

Beyond the resources of the above institutions, the PLGrid infrastructure is also providing access for Polish researchers to the LUMI Supercomputer [2], build by a consortium of 10 European countries in Kajaani, Finland. LUMI is currently ranked third on the Top500 list of the world's fastest supercomputers – and ranked fastest supercomputer in Europe.

Currently, the PLGrid Consortium constitutes The Polish National Competence Center (NCC) in High-Performance Computing (HPC) [3]. As a part of EuroCC and EuroCC2 project, it is responsible for promoting the use of HPC in scientific research and industry in Poland and aims to develop a national research infrastructure in HPC.

The Polish NCC is based at the Academic Computer Center Cyfronet AGH in Krakow, Poland, and provides a range of services and resources to support HPC users in Poland, including access to HPC systems, software and application support, training and education programs, and consultancy services.

Furthermore, the EuroHPC PL project was initiated to build a specialized general-purpose infrastructure for large-scale computing, enabling the undertaking of research challenges in key areas from the point of view of Polish society, the scientific community and the economy. The project is the Polish stage of development of the EuroHPC program.

Involvement in the above projects shows that the PLGrid Infrastructure continues to grow and evolve using national and EU funding, playing a vital role in supporting scientific research in Poland through its provision of high-performance computing resources and services.

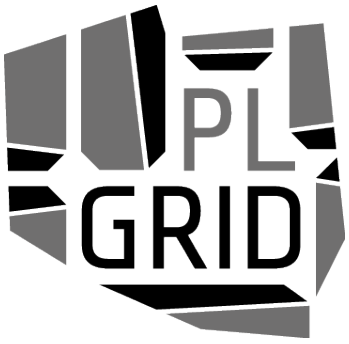


Fig. 1: PLGrid logo.

References

[1] https://www.plgrid.pl/

[2] https://www.lumi-supercomputer.eu/

[3] https://cc.eurohpc.pl/

SMASH – Marie Skłodowska-Curie COFUND project for postdoctoral researchers using HPC VEGA to develop machine learning applications for science and humanities

Gabrijela Zaharijas

Center for Astrophysics and Cosmology, University of Nova Gorica, Slovenia

This article presents the SMASH project [1] - an innovative and inter-sectoral career-development training program for postdoctoral researchers; focused on developing cutting-edge machine learning applications in the sciences and humanities.

The SMASH project is a European project led by the University of Nova Gorica and co-funded by a Horizon Europe, Marie Skłodowska-Curie COFUND Action [2]. SMASH aims to support 50 postdoctoral researchers who will propose innovative and ambitious research projects related to 5 key research areas and 17 research sub-areas, focused on the use of machine learning or, more broadly, artificial intelligence techniques using HPC VEGA to address some of the world's most challenging questions in:

- Data Science - Machine Learning for Scientific Applications
- Fundamental Physics - Machine Learning for Particle Physics, Astrophysics, and Cosmology
- Linguistics - Computing for Human and Animal Communication
- Climate - Machine Learning in Climate Research
- Precision Medicine - Personalized Medicine and Life Sciences

Postdoctoral researchers will have the opportunity to submit their research-oriented projects to three calls launched during the five-year project period and to choose one of the five top-ranked Slovenian research institutions (University of Nova Gorica, University of Ljubljana, Jožef Stefan Institute, Institute for Information Sciences and Slovenian Environment Agency) as a host institution.

Selected applicants will be offered a 2-year full-time employment contract as a Marie Skłodowska-Curie COFUND fellows and full academic freedom to pursue their research projects at one of the SMASH host institutions coupled with personalized training program that meets the needs of each fellow alongside active career mentoring by supervisors from SMASH academic partners and, in addition, 2 short-time trainings (secondments), one in Slovenia at an academic or non-academic organisation and the second at one of the 23 associated partner institutions like UC Berkeley, University of Washington, CERN, Queen Mary University of London, Weizmann Institute, University of Amsterdam, etc.

Our first call is now closed and we are evaluating the applications. We aim at hiring 10-15 fellows in the first call. We plan to have Call 2 and 3 open in July 2023 and 20024 respectively, with application deadlines in October 2023 and 2024.

References

[1] https://smash.ung.si/

[2] https://cordis.europa.eu/project/id/101081355/

Keynote Talk:

Lost in Vega(s): learning how to translate between English and Slovene on Vega

Iztok Lebar Bajec

Faculty of Computer and Information Science, University of Ljubljana, Slovenia

In the digital world, the Slovene language is often overlooked; the big tech companies do not seem to want to invest too much time and effort in such a "small" market. The result are user interfaces that are not translated into Slovene, cell phones, smart assistants and cars that neither understand nor speak Slovene. The main reason is usually the lack of publicly available training datasets with free licenses. An overview of the current state shows that Slovene falls into the category of resource-poor languages here.

To alleviate this problem, the Development of Slovene in a Digital Environment (RSDO) project aims to address the need for computer-based tools and services in the field of Slovene language technologies. The corpora created and the tools developed will be licensed under free licenses and made available to research institutions, businesses, and the general public through public repositories such as CLARIN.si.

The RSDO project is a joint venture of major Slovene universities, institutes, and private companies. Seven main areas of interest have been identified, ranging from language resources, language technologies, semantic resources, machine translation and terminology to infrastructure.

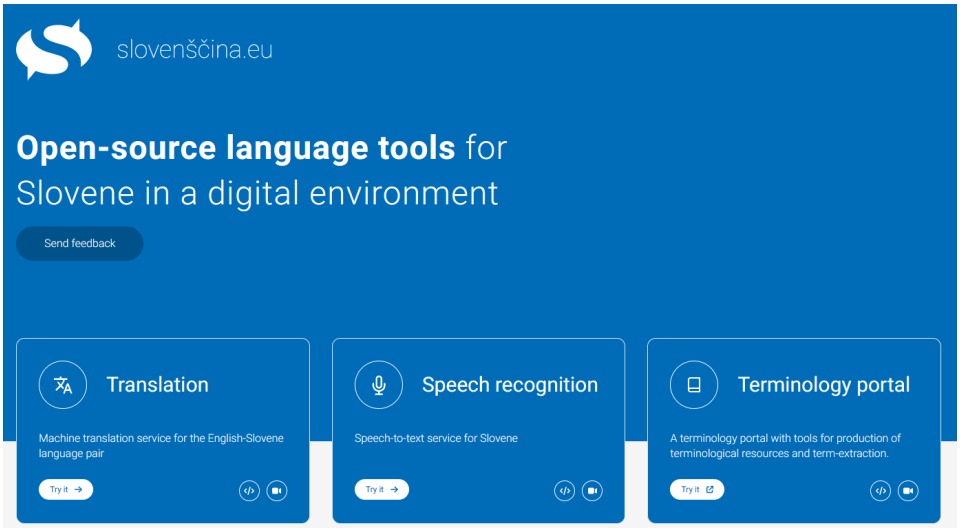


Fig. 1: Demo portal of the Development of Slovene in the Digital Environment (RSDO) project. Free access to demo tools at https://slovenscina.eu.

Here we focus on machine translation, with goals such as collecting texts for the translation corpus, developing tools and defining evaluation methods, testing alternative neural machine translation (NMT) frameworks, developing NMT models in accordance with the growth of the translation corpus, developing a demo web portal, and preparing a long-term plan for the development of future machine translation systems, whether for training or general use.

Our toolkit of choice was NeMo, a computational AI toolkit [1] actively developed by the community with significant participation from NVIDIA researchers and engineers. It is based on Lightning (https://github.com/Lightning-AI/lightning), which itself is a wrapper over Pytorch (https://github.com/pytorch/pytorch). The NeMo Computational AI toolkit thus provides easy scalability to multi-GPU and multi-node/multi-GPU environments. For example, NeMo Megatron LLM models can be trained with up to 1 trillion parameters

using tensor and pipeline model parallelism. These approaches have been used by Microsoft and NVIDIA in training Megatron-Turing NLG 530B, a large-scale generative language model [2].

The NeMo toolkit is written to easily take advantage of modern architectures (Ampere and Hopper) and their support for bf16 [3] and fp8 [4] float representations.

Although we used containers for toolkit setup and experiment reproducibility, training NMT models with a DL toolkit under active development is far from straightforward. The toolkit is constantly being updated, often breaking backward compatibility. This is especially true for HPC systems that traditionally prefer proven, stable code rather than taking advantage of the ever-changing SOTA tools. This is where the help of knowledgeable and helpful HPC support teams becomes an indispensable asset, helping to move faster and achieve results more quickly than if you were on your own.

References

[1] Kuchaiev O, et al. 2019. NeMo: a toolkit for building AI applications using Neural Modules. arXiv:1909.09577

[2] Smith S, et al. 2022. Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, A Large-Scale Generative Language Model. arXiv:2201.11990

[3] Micikevicius P, et al. 2022. FP8 Formats for Deep Learning. arXiv:2209.05433

[4] Kalamkar D, et al. 2019. A Study of BFLOAT16 for Deep Learning Training. arXiv:1905.12322

Time series in HPC: An exploratory study with distributed time series data for energy use cases

Soner Steiner[a], **Rosana de Oliveira Gomes**[b], **Gregor Beyerle**[b], **Dimitar Dimitrov**[b], **and Jan Zabloudil**[a,c]

[a] *VSC Research Center / EuroCC Austria, TU Wien, Austria*
[b] *HAKOM Time Series, Austria*
[c] *BOKU-IT, University of Natural Resources and Life Sciences, Austria*

This project resulted from a collaboration initiated by EuroCC between TU Wien, the University of Natural Resources and Life Sciences (BOKU), and HAKOM Time Series. These three parties submitted together to the call "HPC Exploratory Projects" at the Austrian research promotion agency (FFG). The project has been granted and started in November 2022.

The HAKOM Time Series Manager (TSM) technology is designed to meet the operational needs of all players in the energy industry. It offers users an automated data management tool, able to perform classical tasks such as creating, updating, and aggregating time series data.

The goal of this project is to deploy, benchmark, optimize the TSM framework on the Vienna Scientific Cluster and evaluate computationally highly demanding operations. The high computational processing requirements come both from a large data volume and from the involved algorithmic calculations.

Dynamic time warping (DTW) in the field of time series analysis basically means a measure for time series with different time lags. See for example two time series graphically depicted in Fig. 1. An everyday example of two time series with different time lags would be two persons walking next to each other with different velocities. In such cases DTW is the measure of choice, because it can handle vectors with different lengths. In Euclidean distance or cosine distance metric the vectors need to be of the same size.

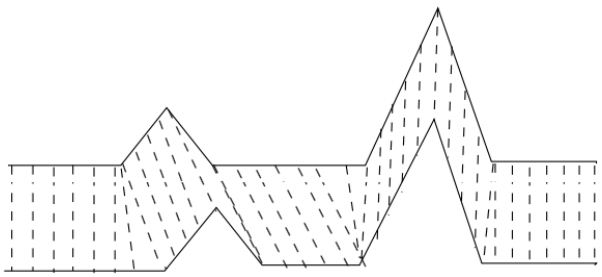


Fig. 1: Depicting DTW visually, taken from [1].

In this work, we apply DTW as a metric to classify time series into different groups, e.g., load profiles of households, of small businesses like hairdressers, and many more.

The data set used in this study comes from a synthetic source, namely the *Last Profile* dataset [2], which simulates the profile of households measured from smart meters. The data set contains time series classified into 27 different profiles. The data volume is enhanced by adding white noise to these profiles. We finally test our algorithm by classifying the generated profiles and evaluating its accuracy.

References

[1] https://en.wikipedia.org/wiki/Dynamic_time_warping

[2] https://www.apcs.at/de/clearing/technisches-clearing/lastprofile

Ensuring the quality of HPC software development

Alja Prah[a,b], **Jan Jona Javoršek**[a], **and Barbara Krašovec**[a]

[a] *Network Infrastructure Centre, Jožef Stefan Institute, Slovenia*
[b] *Theory Department, National Institute of Chemistry, Slovenia*

Utilising DevOps practices to improve speed and quality of software development has been a hot topic the past few years. These practices include the development of code, its review, testing, and deployment. Large projects are broken into smaller tasks, so that it is easier to respond to users' requests and to the changes on the target architecture and environment.

MAX (MAterials design at the eXascale) is a European Centre of Excellence which enables materials modelling, simulations, discovery and design at the frontiers of current and future High Performance Computing (HPC), High Throughput Computing (HTC) and data analytics technologies. Software integration in this context means connecting the development efforts with the environments that are deployed at different supercomputing sites and with the users.

For improving the quality of code and identifying and fixing bugs, it is very important to include testing as part of the development process from the beginning. It is usually done by the QA team or by implementing continuous integration and delivery (CI/CD) into the process.

GitLab is a git-based code hosting service, a platform that supports the entire DevOps cycle. The key to automation is CI/CD, which enables automated testing, building and deployment of code in the HPC environment, including software security testing and performance testing [1,2]. If the test fails or succeeds, feedback is immediately available in GitLab. GitLab enables CI workflows through its runner utility. This is a process that runs on a target machine and is polling the GitLab server for jobs. An event triggers the runner (e.g. a commit), it executes the script and sends the result back to GitLab, which enables immediate feedback on how the code builds in real HPC environment and how well it performs [3]. Shell executor therefore seems like a good solution but it has two major drawbacks: it does not consider multi-user environments and it does not interact with the batch system, which causes another problem, how to ensure security and isolation of the whole testing process. Federated capabilities of the deployment should be studied as well, including integration with the existing LDAP system.

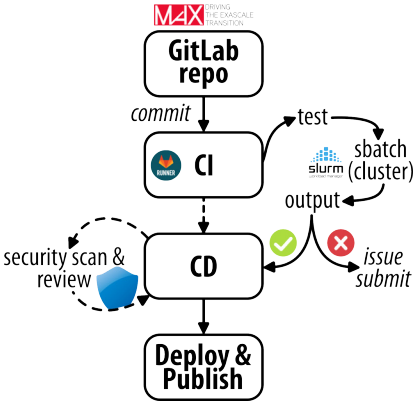


Fig. 1: Simplified representation of the software development scheme.

After the successful DevOps process, the software should be made available for users and host entities. This could be done by publishing in the CVMFS (CernVM File System) repository, providing optimised software for specific CPU architecture and/or GPU compute capabilities. Further deployment includes building containers that will be shared in a container image registry, such as Harbor and in the CVMFS repository as the filesystem, using CVMFS DUCC (Daemon that Unpacks Container Images) [4]. Containers are a good solution for reproducibility and portability of the execution environment.

References

[1] Sampedro, Z., Holt, A., Hauser, T., PEARC '18 (2018).

[2] Robinson, A., Swan, M., Harvey, E., Klein, B., Lawson, G., Milewicz, R., Pedretti, K., Schmitz, M., Warnock, S., Report No. SAND2022-11847 (2022).

[3] Min, E., Liu, F., Rahaman, R., PEARC '22 (2022).

[4] Bocchi, E., Blomer, J., Mosciatti, S., and Valenzuela, A., EPJ Web of Conferences **251**, 02033 (2021).

SCtrain: Supercomputing knowledge partnership

Pavel Tomšič

Faculty for Mechanical Engineering, University of Ljubljana, Slovenia

http://sctrain.eu/

About:

The Erasmus+ SCtrain project is a consortium with the goal of raising HPC knowledge inclusion for future professionals. Through knowledge research and topic development we are raising the level of competences for future science, technology, engineering and mathematics (STEM) professionals between Central European countries, including Italy, Slovenia, Austria and Czech Republic. Our project partners:

• *University of Ljubljana, Slovenia* • *VSB – Technical University of Ostrava, Czech Republic* • *CINECA Consorzio Interuniversitario, Italy* • *TU Wien, Austria* •

Who is it for:

For students: Gaining skills that enable future competitiveness. **For educators:** Raising the level of competences in theoretical, programming, mathematical and teaching skills. **For business experts:** Gaining skills for professional growth and competitiveness with introduction of HPC. **For future HPC HE courses:** A comprehensive set of learning and teaching materials covering HPC in engineering and data science.

What we have achieved:

During the duration of the project there are many goals to be achieved. The intellectual outputs are included in our publicly avaliable knowledge base on our project website in the form of lectures (core knowledge) and exercises/tutorials (hands-on approach). All output material has been tested through training events that serve both the verification of knowledge transfer to students and the enhancement of the educators by gaining experience. Topics include:

- HPC in Engineering – focus on FEM (Finite Element Method) – finished in Autumn 2021,(1)
- HPC in Data Science – focus on Parallelizing with MPI – finished in Spring 2022,(1)
- HPC in Engineering – focus on CFD (Computational Fluid Dynamics) – finished in Autumn 2022,(1)
- HPC in Data Science – focus on IOT and Big Data, *
- A collection of knowledge on the topic of HPC in Engineering and Data Science. *

* *in development Summer 2023*

(1) *the output material is publicly available in the form of videos and material through our website http://sctrain.eu/ (in the menu choose KNOWLEDGE). Further material is available on demand through our E-classroom.*

Future plans:

Currently the SCtrain consortium is developing material for HPC in Data Science - with a focus on IOT and Big Data that will be tested in a one week course that will take place from June 26 to 30, 2023, at IT4Innovations National Supercomputing Center, VSB – Technical University of Ostrava, Czech Republic. The topics aimed at High-Performance Data Analytics (HPDA) are covering data cleaning, exploratory data analysis, modelling using machine and deep learning/AI, and up-scaling the codes to the High-Performance Computing (HPC) clusters, relying on the expertise of HPC specialists coming from four different European countries. At the course we will use SSH or VNC remote connection to HPC clusters hosted in IT4Innovations, VSB-TUO. For further information please visit https://sctrain.eu/course/big-data-and-ai/.

Acknowledgement: The SCtrain project has been funded with support from the European Commission.

The Slovenian National Competence Center for HPC – EuroCC@SLING

Pavel Tomšič[a] **and Damjan Harish**[b]

[a] *Faculty of Mechanical Engineering, University of Ljubljana, Slovenia*
[b] *ARNES, The Academic and Research Network of Slovenia, Slovenia*

The EuroCC 2 project (National Competence Centers within EuroHPC – phase 2) [1] is a continuation of the EuroCC project, which ended at the end of 2022. The NCCs liaise with local communities, in particular SMEs, map HPC competencies and facilitate access to European HPC resources for users from the private and public sector. NCCs deliver training, interact with industry, support the adoption of HPC services and many other things. The Slovenian National Competence Center for HPC - EuroCC@SLING [2] operates within the framework of the project. It consists of a consortium of five SLING partners, led by ARNES. Other partners in the project are Arctur d. o. o., "Jožef Stefan" Institute, University of Ljubljana and University of Maribor.

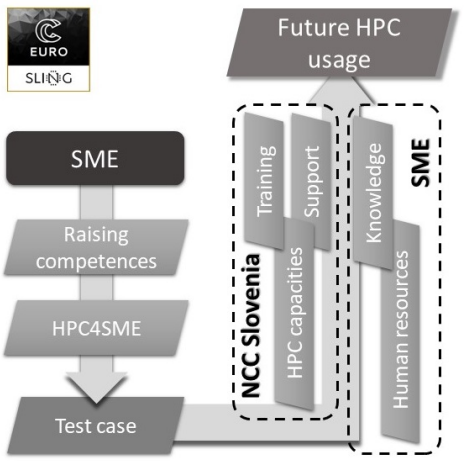
Raising competences

The three-tier workshop system (basic-intermediate-advanced level) aims to support users in adopting HPC resources and strengthening competences. New topics are constantly developed and published on NCCs indico portal [3] by the consortium partners and through international collaboration with other NCCs.
The division of workshops into different levels, according to the knowledge transfer to be obtaine, is set as:

- Basic level; focus on HPC literacy and general HPC-oriented topics
- Advanced level; focus on the STEM area
- Expert level; prior knowledge in HPC, raising competences and competitiveness

For industry

Offering a service portfolio and support for SMEs as well as large industries. Incorporating users and their use-cases which benefits from HPC and related technologies and help them with applications for national and EU funding opportunities.
The workflow and service portfolio includes technical/research expertise and industry consulting services to provide know-how, adequate training and access to existing HPC resources. NCC offers to industrial partners further development, implementation and validation of their problems. NCC Slovenia is ready to provide SMEs support for applying to open calls, applying for the access to the infrastructure and supporting them with consulting. We are also providing infrastructure and support for evaluation of different software and hardware technologies, proof-of-concept, testing and learning activities.



The Automated Assessment Tool HPC4SME [4] has been developed which allows industrial end-users to conduct a self-assessment to determine if they have the potential to use HPC technology.

References

[1] EuroCC access, https://www.eurocc-access.eu/

[2] SLING website, https://www.sling.si/en/

[3] NCC Slovenia indico portal, https://indico.ijs.si/category/29/

[4] HPC4SME tooll, https://www.hpc4sme.eu/

Croatian Competence Centre for HPC - HR HPC CC

Emir Imamagić, Jurica Špoljar, *Martin Belavić*, and Jasna Račić

University Computing Centre, University of Zagreb, Croatia

HR HPC CC was established under the National Competence Centres in the Framework of EuroHPC (EuroCC) project as part of a network of national competence centres for HPC in member states of the European High-Performance Computing Joint Undertaking (EuroHPC JU). Since the beginning of 2023. HR HPC CC is working as part of EuroCC 2 project.

HR HPC CC currently consists of the following Croatian institutions:

- University Computing Centre, University of Zagreb (SRCE)
- Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, Josip Juraj Strossmayer University of Osijek, (FERIT)
- Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split (FESB)
- Rudjer Bošković Institute (RBI)
- Faculty of Engineering, University of Rijeka (RITEH).

HRVATSKI CENTAR KOMPETENCIJA ZA HPC

Fig. 1: HPC CC logo.

SRCE is leading a consortium of Croatian institutions.

The main goal of HR HPC CC is to provide users from scientific and higher education communities, industry and public administration access to high performance computing competences, services and infrastructure. Namely, HR HPC CC provides the following services:

- Creating educational materials in the area odf High Performance Computing, High Performance Data Analytics (HPDA) and Artificial Inteligence (AI).
- Providing information on available courses and workshops in Croatia and European Union.
- Providing assistance in selecting computer resources and software.
- Supporting users in accessing HPC resources in Croatia and the European Union.
- Supporting users in using HPC resoureces.
- Supporting users in drafting project subsidy applications, technology transfer and operational development.
- Connecting users with professionals with competencies in various HPC areas.

In this talk we will present our experiences and successes so far and what we plan to do in future.

References

[1] https://www.hpc-cc.hr

Research Data Infrastructure Landscape at TU Wien

Martin Weise[a,b], Tomasz Miksa[a,b], Tobias Grantner[a], Josef Taha[c], Max Moser[c], Sotirios Tsepelakis[c], Barbara Sanchez-Solis[b], and *Andreas Rauber*[a]

[a] *Research Unit Data Science, TU Wien, Austria*
[b] *Center for Research Data Management, TU Wien, Austria*
[c] *TU.it, TU Wien, Austria*

In the era of big data, research has become increasingly data-driven, with vast amounts of information being generated and analyzed to produce new insights and discoveries. This data deluge requires a combination of methods and technologies to store, process, share and preserve research data. The TU Wien research data infrastructure landscape enables data citation as well as collection of provenance and allows research data to be findable, accessible, interoperable and reusable (FAIR).

The intellectual property of TU Wien is very diverse. Examples include the "Sentinel-1 Global Backscatter Model" datacube [1] that gives a high-quality impression on surface- structures and -patterns of the Earth, a paper self-archive for an "Open-Source River Basin Management System" needed for the yearly intellectual capital statement and Jupyter Notebooks of a recent training event of the Vienna Scientific Cluster. Since there is no single solution that fits our needs of storing, processing, sharing, and preserving this collected research data, a landscape of repositories (c.f. Fig. 1) based on open-source software was, where available, selected, or developed.

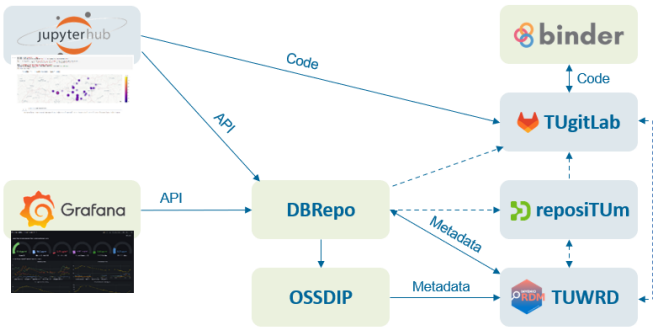


Fig. 1: Research data infrastructure at TU Wien. Nodes in green are currently in development.

File-based research data is managed by the research data repository TUWRD[1], text documents such as publications in the publication repository reposiTUm[2] and analysis steps, such as algorithms and software code should be deposited at the code repository TUgitLab[3] (and not in the research data repository although it is technically possible). These repositories are complemented by a compute platform (JupyterHub), forming a virtual research environment that allows seamless deployment of e.g., Jupyter Notebooks stored in TUgitLab via Binder, which in turn can read data files from TUWRD and/or DBRepo [3], perform some analysis and store results again directly into these repositories. All artifacts can be linked via persistent identifiers (PIDs, e.g. DOI) supporting provenance tracing, citation, and reproducibility of experiments. This virtual research environment (JupyterHub) and visualization tools (e.g., Grafana) can be conveniently accessed via a web browser. The key is to connect each of these solutions and offer seamless integration for the user to support the full lifecycle of research data while increasing external visibility of the data (for sensitive data we are currently developing a secure data infrastructure blueprint [2]).

References

[1] Bernhard Bauer-Marschallinger, Senmao Cao, et al. The Sentinel-1 Global Backscatter Model (S1GBM) - Mapping Earth's Land Surface with C-Band Microwaves, 2021. doi:10.48436/n2d1v-gqb91

[2] Martin Weise, Filip Kovacevic, Nikolas Popper, et al. OSSDIP: Open Source Secure Data Infrastructure and Processes Supporting Data Visiting. Data Science Journal, 21. doi:10.5334/dsj-2022-004

[3] Martin Weise, Moritz Staudinger, et al. DBRepo: a Semantic Digital Repository for Relational Databases. International Journal of Digital Curation, 17, 2022. doi:10.2218/ijdc.v17i1.825

[1] TUWRD is based on InvenioRDM. [Online]. https://researchdata.tuwien.ac.at/
[2] reposiTUm is based on DSpace. [Online]. https://repositum.tuwien.at/
[3] TUgitLab is based on GitLab. [Online]. https://gitlab.tuwien.ac.at/

Data Storage and Analysis Challenges on HPC Systems in Genomic Medicine

Blaž Vrhovšek[a,b], **Jernej Kovač**[a,b], **Tine Tesovnik**[a], **Robert Šket**[a], **and Barbara Jenko Bizjan**[a,b]

[a] *Clinical Institute of Special Laboratory Diagnostics, University Children's Hospital, UMC Ljubljana, Slovenia*
[b] *Faculty of Medicine, University of Ljubljana, Slovenia*

Introduction: Advances in sequencing technologies have opened up new possibilities in medical diagnostics. With the increasing amount of data generated, data analysis and data management have become the main bottleneck in diagnostic procedures. In clinical practice, accurate diagnosis is crucial, and analysis speed is of utmost importance. Furthermore, privacy and security measures are necessary to protect sensitive data. In this abstract, we highlight some of the most common problems.

Data Transfer and Analysis Challenges: Sequencing machines produce vast amounts of data that must be transmitted to the high-performance computing (HPC) system for fast and accurate analysis. Analysis of genomic data involves multiple steps, each requiring a unique set of bioinformatics tools supported by specific dependencies. Various diagnostic pipelines, such as short and long read analysis, single nucleotide and structural variant calling, bulk RNA or single-cell RNA expression or fusion analysis, must follow common pipeline guidelines and result in a standardized data storage structure. Since GPUs are enabling significant improvement in speed of analysis, it is preferable to use bioinformatics tools with GPU acceleration, as this can significantly reduce the time to the final clinical report. Specific analysis and data transfer must be enabled for the user, and extensive bookkeeping of access, pipelines, and tool versions must be maintained for reproducibility of analysis. Finally, the stored data must support periodic re-analysis using newer tools or versions [1, 2].

Data Storage Challenges: The vast amount of data generated by sequencing platforms requires massive storage capacity, which can strain the storage infrastructure of an HPC system. To reduce storage requirements, genomic data is often stored in compressed file formats, which can lead to additional computational overhead during data transfer and processing. Additionally, the high-throughput nature of genomic data can cause a significant input/output (I/O) bottleneck during data transfer and processing. This can be particularly challenging when working with shared storage systems, where multiple users may access the same data simultaneously. Therefore, efficient data management and processing requires optimizing storage and I/O operations [3].

Conclusion: In conclusion, genomic data storage and analysis present significant challenges for HPC systems. Successful and fast analysis requires efficient data management, optimization of storage and I/O operations, and the development of specialized analysis tools and pipelines. As genomic medicine research continues to generate increasingly large and complex datasets, addressing these challenges becomes even more crucial for efficient and effective analysis.

References

[1] Kulkarni P., Frommolt P., Comp. and Str. Biotech. J. **15**, 471 (2017).

[2] Tanjo T., Kawai Y., Tokunaga K., Ogasawara O., Nagasaki M., J. Hum. Genet. **66**, 39 (2021)

[3] Papageorgiou L., Eleni P., Raftopoulou S., Mantaiou M., Megalooikonomou V., Vlachakis D., EMBnet J. **24**, e910 (2018).

Python, Conda, and Virtual Environments on the Vienna Scientific Cluster

Katrin Muck

VSC Research Center, TU Wien, Austria

Traditionally HPC centers like the Vienna Scientific Cluster (VSC) were mainly focused on supporting users with highly optimized distributed code written in compiled languages such as Fortran and C++ using batch processing together with e.g. MPI. But in the last decade Python has gained immensely in popularity in the scientific community. Not only because of the ease of use of the language itself but also because of tools such as *Conda* that help the users to setup complicated environments from scratch without major hassles. Many researchers have since adopted Python for the use in HPC environments as well. Apart from the obvious benefits this trend also created a couple of problems for cluster maintainers. Earlier the VSC provided certain Python packages and matching Python libraries via *Spack* and didn't suggest guidelines or provide configuration for using Conda. However, this approach is hard to maintain since every Python package needs to be installed manually via spack, multiple versions of each package for multiple Python versions need to be provided and so on. Some packages are especially hard to come by since they are not provided out of the box by spack. Furthermore this approach doesn't help people that were already using Conda. The problem with the latter being, that users have lots of copies of the same software packages in their home directories, sometimes even accross the same project. In this short talk I want to address the issues we identified so far and provide solutions to improve the situation at the VSC.

Fig. 1: Python, Conda, Micromamba illustrations.

Computer Aided Engineering in Augmented Reality: Flow Visualizations for Hydro Power Applications

Diego Dalla Costa[a], **Franz Josef Johann Hahn**[b], **and Irene Reichl**[a]

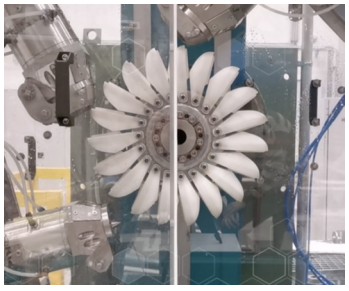
[a] *VSC Research Center, TU Wien, Austria*
[b] *Research Unit of Fluid Flow Machinery, Institute of Energy Systems and Thermodynamics, TU Wien, Austria*

Augmented Reality (AR) technology has attracted a great deal of attention in a number of fields, including engineering. AR provides a unique opportunity to diversify the traditional methods implemented in engineering by overlaying digital information onto the physical world. Displaying CAD geometries in AR has already become a standard feature in building information modeling (BIM) or Industry 4.0. The visualization of Computer Aided Engineering (CAE) results in virtual, augmented, or mixed reality (VR, AR, XR) is not yet standard in most of the common engineering software packages. Our current work is therefore devoted to the development of an interface between multi-physics software and AR, where there is none yet.

The present contribution demonstrates that procedure for an example from Computational Fluid Dynamics (CFD), namely Pelton turbines. They convert the potential energy of water stored at high-altitude reservoirs into mechanical power. At the end of the penstock, which guides the water from the reservoir to the turbine, nozzles generate high-velocity water jets that act on the buckets of the Pelton turbine runner. When passing through the water jets, the buckets are admitted to the flow impulse and participate in the energy transfer. In consequence, CFD simulations of the runner need to consider the unsteady character of the jet-bucket interaction. This, together with the water-air multiphase fluid mixture, poses an inherent challenge for simulation and visualisation of the flow. However, by employing advanced techniques such as CFD in the early design phase, hydraulic engineers are given a very helpful tool to calculate the efficiency of a Pelton turbine runner and to study how a change of runner parameters affects the efficiency of the runner. Morevover, the number of costly experiments can be greatly reduced in the design and development phase of a turbine.

The simulation itself was carried out on High-Performance Computing (HPC) resources of the *Vienna Scientific Cluster (VSC)*, including a virtual desktop interface set up by the remote access software *NoMachine* for direct pre- and post-processing on cluster nodes. Aiming for a light-weight AR application (AR-App) on a tablet or mobile phone, the 3D Geometry and selected representations of the data (e.g., like in Fig. 1, an isosurface of one quantity colored by the value of another quantity) were copied to a local machine in order to build an Android AR-App by means of the game engine *Unity*.

For validation purposes, in addition to runner simulations executed in *ANSYS CFX 19.2*, efficiency measurements in the laboratory of Fluid Flow Machinery were conducted. This allowed us, to augment the *real world* experimental setup with simulation data, i.e., the time evolution of the water jet hitting the rotating runner, compare Fig. 1. As soon as the AR-App recognizes the real target, the projected AR object can be rotated and translated via touch screen. Due to the modest device requirements, the presented AR-App is well suited to be used on demand in presentations, to enrich the general understanding of the audience for a multi-physics topic. Further, the application is suited for academic teaching and research, as well as for public communication. Where the present approach requires to select data representations beforehand, our future investigations will explore the direct interaction of the XR engine with the data.



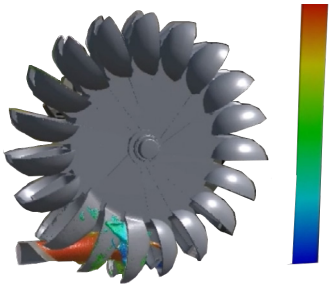


Fig. 1: Left: Test setup in the hydraulic laboratory of IET. Right: Isosurface of water coloured by the normalized flow velocity for one time step. The AR app visualizes the time series of the transient calculation results and provides the user a 3D view of the AR object, that can be zoomed and rotated.

Real-time human pose estimation using YOLOv7 on HPC

Leon Bogdanović **and Leon Kos**

LECAD laboratory, Faculty of Mechanical Engineering, University of Ljubljana, Slovenia

In the last decade, human pose estimation in videos has become more and more accurate with the use of Deep Learning models. These models are usually implemented in two stages: person detection and key-point localization. However, their complexity limits real-time performance, since they are computationally expensive. Hence, these models are usually much faster on GPUs than on CPUs. The YOLOv7 framework [1], released last year, is the first in the "You only look once" (YOLO) family of real-time object detection frameworks to include a human pose estimation model. It is implemented in PyTorch, a Python machine learning framework, what makes it highly customizable and easy to use.

HPC clusters, with hardware dedicated primarily to intensive computations, can be used to train deep learning models, but can also be used for real-time visualization of object detection, e.g., through remote virtual desktops that allow persistent and graphics desktop connection with installation of remote access clients that compress the hardware accelerated (VirtualGL) graphical desktops over internet. The HPCFS cluster at the University of Ljubljana uses NoMachine remote virtual desktop to achieve this. Real-time human pose estimation tests with YOLOv7 were performed on HPCFS's gpu02 login node equipped with an NVIDIA A100 GPU. For that purpose, a Python virtual environment with PyTorch libraries was created.

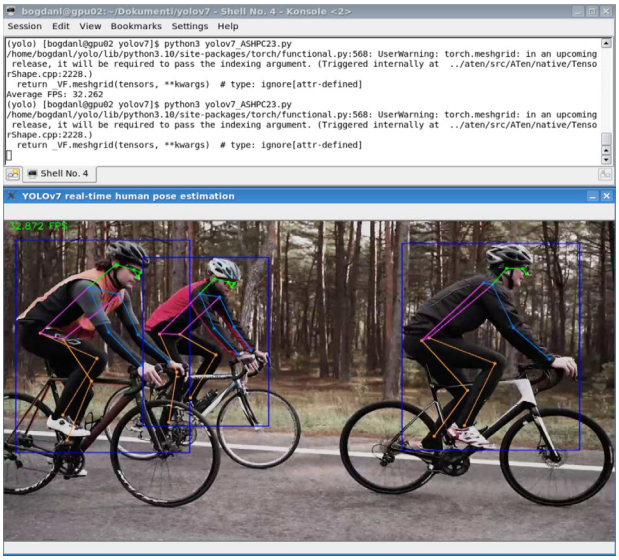


Fig. 1: Real-time human pose estimation on HPC.

Since the default CUDA version (11.5) in the NVHPC/21.11 module used does not work with the default PyTorch libraries on A100 GPU, a specific combination of PyTorch libraries with CUDA 11.3 support, i.e., `torch==1.11.0+cu113`, `torchvision==0.12.0+cu113`, `torchaudio==0.11.0`, were installed. A script that uses a pre-trained keypoint detection model `yolov7-w6-pose.pth` [2] was written to perform real-time human pose estimation on a full HD video (Fig. 1). Average rates of about 33 FPS were achieved in real-time. The script also writes the results of human pose estimation of every frame to an output video in `*.mp4` format, which can be reviewed in a video program [3]. This contribution demonstrates that HPC clusters can be used for computationally intensive jobs, such as real-time human pose estimation in videos, that also require on-the-fly graphical visualization. On login nodes such jobs should be carefully executed since they can affect other users of the cluster.

References

[1] C.-Y. Wang, A. Bochkovskiy, H.-Y. Mark Liao, YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, arXiv, 2207.02696 (2022).

[2] Official YOLOv7, https://github.com/WongKinYiu/yolov7/, 2022.

[3] YOLOv7 on HPC, https://bitbucket.org/lecad-peg/hpc-yolov7/src/master/, 2023.

Cryo-EM Software Packages: A Sys-admin's Point of View

Stefano Elefante, **Stephan Stadlbauer, Michael Alexander, and Alois Schlögl**

Institute of Science and Technology Austria, Austria

Cryo-electron microscopy (CryoEM) is a technique to determine the high-resolution structure of biomolecules in solution [1]. The method is based on the transmission electron cryomicroscopy (CryoTEM) that is also usually denoted as CryoEM. At the authors' institute (ISTA), the electron microscopy facility is equipped with two CryoEM devices: 200kV Cryo-TEM Glacios (Fig. 1) and 300kV Cryo-TEM Titan Krios G3i (Fig. 2). CryoEM samples can preserve cryogenic conditions up to 72 hours within the Krios Cryo-TEM and 24 hours within the Glacios Cryo-TEM system. These features allow researchers to collect all data that are necessary to achieve high-quality biomolecules reconstruction. However, the amount of cryoEM data required to obtain high-resolution analysis can be very large. As a consequence, the computational load can be very high requiring the use of high-performance computing (HPC) ressources. Several software packages are available to process cryo-EM data: general purpose packages such as Relion [2], WarpEM [3], and CryoSPARC [4] and also software to address more specialized tasks like crYOLO, TOPAZ, etc.

Relion v3.0-v4.0 includes an end-to-end processing of raw cryo-EM data based on an empirical Bayesian approach. **WarpEM** v1.0.9 has been developed for the MS-Windows platform and has on-the-fly CryoEM data processing capabilities. **CryoSPARC** v3.0-v4.2.1 is a proprietary software platform for cryoEM data processing and analysis. **crYOLO** v1.9.3 is based on an algorithm for particle picking which relies on a convolution neural network. **TOPAZ** v0.2.5 is a pipeline for particle picking in cryoEM micrographs which employs neural networks.

Fig. 1: ISTA 200kV Cryo-TEM Glacios.

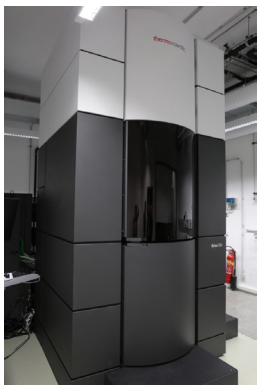


Fig. 2: ISTA 300kV Cryo-TEM Titan Krios G3i

References

[1] The development of Cryo-Electron microscopy, Scientific Background on the Nobel Prize in Chemistry 2017. https://www.nobelprize.org/uploads/2018/06/advanced-chemistryprize2017-1.pdf

[2] Sjors H.W. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure determination, Journal of Structural Biology, Volume 180, Issue 3, December 2012, Pages 519-530

[3] Tegunov, D., Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. Nat Methods 16, 1146–1152 (2019). https://doi.org/10.1038

[4] Punjani, A., Rubinstein, J., Fleet, D. et al. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods 14, 290–296 (2017). https://doi.org/10.1038/nmeth.4169

Inverse Molecular Docking Approaches on HPC Systems

Marko Jukič[a,b], **Vid Ravnik**[a], **and Urban Bren**[a,b,c]

[a] *Faculty of Chemistry and Chemical Engineering, University of Maribor, Slovenia*
[b] *Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Slovenia*
[c] *Institute for Environmental Protection and Sensors, Slovenia*

The method of inverse molecular docking, or target fishing, is one of the historically explored applications of molecular docking. Indeed, in silico target prediction for bioactive compounds is a promising field and could help in the development of selective drugs and the prediction of off-target effects. However, the classical approach of inverse methodology is not without problems, as interpretation of results, assessment of biological context, and in vitro confirmation of hits on a pleiade of targets remain challenging issues. To move away from single-point literature confirmations in the resulting inverse docking matrices, we developed an inverse fingerprinting approach that can build on the methodological strengths mentioned above. Namely, we postulate, we can leverage the imposed binding modes of a number of proteins to our query molecules as selectors for potential activity. In this way, inverse docking fingerprints are generated for query molecules on a set of targets under investigation. Subsequently, the space where the target molecules meet can be examined and compared in silico, focusing on complete fingerprint profiles rather than individual systems for single point validations. This computationally intensive approach is particularly relevant with the availability of high-performance computing (HPC) systems that can effectively advance drug discovery. Timely computations of complete libraries of query compounds against complete libraries of targets are possible, and receptor/ligand binding affinities can be assessed in an inverse fingerprinting manner. In this work, we present HPC Vega-assisted classical inverse docking approaches and compare them to fingerprinting experiments to highlight their strengths in identifying active molecules [1,2].

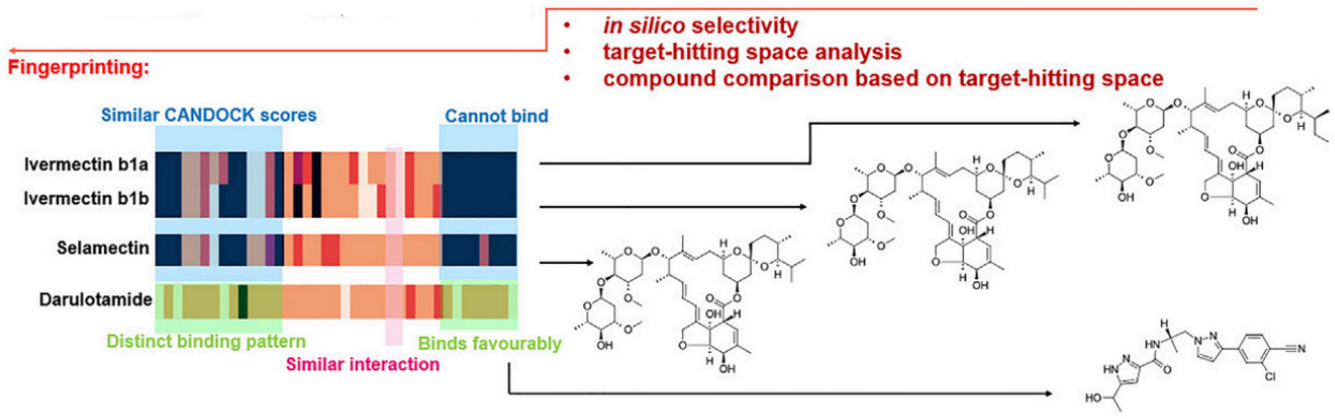


Fig. 1: Inverse fingerprinting approach using HPC.

References

[1] Jukič, M., Kores, K., Janežič, D., & Bren, U. (2021). Repurposing of drugs for SARS-CoV-2 using inverse docking fingerprints. Frontiers in Chemistry, 1059.

[2] Lešnik, S., Jukič, M., & Bren, U. (2023). Mechanistic Insights of Polyphenolic Compounds from Rosemary Bound to Their Protein Targets Obtained by Molecular Dynamics Simulations and Free-Energy Calculations. Foods, 12(2), 408.

Open boundary molecular dynamics with GPU parallel computation

Andraž Filipčič[a], **Matej Praprotnik**[b], **and Uroš Lotrič**[c]

[a] *Jožef Stefan Institute, Slovenia;* [b] *National Institute of Chemistry, Slovenia*
[c] *Faculty of Computer and Information Science, University of Ljubljana, Slovenia*

Molecular dynamics is a computer simulation method for analysing the movement of atoms and molecules in an area. It does so by incrementing through time with a set time step and individually moving all the particles within the simulation box according to their velocity, then computing interatomic potential between the particles and adjusting their velocities accordingly.

Most use cases implement a periodic boundary condition, in which the simulation box borders copies of itself (periodic images) on each of its sides. As such, a particle that escapes the boundaries is wrapped back into the simulation box, as pictured by the blue particles in Figure 1. Another possible implementation is using an open boundary condition, in which particles are free to leave the simulation box, and are not replaced by copies of the particles in bordering periodic images.

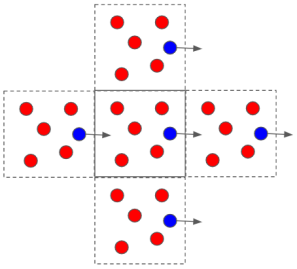


Fig. 1: Periodic boundary conditions.

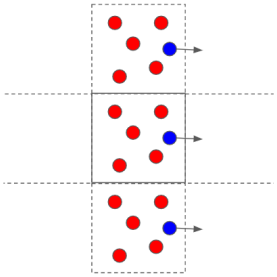


Fig. 2: Open and periodic boundary conditions.

Our implementation uses open boundary conditions on the left and right borders of the primary simulation box (the x-axis), and periodic boundary conditions on the other borders, as pictured in Figure 2. The particles that escape the primary simulation box are not immediately reinserted in any way, but when enough of them escape, new particles are inserted near the open border so that the total momentum and mass of the system are conserved [1].

With an existing implementation of our boundary conditions in a CPU-driven software (ESPResSo++, [2]), our goal was to reproduce this implementation in software which makes use of hardware acceleration with the help of CUDA (Mirheo, [3]). Molecular dynamics simulations are a great candidate for GPU parallelisation, as each of the several thousand particles' trajectories has to be calculated individually.

The first parallelisation is performed by splitting up the primary simulation box into multiple sub-boxes and assigning each to a CPU thread, and communication between compute threads is implemented with MPI. The second parallelisation is performed during the integration part of each time step of the simulation, where each individual particle's trajectory is calculated, and each particle is assigned its own GPU thread, which allows for far greater parallelisation efficiency than using only CPU threads.

The main limit of CPU parallelisation is the splitting of the primary simulation box into sub-boxes, as splitting it also increases the amount of borders between the sub-boxes, which results in more communication between CPU threads as particles pass between them. This is a significant barrier when using only CPU parallelisation, as the efficiency of sub-box parallelisation greatly drops when trying to split into more than 12 sub-boxes, the total time to run simulations even starts to increase. But with the assistance of GPU parallelisation, the calculation of particle trajectories is not limited by the amount of CPU threads participating, as it can simply be offloaded to the GPU.

References

[1] Delgado-Buscalioni, R., Sablić, J., and Praprotnik, M. Open boundary molecular dynamics. Eur. Phys. J. Spec. Top. 224, 2331–2349 (2015). https://doi.org/10.1140/epjst/e2015-02415-x

[2] http://www.espresso-pp.de/

[3] https://mirheo.readthedocs.io/

Rand_gpu – A C/C++ library for generating pseudo-random numbers on the GPU

Jakob Kenda, Uroš Lotrič, and *Davor Sluga*

Faculty of Computer and Information Science, University of Ljubljana, Slovenia

Pseudo-random numbers (hereafter *random numbers*) are widely used in scientific applications. They are provided by random number generators – RNGs. The generation of random numbers can represent a considerable fraction of the execution time of an application. If the time spent on generation could be somehow reduced, the overall performance could be increased.

One way to achieve faster execution times is by developing new algorithms, but this approach has drawbacks – faster algorithms tend to be less robust and have inferior statistical properties.

Another approach is to use accelerators, such as graphics processing units (GPUs), to speed up random number generation. However, adapting existing applications to use GPUs can be challenging because it requires knowledge of programming GPUs, significant changes to the codebase, and often results in suboptimal solutions. Additionally, support for different hardware vendors can complicate implementation.

To address these challenges, we used an existing codebase of RNGs in OpenCL [1] to create a cross-platform C/C++ library that uses GPUs to generate random numbers and offers the user a familiar interface. The GPU can generate an array of numbers in advance and transfer them to the system memory for the CPU to access them when needed. We can achieve this because the sequence of random numbers is deterministic given the same seed.

In Fig. 1 we can see the speedup relative to the `rand()` function from C's stdlib achieved on different platforms we have tested the library on. We measured 15 random number generation algorithms we implemented in our library. These measurements represent the minimum, average, and maximum speedups achieved. As we can see, there is considerable variation in the speedups achieved on different hardware due to the different computational requirements of the various algorithms and the capabilities of different GPUs. Overall, we reduced execution time by a factor of 3-5.

Our library interface was designed to be similar to the C++ standard library making it convenient to use in existing software. It offers a wide variety of RNGs to meet the requirements of different applications. Although any of the 15 implemented algorithms can be chosen, Tyche was designed for parallel use from the get-go [2] and gives one of the best performances while having good statistical properties.

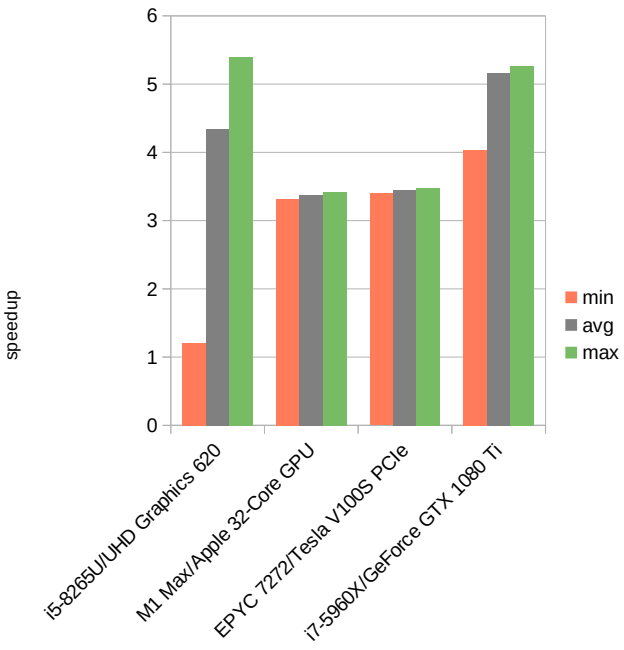


Fig. 1: Speedup on different platforms.

References

[1] Ciglarič, T., Češnovar, R., and Štrumbelj, E. An OpenCL library for parallel random number generators. The Journal of Supercomputing. **75.7**, 3866-3881 (2019).

[2] Neves, S. and Araujo, F. Fast and Small Nonlinear Pseudorandom Number Generators for Computer Simulation. International Conference on Parallel Processing and Applied Mathematics, 92 (2011).

Computing in DAPHNE: Integrated Data Analysis Pipelines for Large-Scale Data Management, HPC and Machine Learning

Aleš Zamuda

Faculty of Electrical Engineering and Computer Science, University of Maribor, Slovenia

This contribution highlights latest developments in the DAPHNE (Integrated Data Analysis Pipelines for Large-Scale Data Management, HPC and Machine Learning) project, https://daphne-eu.eu. The project is focused on High-Performance Computing (HPC) and is publishing code at https://github.com/daphne-eu/daphne under the open source Apache License (version 2.0). The project is developing a Domain-Specific Language (DSL) for HPC. This DSL is capable of integrating different pipelines for data analysis and supports large-scale data management. Early description and results of DAPHNE were presented in [1] and now these are being developed further to support more functionalities, capabilities, and use cases. The remaining of the text below hence reports on possible further steps in the project, especially some computing use cases.

Among other, some potential use cases of the DSL are Machine Learning (ML) and Genetic and Evolutionary Computation (GECCO) [3]. Namely, an HPC-applicable code that is integrated as a platform with data reproducibility, support for experimentation speedup through integrated data pipelines for modular integration of algorithms and benchmarking frameworks, could, through new analysis capabilities, offer further advances in ML and GECCO benchmarking. With integration of new hardware architectures, cloud computing, supercomputers, and software [1], such benchmarking would offer new capabilities and hence advance such ML and GECCO use cases.

Fig. 1: DAPHNE.

With these capabilities and the existing body of work in the DAPHNE architecture [1] and GECCO benchmarks (like IEEE CEC and ACM GECCO functions implemented in COCO, IOHprofiler, etc. [2]), here are some examples of suggested future work to be done for supporting computing in DAPHNE for GECCO (GECCO.DAPHNE):

- execution of modern high-performance multi-node workloads of GECCO optimizers, fitness functions, and benchmarking [2]; with sparse data representations support, especially in the context of ML,
- integrated pipeline code support for distributed and heterogeneous GECCO workloads on GPUs and FPGAs, including microbenchmarks and use cases like ejector optimization and automotive vehicle development [1], and
- supporting specific accelerated operations on data ranging from floating-point to text [2], as dedicated Multi-Level Intermediate Representation (MLIR, https://mlir.llvm.org/) kernels and functions.

Acknowledgement: this work is supported by EU project no. 957407 and ARRS programme P2-0041.

References

[1] P. Damme, M. Birkenbach, C. Bitsakos, M. Boehm, P. Bonnet, F. Ciorba, M. Dokter, P. Dowgiallo, A. Eleliemy, C. Faerber, G. Goumas, D. Habich, N. Hedam, M. Hofer, W. Huang, K. Innerebner, V. Karakostas, R. Kern, T. Kosar, A. Krause, D. Krems, A. Laber, W. Lehner, E. Mier, M. Paradies, B. Peischl, G. Poerwawinata, S. Psomadakis, T. Rabl, P. Ratuszniak, P. Silva, N. Skuppin, A. Starzacher, B. Steinwender, I. Tolovski, P. Tözün, W. Ulatowski, Y. Wang, I. Wrosz, A. Zamuda, C. Zhang, and X. Xiang Zhu. 12th Conference on Innovative Data Systems Research, 12 pgs. (2022).

[2] A. Zamuda, M. Nicolau, and C. Zarges. Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO), pgs. 1777–1782 (2018).

[3] J. E. Fieldsend (Ed.). GECCO '22: Proceedings of the Genetic and Evolutionary Computation Conference (Boston, Massachusetts). Association for Computing Machinery, New York, NY, USA (2022).

SPACE - Scalable Parallel Astrophysical Codes for Exascale Center of Excelence

Lubomír Říha

IT4Innovations, VŠB - Technical University of Ostrava, Czech Republic

In astrophysics and cosmology (A&C), HPC (high-performance computing)-based numerical simulations are invaluable instruments to support scientific discovery. Given the complexities of the problems, they represent essential tools for modeling, interpreting, and understanding the physical processes behind the observed sky.

Advances in computational power promise a wealth of groundbreaking, new scientific discoveries by making ever greater numerical simulations feasible, provided that equally advanced tools are created to exploit these computational resources. For a field like astrophysics, which is generally unsuited to laboratory experiments, the significance of this development cannot be overestimated.

Future exascale computing systems are expected to have extremely complex, heterogeneous architectures. The currently used numerical simulation codes are unsuitable for use on these systems since they were not purposely designed for them and, therefore cannot effectively take advantage of the superior processing capabilities promised.

SPACE's main goal, therefore, is to enable current astrophysical and cosmological codes (Open GADGET, CHANGA/GASOLINE, PLUTO, iPic3D, RAMSES, WhiskyTHC, FIL, and BHAC) to be used on the pre-exascale HPC architectures funded by the EuroHPC JU and made available at the end of 2022, as well as on future architectures by re-designing or adapting the existing computational tools for this next-generation hardware. It will bring together scientists, community code developers, HPC experts, hardware manufacturers, and software developers in co-design activities to re-engineer eight of the most widely used European A&C HPC codes into new products that can efficiently exploit future computing architectures. These eight A&C HPC codes represent 70% of the HPC A&C simulations and were selected after an extensive analysis of their features and capabilities. They will initially be prepared to adequately exploit the pre-exascale systems with a view to their transition to exascale systems and beyond. At the same time, SPACE will work to advance workflows and data processing based on machine learning and visualization applications and to enhance their exascale capabilities. Furthermore, the SPACE Centre of Excellence (CoE) will promote the adoption of general and community standards for data products based on FAIR principles and the interoperability of data and applications based on the technology standards and best practices of the International Virtual Observatory Alliance (IVOA). SPACE will also implement the selected applications and foster their use by means of a specific outreach and training program aimed at creating a broad and skilled talent pool in Europe to boost the use of high-performance and high-throughput solutions in academia in order to pave the way for the transition to exascale technologies and beyond.

Acknowledgement: SPACE CoE is funded by the European Union. It has received funding from the European High Performance Computing Joint Undertaking and from Belgium, the Czech Republic, France, Germany, Greece, Italy, Norway, and Spain under grant agreement No. 101093441.

References

[1] SPACE CoE Project website: www.space-coe.eu (2023).

Accelerating Astrophysical Code Using CUDA Fortran

James McKevitt **and Eduard Vorobyov**

Department of Astrophysics, University of Vienna, Austria

Astrophysical simulations often require the calculation of gravitational potential across a self-gravitating domain, something made more efficient through the use of nested meshes. This work represents a novel, fast, and accurate method for computing gravitational potentials in such cases which is highly parallelisable and applicable to graphics processing units (GPUs) for further acceleration. Fortran remains a popular programming language in the astrophysical community with many codes written in the language and as such, we implement our method using CUDA Fortran, sharing the familiar standard Fortran syntax. CUDA Fortran offers an efficient way to code Nvidia GPUs which is easily compatible with these astrophysical codes, requires minimal additional syntax, and does not require the use of the C language.

Leveraging the convolution theorem, we compute the gravitational potential using its integral form, avoiding the iterative approaches typically employed. Comparative studies against traditional iterative methods demonstrate not only comparable accuracy but also a markedly improved runtime performance - a factor of 200 times faster when deployed on GPUs.

The convolution method's high computational speed, coupled with the relative ease of implementation with CUDA Fortran, makes it an appealing choice for astrophysical simulations involving many nested grids.

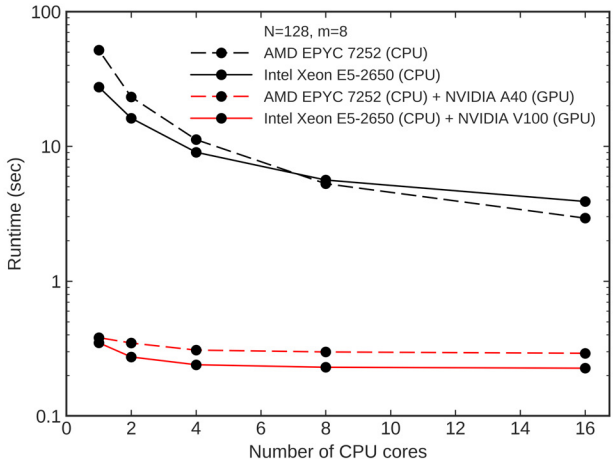


Fig. 1: Runtime vs. the number of CPU cores in computing the gravitational potential on 8 nested grids. [1]

Additionally, we demonstrate the use of CUDA Fortran alongside distributed memory parallelism using both the standard Message Passing Interface (MPI) and the native Fortran, syntax-friendly and easy-to-implement Coarray Fortran (CAF). We demonstrate a comparable performance and, in the case of CAF, a technique allowing the combination of the proprietary and high-speed CUDA Fortran offered by Nvidia and the high-speed CAF implementation offered by Intel.

Acknowledgement: This work was supported by FWF project I4311-N27. The simulations were performed on the VSC Vienna Scientific Cluster.

References

[1] Vorobyov, E.I., McKevitt, J.E. et. al, Astronomy and Astrophysics **671**, A81 (2023).

SIMPIC: A simplified particle-in-cell plasma physics code optimized for multiple GPUs

Ivona Vasileska[a], **Ezhilmathi Krishnasamy**[b], **Leon Kos**[a], **Pascal Bouvry**[b], **and Pavel Tomšič**[a]

[a] *Faculty of Mechanical engineering, University of Ljubljana, Slovenia*
[b] *Parallel Computing and Optimization Group, University of Luxembourg, Luxembourg*

Numerical simulations, especially those based on a kinetic plasma model, play a crucial role in the study of fusion devices with their highly nonlinear plasma behavior. The particle-in-cell (PIC) method is a well-established approach for kinetic plasma simulation codes. Simulations based on such codes compute the evolution of up to 10^{12} particles using 10^6 CPU cores and more, creating a significant workload on even the largest HPC clusters to solve the resulting multivariable systems of equations. Our research problem: Optimizing the efficiency and cost-effectiveness of an HPC system by using GPUs in PIC simulation scenarios.

We present an implementation of the Simple Particle-In-Cell (SIMPIC) code - derived from common PIC algorithms - for a multi-GPU environment and investigate the computational time required for simulations with different input parameters. The SIMPIC code addresses electrostatic plasmy physics scenarios, i.e. it neglects particle collisions and magnetic fields to compute only the evolution of electron (super)particles (representing bunches of real electrons), while ignoring the motion of the ions. Under these assumptions, the solution of Maxwell's equations is simplified to the calculation of a Poisson equation. Details of our hybrid CPU-GPU parallelized SIMPIC implementation can be found in [1]. We have analyzed the time needed to simulate 10^3 particles for different numbers of grid cells, running the simulations on either one or four (then Nvlink 2.0 connected) Nvidia V100 GPUs on the MARCONI100 HPC cluster [2], using the OpenACC model of SIMPIC. The results are shown in Figures 1 and 2.

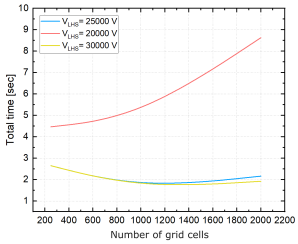


Fig. 1: Runtimes on a single GPU

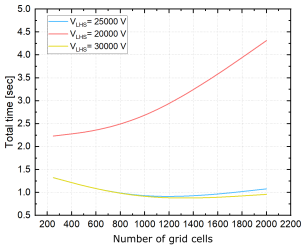


Fig. 2: Runtimes on four GPUs

Our results show that using four GPUs (see Fig. 2) - when simulating 10^3 particles with different numbers of grid cells - is twice as fast as using a single GPU (see Fig. 1) alone. However, communication between the GPUs adds a small overhead to the simulation. As the number of grid cells increases, the simulation time increases slightly. Note that the decrease in simulation time as the voltage applied to the particles is increased is simply related to the particle velocities. As the voltage increases, the particles move faster, resulting in a shorter overall simulation time.

Our conclusion: Properly adapted PIC codes have the potential to benefit from transfer to complex heterogeneous high-performance computing machines. Future work will address the incorporation of more sophisticated physics into the SIMPIC code and the transfer of simulations to larger machines with parallel architectures.

References

[1] E. Krishnasamy, I. Vasileska, L.Kos, P. Bouvry, Hybrid programming and multiple GPUs implementation for Particle-In-Cell, The 8th International Conference on Computer and Communication Systems, 2023

[2] https://www.hpc.cineca.it/hardware/marconi100

GPU-accelerated matrix-free solvers for the efficient solution of cardiac electrophysiology in `life`^{x}

Francesco Carlo Mantegazza[a], **Alfio Maria Quarteroni**[d], **Paola Gervasio**[b], **Pasquale Claudio Africa**[e], **Matteo Salvador**[c], **and Elias Karabelas**[a]

[a] *Universität Graz & BioTechMed-Graz, Austria;* [b] *Università di Brescia, Italy;* [c] *Stanford University, USA;*
[d] *École Polytechnique Fédérale de Lausanne, Switzerland & Politecnico di Milano, Italy*
[e] *Scuola Internazionale Superiore di Studi Avanzati di Trieste, Italy*

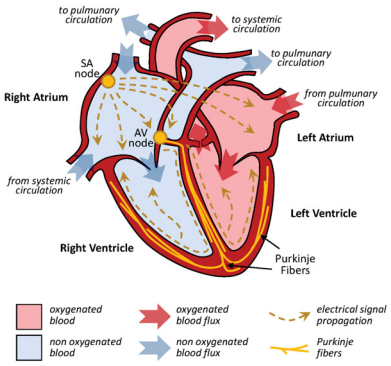


Fig. 1: Cardiac electrophysiology.

Computational cardiology is a growing field that enables to address physiological and pathological scenarios by exploiting suitable mathematical models and numerical methods. Cardiac electrophysiology can be modeled by the monodomain equation, a time-dependent partial differential equation (PDE) describing the space–time evolution of the transmembrane potential, coupled with proper ionic models, *i.e.*, systems of ordinary differential equations (ODEs) representing how chemical species move across ionic channels. The most challenging aspect is the requirement of high temporal and spatial resolution dictated by the solution to such problem, as it resembles a propagating wavefront with very steep gradients. The computational cost of solving the discretized problem is such that direct methods quickly become too expensive. We exploit modern hardware architectures to develop a GPU-accelerated matrix–free solver that can be readily employed for mixed CPU-based and GPU–based, massively parallel, large-scale electrophysiology simulations. We rely on matrix-free solvers and on the sum-factorization, two computational techniques that have proven to be effective ways to reduce the computational complexity. In this framework, we propose and implement two different time discretization schemes for the monodomain equation: the 2nd-order Backward Differentiation Formula (BDF2), resulting in a semi-implicit time scheme, and the 2nd-order Runge-Kutta scheme (RK2), giving an explicit scheme. While the former requires the solution of a linear system at each time-step, the latter can perform a DoF-by-DoF update, as the space discretization is based on Spectral Element Method with Numerical Integration (SEM-NI). Moreover, the RK2 scheme is theoretically much faster, but constrained by a severe stability condition on Δt [2]. We perform a thorough comparison among the different solvers under p-refinement, *i.e.*, by increasing the polynomial degree of the basis functions, in terms of computational cost and accuracy.

The mathematical models and the numerical methods presented in this thesis have been implemented in `life`^{x} (https://lifex.gitlab.io/) [3], a high performance `C++` library developed within the iHEART project and based on the `deal.II` (https://www.dealii.org) Finite Element core [1]. The architecture we have relied upon to run the numerical simulations is a node part of the cluster present at MOX, Dipartimento di Matematica, Politecnico di Milano. It is equipped with 2 CPUs Intel Xeon Gold 6238R (2.2GHz) and a GPU NVIDIA GV100GL - Tesla V100 PCIe 32GB (512 `CUDA` cores).

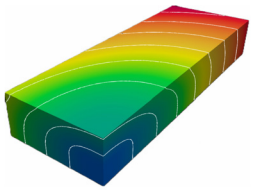


Fig. 2: Activation time.

References

[1] D. Arndt, W. Bangerth, M. Feder, M. Fehling, et al. *The* `deal.II` *Library, Version 9.4*, Journal of Numerical Mathematics, (2022).

[2] A.M. Quarteroni, A. Valli, *Numerical Approximation of Partial Differential Equations*, Springer Berlin Heidelberg, (1994).

[3] P.C. Africa, *life*^{x}*: A flexible, high performance library for the numerical solution of complex finite element problems*, SoftwareX, 20 (2022).

Massively Scaling Molecular Screening Workloads on EuroHPC Supercomputers

Sascha Hunold[a], **Ioannis Vardas**[a], **Gökhan Ibis**[b], **and Thierry Langer**[b,c]

[a] *Research Group for Parallel Computing, TU Wien, Austria*
[b] *Inte:Ligand Software Development and Consulting GmbH, Austria*
[c] *Department of Pharmaceutical Sciences, University of Vienna, Austria*

LigandScout is an *advanced molecular modeling and design software suite*, which aims at enhancing early-stage drug discovery efficiency by providing in-silico experiments that help reducing the number of expensive in-vitro and in-vivo assays. An essential method of LigandScout is *virtual screening*, where pharmacophore models are used for identifying hits from large compound databases. Due to the computational requirements of this screening process, LigandScout already provides parallelization strategies, which were primarily *designed for elastic computations* on dedicated HPC or Cloud resources [1].

Goal: In this work, we consider the problem of *executing* an adapted, parallel version of *LigandScout* on a *dedicated supercomputer*, in particular, the Vienna Scientific Cluster 5 (VSC-5). Running LigandScout on a multi-user supercomputer is a challenging task, which requires a careful redesign of the parallelization workflow. We address these challenges and propose several solution strategies.

Challenges: A fundamental question is how to efficiently perform application-level load balancing. In dedicated HPC or Cloud environments, the task of load balancing can be offloaded to the batch scheduler (e.g., SLURM), i.e., the application sends computational tasks as single jobs to the batch scheduler, which takes care of executing these tasks whenever machines become available. This strategy is less suited for supercomputers such as the VSC-5, where batch jobs typically stay minutes or hours in the queue before being executed.

Approach: A possible solution to the load balancing problem at application level is to perform the scheduling within a batch job. To that end, we have ported our traditional parallelization approach to a job-level load balancer that was built on top of a distributed task queue (using the open-source Celery framework).

Results: The strong scaling results (fixed input) in Fig. 1 indicate that there is a sweet spot when selecting an effective number of compute nodes. The relatively low parallel efficiency mainly stems from the fact that the frequency of one core of the EPYC 7003 series processor, used in VSC-5, can be immensely boosted. Fig. 2 highlights that a good mix of threads and processes is needed to effectively use the large multi-core compute nodes of VSC-5.

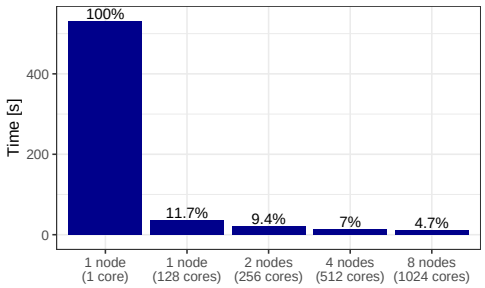


Fig. 1: Parallel runtime and efficiency (on top) of LigandScout for an increasing number of compute nodes (cores) on VSC-5.

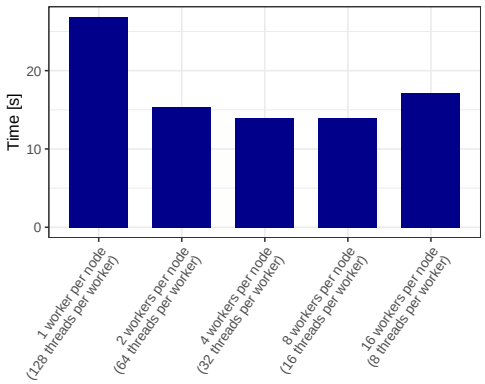


Fig. 2: Runtime of LigandScout when varying the number of workers and threads per worker. Notice that a total of 128 threads is always used per compute node.

Acknowledgements: This research was funded by the Austrian Research Promotion Agency (FFG).

References

[1] Thomas Kainrad, Sascha Hunold, Thomas Seidel, Thierry Langer: LigandScout Remote: A New User-Friendly Interface for HPC and Cloud Resources. *J. Chem. Inf. Model.*, 59(1): 31–37, 2019.

Benchmarking the CmDock Molecular Docking Software on HPC Systems

Vid Ravnik[a], **Marko Jukič**[a,b], **and Urban Bren**[a,b,c]

[a] *Faculty of Chemistry and Chemical Engineering, University of Maribor, Slovenia*
[b] *Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Slovenia*
[c] *Institute for Environmental Protection and Sensors, Slovenia*

Computer aided drug design has proven itself to be a powerful technique in the drug design toolbox. High throughput virtual screening (HTVS) of potential ligand candidates allows us to sift through orders of magnitude more compounds compared to classical *in vitro* biophysical assays (HTS, High Throughput Screening) [1]. Molecular docking represents a computational method for predicting the binding mode and affinity of a ligand to a receptor and is commonly applied for HTVS. Due to the computation time constraints molecular docking software often cannot afford to consider the level of detail which would be required for truly accurate results. Docking methods must therefore strike a balance between chemical accuracy and computation speed. Due to the approximations and neglected terms docking methods cannot reliably order molecules by their binding energy. In spite of this, molecular docking can be applied in screening large libraries for a small fraction of molecules which are more likely to bind strongly to the target. The goal of HTVS with molecular docking is therefore to obrain such an enriched subset of potential strong binders [1,2].

We are developing the CmDock molecular docking software, which represents an open source versatile molecular docking program based on rDock [2,3]. We have applied CmDock to the DUD-E and LIT-PCBA datasets to benchmark the software. These represent datasets of receptor structures containing known binders (actives) as well as a larger set of decoys with similar chemical and physical properties which do not bind with the receptors. We measured how well CmDock can enrich the ligands in the benchmark sets by evaluating receiver operating characteristic (ROC) curves, as well as the BEDROC and enrichment factor metrics.

We performed the calculations on the HPC Vega supercomputer, thoroughly examining the influence of the number of docking runs on the virtual screening results. We have also performed docking with the very popular open source AutoDock Vina docking software in order to compare the results with CmDock. LIT-PCBA contains multiple structures for each target, we docked the actives and decoys to each structure, this represents around 16.6 million ligands docked with receptors. DUD-E contains 102 targets and in total around 1.4 million ligands. We estimate that we have performed around 50 million individual receptor ligand dockings with CmDock and Vina together. Our results show that CmDock performs slightly better than Vina on the DUD-E benchmark, while Vina gives marginally better results on the LIT-PCBA benchmark.

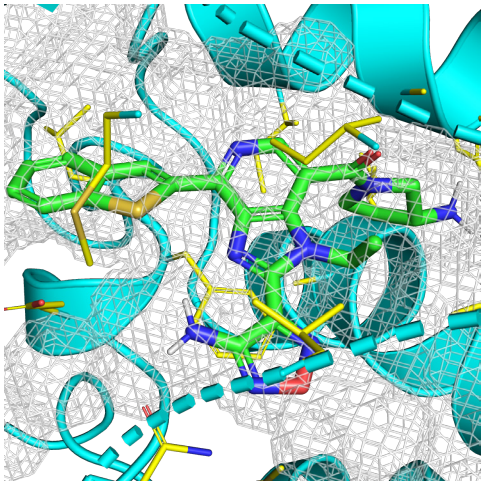


Fig. 1: An example of a favorable docking pose calculated by CmDock.

References

[1] Bender, B.J., Gahbauer, S., Luttens, A. et al., Nat Protoc 16, 4799 (2021).

[2] Jukič, M., Škrlj, B., Tomšič, G., Pleško, S., Podlipnik, Č. and Bren, U., Molecules 26, 3003 (2021)

[3] Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, Garmendia-Doval AB, Juhos S, Schmidtke P, et al., PLoS computational biology 10, e1003571 (2014)

OpenScienceLabs for HPC on the European Open Science Cloud

Constanze Roedig [a], **Kareem Sorathia**[b], **and the AOCC team**[c]

[a] *VSC Research Center, TU Wien, Austria*
[b] *Center for Geospace Storms, Applied Physics Lab, Johns Hopkins University, USA*
[c] *The Austrian Open Cloud Community (AOCC) is the continuation of the Austrian-wide digitalization project "Cluster Forschungsdaten", it is currently facilitated at TU Wien as part of the VSC Research Center, Austria*

We present a method [1] for publishing high-performance computing (HPC) code and results in a scalable, portable and ready-to-use interactive environment in order to enable sharing, collaborating, peer-reviewing and teaching. We show how we utilize cloud-native elements such as Kubernetes, containerization, automation and webshells to achieve this and demonstrate such an OpenScienceLab for the MAGE (Multiscale Atmosphere Geospace Environment) model, being developed by the recently selected NASA DRIVE Center for Geospace Storms. We argue that a key factor in the successful design of such an environment is its (cyber)-security, as these labs require non-trivial computing resources open to a vast audience. Benefits as well as implied costs of different hosting options are discussed, comparing public cloud, hybrid, private cloud and even large desktops. We encourage HPC centers to test our method using our fully open-source blueprints. We hope to thus unburden the research staff and scientists to follow FAIR principles and support open source goals without needing a deep knowledge of cloud computing.

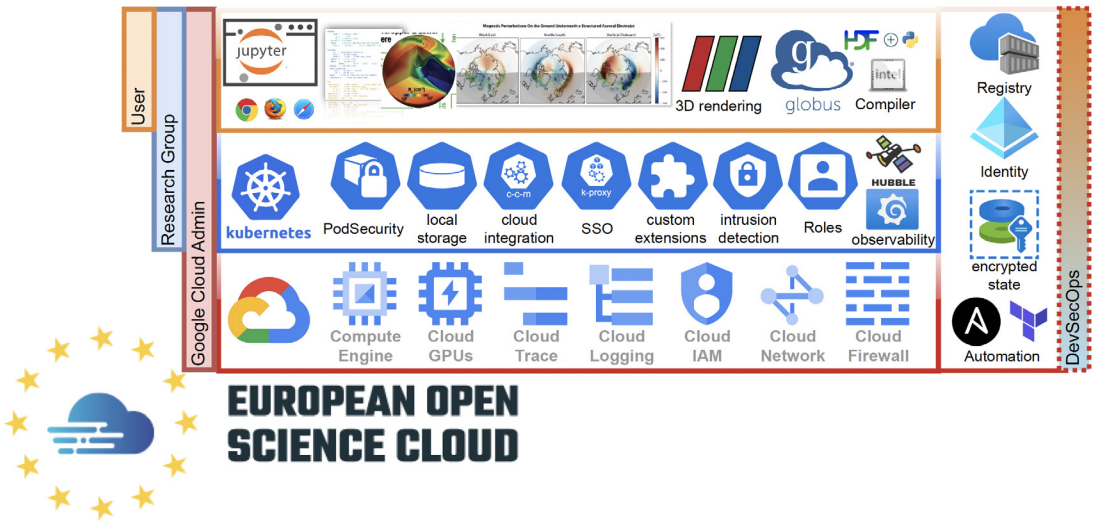


Fig. 1: Layers and Components to run an OpenScienceLab on Google Cloud Platform for the European Open Science Cloud (EOSC [2]): Users access their containerized HPC code via the web-browser wrapped into a Jupyter-Extension and have access to visualization, data (via e.g. Globus), their compilers and their own pre-built code (orange). Each Research Group self-manages their tenant (blue), which contains an opinionated Kubernetes thus abstracting low-level IT components from the users. Each tenant is ephemeral and of dynamic size. The cloud service provider (CSP) (red) is GCP in the EOSC case, however, at the VSC there is an alternative to run it on Openstack.

References

[1] Constanze Roedig, Kareem Sorathia, Michael Wiltberger, Elias Wimmer, Thomas Weber, Peter Kandolf, Franz Marko, Vlad Popescu-Vifor, Carina Urbanke, Brent Smith, Mojib Wali, Thomas Lahmer. Cloudnative OpenScienceLabs: Effectively sharing HPC results and knowledge, Proceedings of the 2023 Improving Scientific Software Conference, *submitted*

[2] https://eoscfuture.eu

Slovenian open access infrastructure – storing and sharing big data

***Marko Ferme*, Janez Brezovnik, and Milan Ojsteršek**

Faculty of Electrical Engineering and Computer Science, University of Maribor, Slovenia

Slovenian open access infrastructure [1] consists of institutional repositories (IRs) and a national portal (NP) that aggregates content from the repositories to provide a common search engine, recommendations of similar documents, and similar text detection. IRs store full-text versions of final study works, and research publications as well as other intellectual productions from universities (teaching materials, project reports, studies, monographs, research data, etc.).

With the increased availability of HPCs in Slovenia [2] the production of datasets has also increased and with that, also the demand for their storage. The Slovenian infrastructure had to be upgraded to support the research processes in terms of data management and providing the researchers an efficient way to store and publish their datasets in a way, that complies with the FAIR [3] principles.

To achieve this goal, we upgraded our Slovenian open access infrastructure, to store and share big data. We set up mechanisms for efficient transfer of big data, provided services that equip the datasets with permanent unique identifiers (PIDs), created metadata schemas and user interfaces which equip the datasets with metadata, automized publishing the datasets in the IRs and NP and enabled sharing the data with other researchers and archived all the existing digital object in the infrastructure and equipped them with PIDs.

We also set up the process for the researchers to publish their data and upgraded the IRs to support such a workflow. This enables the researchers to publish their datasets efficiently and enables them to easily use datasets found in NP or IRs on the HPC systems.

References

[1] M. Ojsteršek, J. Brezovnik, M. Kotar, M. Ferme, G. Hrovat, A. Bregant, M. Borović, Establishing of a Slovenian open access infrastructure: A technical point of view," Program, vol. 48, no. 4, pp. 394-412, Aug. 2014.

[2] Vega description: https://www.izum.si/en/hpc-en/

[3] Fair principles: https://www.go-fair.org/fair-principles/

Federated storage status and plans in SLING centres

Barbara Krašovec, Andrej Filipčič, and *Dejan Lesjak*

Jožef Stefan Institute, Slovenia

Similar to compute capacity, there is an increasing need for larger storage for input and output data among various user groups that is then processed at various HPC centres [1]. As the data is needed at different HPC sites, it should then be easily accessible to the users and their jobs as conveniently as possible, while taking care not to complicate maintenance for the administrators and of course not to compromise security. Hence, the need for federated storage in the HPC landscape. This is also the topic of interest in multiple collaborative projects (e.g. EOSC data lakes [2]) and could be an advantage for attracting industry users with lower costs for high-bandwidth connections.

Implementing such storage of course assumes the availability of some federated identity and credential service that then enables users to seamlessly access their data from participating sites. Since access can be both interactive and in batch mode in the case of submitted jobs, the method of identifying and authorizing a user should be usable in both modes of access. Support of LDAP with Kerberos, X.509 certificates with VOMS attributes and scitokens are some of the available options.

While some applications can now access data via newer protocols like S3, a lot of them only work on a POSIX filesystem, which has to be taken into account when considering options. There are, however, solutions that offer different modes of access in the same implementation. Additionally, access is often needed outside of HPC sites, for example at laboratories and detectors where data taking takes place. Therefore, access to storage should allow for secure and practical usage outside of the controlled environment of HPC systems.

An important aspect of offerings in this context is also the possibility to access data from HPC compute nodes that have limited or no access to the internet outside the site. Gateways and proxy servers can be considered to alleviate this problem giving advantage to systems that can use such options.

SLING sites offer users several ways to access and use available storage capacity. This talk attempts to highlight their benefits and weaknesses as well as outline current plans for storage access offerings. Differences between different usage modes (e.g. mostly read, mixed read-write) are explored to show that not every usage must be satisfied with a single system. Among the solutions and protocols explored are dCache, NFS, Ceph, S3 and EOS [3]. The latter currently taking main focus in implementation and deploying across participating HPC sites.

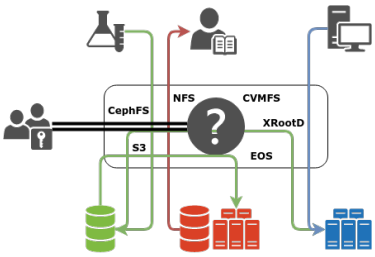


Fig. 1: Providing users with sensible access to their data is one of the interesting challenges for modern HPC centres and their collaborations.

References

[1] Ray, B.R., Chowdhury, M., Atif, U. (2017). Is High Performance Computing (HPC) Ready to Handle Big Data?, Communications in Computer and Information Science, vol 759. Springer, Cham. (2017)

[2] Di Maria, R., Dona, R., on behalf of the ESCAPE project, ESCAPE Data Lake, EPJ Web Conf., 02056 (2021).

[3] Mascetti, et al., CERN disk storage services: EPJ Web Conf. 245, 04038 (2020).

Keynote Talk:

European Processor Initiative

Mario Kovač

Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia

Digital sovereignty has become a key requirement for the EU, which is now striving to move from a simple user of ICT technologies to a main player on the digital market, both for products and human skills. The need for the EU (and the EuroHPC JU) to deploy European world-class exascale supercomputers in Europe is urging. The European Processor Initiative (EPI) develops a roadmap for future European high-performance low-power processors aligned with the EuroHPC roadmap.

EPI [1] focuses on processor technologies based on the ARM HPC processor architecture, and on architecture and HPC accelerator technologies based on RISC-V, the open-source ISA. RISC-V is also the candidate to become Europe's own technology for covering long term objectives. This phase also seeks to expand the scope of the project into adjacent European and global vertical markets that can leverage HPC to enable new and improved systems and solutions.

EPI has the following objectives:

- Finalize the development and the bring-up of the first generation of low-power processor units developed in SGA1,
- Develop the second generation of the General Purpose Processor (GPP) applying technological enhancements targeting the European Exascale machines with respect to the GPP (Rhea) of SGA1,
- Develop the second generation of low-power accelerator test chips, usable by the HPC community for tests, and
- Develop sound and realistic industrialization and commercialization paths, and enable long-term economical sustainability with an industrialization path in the edge computing area demonstrated in a few well-chosen proofs of concepts.

Fig. 1: EPAC test sample.

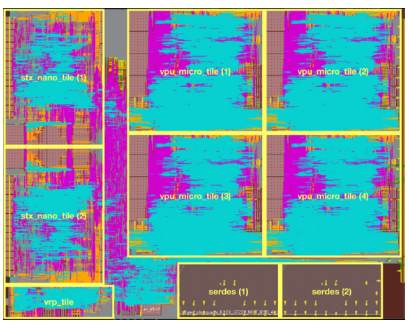


Fig. 2: EPAC layout with VPU, STX and VRP accelerators with 25 mm2 in GF 22FDX technology.

References

[1] https://www.european-processor-initiative.eu

Energy Efficiency in HPC - European, Global and National Perspective

Markus Hickel[a] **and *Siegfried Höfinger***[a,b]

[a] *VSC Research Centre, TU Wien, Austria*
[b] *Department of Physics, Michigan Technological University, USA*

Prices for electric energy have been soaring in Europe in recent times [1] (Fig. 1). Austria had been hit particularly hard as evidenced by the prices already doubling prior to the Russian attack on Ukraine and the resulting embargo on natural gas imports [2] (Fig. 2). Obviously, this has important consequences for HPC centers and their operational costs, and the central question arising is how to best respond to such a crisis in terms of user awareness and potential adaptations to usage models with a focus on energy efficient computing – hence subjects traditionally ignored by the conventional HPC user in the academic setting.

In this present contribution we shall be trying to establish a common viewpoint on whether or not – and in case how – to develop energy efficient computing on VSC systems. The basis will be formed by two important resources,

1) a recently hold BoF* at ISC23 (European perspective) where one of the topics discussed was "training in the context of energy efficient HPC",

2) The energy efficient HPC working Group [3] (global perspective) with frequent workshops and BoFs at SC/ISC/ISCG/HPC-Asia events and regular usergroup meetings discussing important aspects in specific sub-divisions.

* "Birds of a Feather" – a user group discussion format used at major supercomputing conferences.

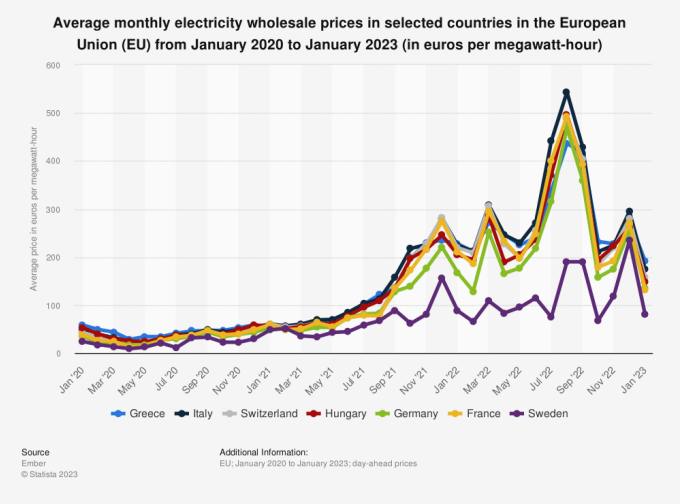


Fig. 1: Monthly electricity prices in selected EU countries

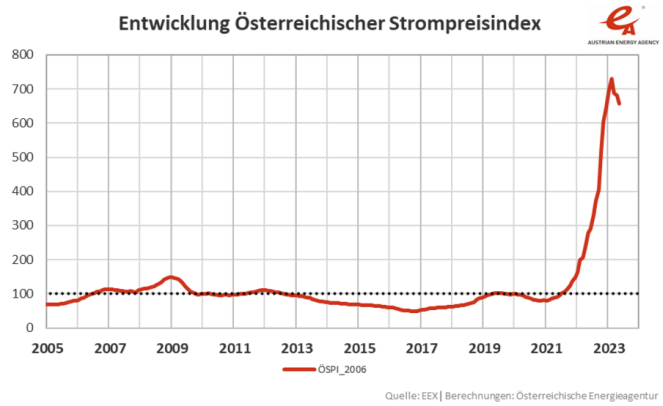


Fig. 2: Development of costs for electric energy in Austria

References

[1] Statista Ltd. 209-215 Blackfriars Road, SE1 8NL London, UK https://www.statista.com (2023).

[2] Österreichische Energieagentur https://www.energyagency.at (2023).

[3] Energy Efficient High Performance Computing Working Group https://eehpcwg.llnl.gov (2023).

A GPU Cluster for Small Teams to Efficiently Conduct Deep Learning Research and Development

Domen Tabernik **and Luka Čehovin Zajc**

Faculty of Computer and Information Science, University of Ljubljana, Slovenia

The rapid advancement of deep learning techniques has led to a growing demand for efficient and flexible GPU-based computing infrastructure to support the development and research of deep learning-based solutions. We present our in-house developed cluster architecture, designed specifically to cater the requirements of small teams working on various deep learning projects. The primary challenge faced by these teams is the need for rapid iteration and direct debugging on compute nodes during the development process, where the access to resources (primarily GPUs) is not constant for a given time-span, but is required sporadically without a well defined schedule. Traditional job-based architectures, such as SLURM [1], are not suitable for this use-case due to rigid resource allocation process with exclusive locks, potential delays in resource allocation and the lack of support for long-lived jobs that may only occasionally require access to resources, as is the case in the development process using an IDE.

Architecture: Our HPC architecture is an attempt to address the challenges that we encounter in our every day work, namely the need for a flexible user/project-wide isolation and flexible resource allocation for long-running development sessions. Our system consists of several independent mutli-GPU nodes. The nodes only manage local GPUs, while leaving CPU/RAM management to the OS. Different to GPU resources, CPU and RAM can be shared between processes on a system level without significant performance penalties. This decision was validated in practice. We have observed that deep-learning GPU-bound tasks are mostly limited by the GPU availability and rarely by RAM and CPU. We have developed a basic deployment system [2] based on Ansible and YAML configuration, which also manages automatic reconfiguration of the system based on user-provided settings.

Allocation: For maximum flexibility, GPU allocation is decoupled from the rest of the system. A service on each node monitors GPU allocations and can be used to request free GPU devices automatically or wait for free resources in case of congestion. The usage of the system is optional, each user can also acquire GPU resources manually. Compared to job-queue based architectures, such as SLURM, our approach directly monitors the actual use of resources thus avoiding predefined resource allocation.

Isolation: Our isolation system is called Conda Compute Containers [3], it is based on a Docker image hierarchy that provides a Conda environment and can be extended with other services (i.e. IDE, Jupyter notebooks). The Conda environment is shared between nodes using NFS. Containers are deployed on all nodes and are persistent. Access to individual environments is provided through a reverse TCP proxy using SSH as well as forwarding for HTTP services. Containers on different nodes are also connected locally through Docker virtual networks, this enables the use of multiple GPUs on multiple nodes to train large models with large batch-sizes.

Conclusion: Our custom HPC architecture offers a flexible and efficient GPU sharing solution for small teams working on deep learning projects, specifically addressing the challenge of rapid iteration and direct GPU debugging. Several compromises had to be made to ensure the flexibility needed. The system assumes cooperative behavior as well as some manual involvement in resource allocation.

References

[1] Yoo, A. B., Jette, M. A., & Grondona, M. (2003). SLURM: Simple Linux Utility for Resource Management. In Job Scheduling Strategies for Parallel Processing (pp. 44-60).

[2] https://github.com/vicoslab/ccc-deployment

[3] https://github.com/vicoslab/ccc

Running Windows-applications on a Linux HPC cluster using WINE

***Alois Schlögl*, Stefano Elefante, and Victor-Valentin Hodirnau**

Institute of Science and Technology Austria, Austria

High-Performance Computing (HPC) is an area of computer science where Linux-based operating systems have established their dominance [1]. As a consequence, applications designed purely for Microsoft (MS) Windows do not run on GNU/Linux HPC machines. This is often not an issue if the developers provide binaries for all platforms or if the source code is available, as it might be recompiled for GNU/Linux. However, there are still widely-used applications, such as WarpEM [2] or Imaris, where developers provide only a Windows version, and recompilation on Linux is not always feasible. Thus, there is a need to bridge compatibility between Windows-based applications and GNU/Linux HPC computers.

Solutions based on some kind of virtualization engine might be available but have their own drawbacks (performance penalties, increased memory requirements). In addition, these engines usually don't run in user-space, which is problematic when trying to integrate them with schedulers like SLURM. Therefore, we tested whether such MS-Windows applications could run on our HPC machines using WINE [2], a thin compatibility layer that allows Windows applications to run under GNU/Linux.

We tested whether Warp [2] can run under Wine [3] on our HPC cluster [4] with the following setups: Debian 11 (Bullseye), Wine compiled from source, Wine Staging patches (in order to enable GPU support), X-tunneling through SSH, x2go, and VGL/VNC for remote GUI-based access. So far, we have had success in running WarpEM using Wine Staging v7.17 to v7.21 (with version 7.22 the set of patches for *nvidia-cuda* has been disabled).

Imaris 9.9.1, Imaris Viewer 9.9.1 and 10.0.0 run without issue on newer versions of Wine.

Of course some Windows applications do not run on Wine. For those that we tested and failed, we submitted test reports to the WINE application database [3] (https://appdb.winehq.org/).

Specifically, issues have been observed with the graphics stack over remote connections to the cluster. This affects Mesa, x2go, and VirtualGL. To address some of these issues, some newer versions of Mesa have been compiled by us and are provided via the Lmod module system.

Wine and its applications run purely in user-space. This has several advantages, (i) it works well with our scheduling system, (ii) the security assessment is easy, (iii), there is very little computation overhead, (iv) addition layers in the software (hypervisor/emulator) and an additional operating system (MS Windows) can be avoided. Therefore, Wine seems to be a viable option for us to run Windows applications on a GNU Linux cluster.

Acknowledgement: Thanks to Jesse Hansen for his suggestions on improving the abstract.

References

[1] https://top500.org/statistics/overtime/ - Statistics - Development over Time Category - Operating System Family, (March 2023).

[2] Dimitry Tegunov, D., and Cramer, P. Nature Methods **16**, 1146–1152 (2019).

[3] https://www.winehq.org/

[4] Schlögl, A., Janos Kiss, J., and Elefante, S., Is Debian suitable for running an HPC Cluster?, Austrian High-Performance Computing meeting AHPC19, Grundlsee, Austria, 25th-27th Feb 2019.

MPI is Good, Control is Better: Checking Performance Guidelines of Collectives

Sascha Hunold **and Maximilian Hagn**

Research Group for Parallel Computing, Faculty of Informatics, TU Wien, Austria

MPI Collective Operations, such as `MPI_Bcast` and `MPI_Allreduce`, are important building blocks for many HPC applications. It is important to point out that only the semantics of each collective operation are defined in the MPI standard. Therefore, each MPI library may use different algorithms to implement MPI collectives. Additionally, some algorithms only work efficiently on a subset of communication problems, e.g., they were designed for communicating small messages, but they are rather inefficient for large messages.

Self-consistent Performance Guidelines can be formulated for each collective call [1]. These guidelines define naturally arising performance expectations for specialized collectives, e.g.,

$$\texttt{MPI\_Allreduce}(n) \preceq \texttt{MPI\_Reduce}(n) + \texttt{MPI\_Bcast}(n) \text{ or}$$
$$\texttt{MPI\_Scatter}(n) \preceq \texttt{MPI\_Bcast}(n).$$

A former guideline states that an Allreduce call with a payload of n data items should not be slower than the subsequent calls to Reduce and Bcast, which semantically perform the same operation. Another example is shown in Fig. 1, where the processes mimic the Scatter functionality by (1) first calling Bcast and then (2) copying the relevant pieces to form the correct result buffer.

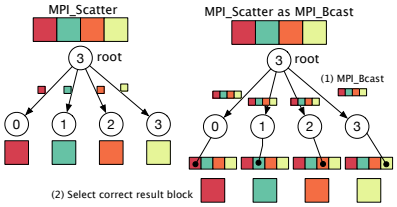


Fig. 1: MPI_Scatter as MPI_Bcast guideline.

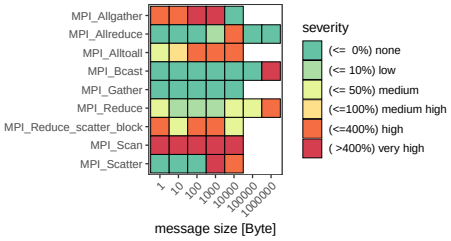


Fig. 2: `pgchecker` results for `16x128` processes on VSC-5, Open MPI 4.1.4. The severity denotes the slowdown of default MPI (in percent) compared to the best guideline implementation.

Goal and Approach: We examine the performance-consistency of modern MPI libraries with respect to defined performance guidelines [2] on current supercomputers, e.g., VSC-5 (Austria). We have implemented a tool, called `pgchecker`, which analyzes the performance consistency of MPI libraries. It is built on top of the ReproMPI benchmark [3] and implements several statistical tests (e.g., Mann-Whitney), which allows for an automatic and non-parametric testing of MPI libraries' performance-consistency.

Results: Fig. 2 shows `pgchecker` results for the VSC-5 and Open MPI, where the individual colors denote the severity of performance-guideline violations. Orange and red squares represent serious performance deficits. We observe several severe performance issues with Open MPI on VSC-5, e.g., with a payload of 1 MB, the default Bcast is more than 4 times slower than one of the mock-up implementations.

Acknowledgement: This work was partially supported by the Austrian Science Fund (FWF) project P33884-N.

References

[1] Jesper Larsson Träff, William D. Gropp, Rajeev Thakur: Self-Consistent MPI Performance Guidelines. *IEEE Trans. Parallel Distributed Syst.*, 21(5): 698–709, 2010.

[2] Sascha Hunold and Alexandra Carpen-Amarie: Autotuning MPI collectives using performance guidelines. In *Proceedings of the HPC Asia 2018*, pages 64–74. ACM, 2018.

[3] Sascha Hunold and Alexandra Carpen-Amarie: Reproducible MPI benchmarking is still not as easy as you think. *IEEE Trans. Parallel Distributed Syst.*, 27(12):3617–3630, 2016.

Rank Reordering within MPI Communicators to Exploit Deep Hierarchal Architectures of Supercomputers

***Philippe Swartvagher*, Ioannis Vardas, Sascha Hunold, and Jesper Larsson Träff**

Faculty of Informatics, TU Wien, Austria

HPC machines are more and more hierarchical: they feature nodes, processors, NUMA domains, L3 caches, ... The performance of applications can depend on how computations and data are mapped to this hierarchy, thus it is important to consider it. Regarding how MPI processes are distributed across available cores, the performance of communications can change, because of the capacity of links on each hierarchy level or contention on these links. Thus, each application has its own optimal process mapping, influenced by the capacities of the hardware and the characteristics of computations and communications. Unfortunately, expressing a mapping based on the hierarchical topology of the system is not straightforward, especially to later execute collective operations [1].

Here, we propose an algorithm to reorder ranks of MPI processes inside a communicator. The idea is to decompose the number corresponding to the MPI rank into a mixed-radix number with a numerical base built from the characteristics of the hierarchy. Changing the order of the base allows to change how the cores in the system hierarchy are enumerated to relabel their ranks. For instance, an order can make two consecutive ranks to always be mapped on different nodes, while another order can make them mapped only on different sockets. Each order is described as a sequence of integers corresponding to the levels of the system hierarchy. For instance, the default order commonly used to enumerate MPI processes is $3-2-1-0$ for a system with a hierarchy composed of 4 levels, such as Node − Socket − NUMA − Core.

These different mappings can have a large effect on overall communication performance. In our experiments, we use this *reordering* method to create sub-communicators. The used order allows to control how MPI processes belonging to a communicator are spread all over the available nodes or rather packed as much as possible inside a single level of the hierarchy. As depicted in Fig. 1 and Fig. 2, results of microbenchmarks executing `MPI_Alltoall` operations in sub-communicators show a performance difference of factor 2 between the best and the worst order when only one communicator executes the collective operation, and a factor of 4 when all communicators simultaneously execute it, respectively. When only one sub-communicator executes collective operations, it is better to have a spread mapping to maximize available bandwidth; while a packed mapping is better when all sub-communicators simultaneously execute collective operations, to minimize contention. By changing the order of rank numbering, we also see an impact on the performance of proxy-applications which execute collective operations in sub-communicators.

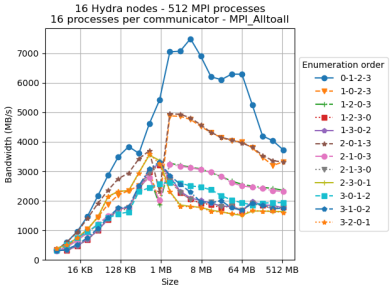


Fig. 1: Impact of different reorderings on communication bandwidth when executing `MPI_Alltoall` operation in one communicator of 16 processes (out of 512).

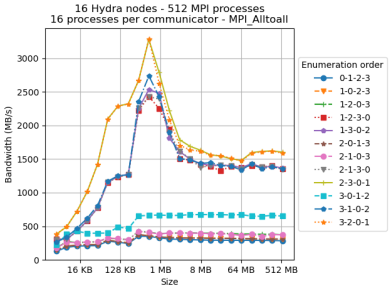


Fig. 2: Impact of different reorderings on communication bandwidth when simultaneously executing `MPI_Alltoall` operations in 16 communicators, each with 16 processes.

References

[1] Mirsadeghi, S.H. and Afsahi A., Topology-Aware Rank Reordering for MPI Collectives. IEEE International Parallel and Distributed Processing Symposium Workshops (2016).

Joint propagation characteristics of acoustic and electromagnetic waves in shallow ocean

Alena Melic[a] **and Christoph F. Mecklenbräuker**[b]

[a] *VSC Research Center, TU Wien, Austria*
[b] *Institute of Telecommunications, TU Wien, Austria*

Most underwater communication and sensing systems use acoustic waves exclusively because of their low loss propagation characteristics. In this work, we aim at joint processing of sensor array data for acoustic and electromagnetic/optical waves for estimating constitutive parameters of a layered shallow ocean model. Key acoustic propagation media characteristics are sound speed and density profiles, whereas permittivity and conductivity are key characteristics for electromagnetic propagation. Many approximations to the modelling of wave propagation have been used: analytical (e.g. Ray Methods, Normal Modes) and numerical (Finite Elements, Finite Differences) [1]. We use the finite element method to compute the sound pressure field in response to an isotropic point source emitting a sinusoidal spherical wave in a shallow ocean with sound speed $c = 1500\,\mathrm{m/s}$, water depth $d = 130\,\mathrm{m}$ at a frequency of $f = 170\,\mathrm{Hz}$.

Initial results for acoustic propagation are obtained with a commercially available three-dimensional finite element solver (Iterative Solver Shifted Laplace) using the Vienna Scientific Cluster (VSC) as numerical platform. This method is computationally expensive, therefore the calculation at the VSC. The finite element solver is used for finite volume calculations; a problem with the ocean. Through boundary conditions, such as the Perfectly Matched Layer (PML), acoustic propagation calculations and approximations are nevertheless achievable.

The PML setup allows one or two modes with a specific wave number to leave the modeling domain with minimal reflections. We specify the size of the modeling domain to be 130 m in length times 130 m in width times 130 m in height and the maximum tolerable element size $\lambda/5 \approx 1.8\,\mathrm{m}$. The complete mesh consists of 6991923 domain elements, 76460 boundary elements, and 900 edge elements. The boundary conditions for this model: The sea surface is simulated by the Dirlichlet boundary, which satisfies $p = 0$. The seabed is simulated by the Neumann boundary, which satisfies $dp/dn = 0$. We deal with the open boundaries, with the PML approach [2].

The numerical result for the sound pressure field is visualised in Fig. 1 and compared to the normal mode approximation.

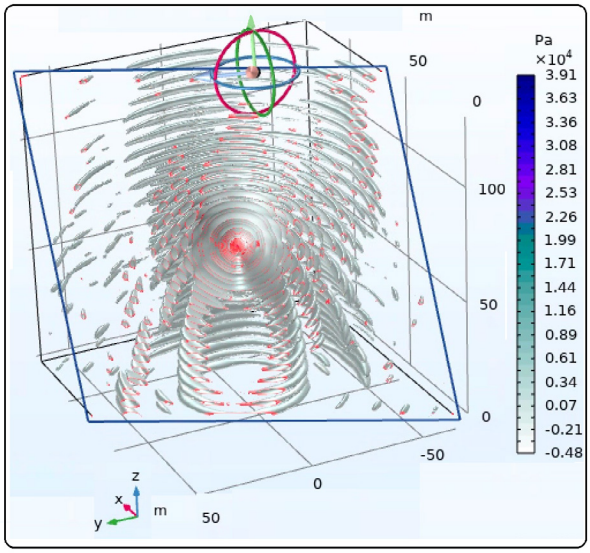


Fig. 1: Surface: Total acoustic pressure; Simulation with isotropic point source and ocean attenuation, estimated with finite element solver

References

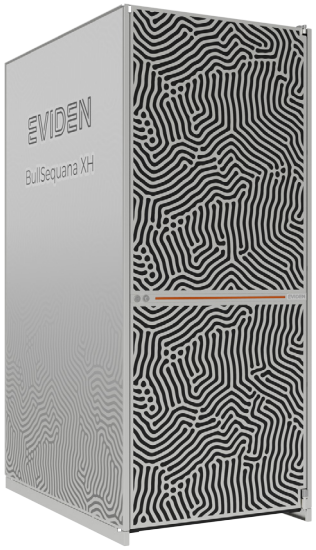
[1] F.B. Jensen, W.A. Kuperman, M.B. Porter, H. Schmidt, Computational Oceans Acoustics AIP Press, (1994).

[2] Y. Zhou, Wen-Yu Luo, A Finite Element Model for Underwater Sound Propagation in 2-D Environment Journal of Marine Science and Engineering, (2021).

ASHPC23 is sponsored by:

EVIDEN

Eviden is an Atos business that brings together its digital, cloud and big data & security business lines. It will be a global leader in data-driven, trusted and sustainable digital transformation. A next-generation digital business, Eviden stems from worldwide leading positions in digital, cloud, data, advanced computing and security. Our HPC, HPCaaS, Quantum computing, Edge & Business computing and AI solutions push the potential of next-gen computing. Securely capture, compute and analyze data at the edge, in the cloud, and in data centers.



Explore endless possibilities by harnessing HPC. Whether it's about more accurate weather forecast, new drug discovery, smart city, or nuclear defense breakthrough, Digital simulation is at the center of our daily life. Harnessing the power of HPC toward Exascale while taking care of its environmental impact will unlock endless new possibilities to us all. Discover Atos' line of HPC hardware and software products including BullSequana XH3000, Nimbix Supercomputing Suite, ThinkAI, and JARVICE XE HPC Enterprise Software.

BullSequana XH3000 Combining the cutting-edge processing technologies and AI mechanism and the Atos Direct Liquid Cooling (DLC) technology, BullSequana XH3000 delivers both unprecedented performance and unrivaled efficiency. With the Open Sequana program, Atos allows for 3rd party technology ecosystem participation and delivers to customers a future-proof platform ready for quantum accelerators.

ThinkAI Atos ThinkAI accompanies enterprises and researchers to contemplate, architect, build, and deliver scalable, energy-efficient, high-performance AI solutions in weeks. The Atos Center of Excellence in Advanced Computing combines in-depth industry knowledge with data science proficiency to strategize a purposeful roadmap for your AI journey. Atos supports enterprises and academic institutions by orchestrating hardware, software, and expertise in digital security, resulting in accelerated time to AI operationalization and industrialization at optimized cost.

Index of presenting authors

List of ASHPC23 participants

Ady, László	adylaszlo@nexttechnologies.hu	Karman Mechanics Ltd., Hungary
Azab Mohamed, Abdelrahman	azab@uio.no	University of Oslo, Norway
Belavić, Martin	martin.belavic@srce.hr	University of Zagreb, Croatia (Local Name: Hrvatska)
Blaas-Schenner, Claudia	claudia.blaas-schenner@tuwien.ac.at	Technische Universität Wien, Austria
Bogdanović, Leon	leon.bogdanovic@lecad.fs.uni-lj.si	University of Ljubljana, Slovenia
Borovinšek, Matej	matej.borovinsek@um.si	University of Maribor, Slovenia
Caravita, Caterina	c.caravita@cineca.it	CINECA, Italy
Dalla Costa, Diego	e1646641@student.tuwien.ac.at	Technische Universität Wien, Austria
Dellago, Christoph	Christoph.Dellago@univie.ac.at	University of Vienna, Austria
Egger, Daniel	deg@zurich.ibm.com	IBM Quantum, IBM Research Europe - Zurich, Switzerland
Elefante, Stefano	stefano.elefante@ista.ac.at	Institute of Science and Technology Austria (ISTA), Austria
Felhos, David	david.felhos@karman-mechanics.hu	Karman Mechanics Ltd., Hungary
Ferme, Marko	marko.ferme@um.si	University of Maribor, Slovenia
Filipčič, Andraž	andraz.f lipcic@ijs.si	Jožef Stefan Institute, Slovenia
Fink, Michael	Michael.Fink@uibk.ac.at	Universität Innsbruck, Austria
Ganguly, Raman	raman.ganguly@univie.ac.at	University of Vienna, Austria
Gergely, Eva	eva.gergely@univie.ac.at	University of Vienna, Austria
Goiser, Malgorzata	malgorzata.goiser@tuwien.ac.at	Technische Universität Wien, Austria
Goldenberg, Florian	f orian@goldenberg.at	Technische Universität Wien, Austria
Grossoehme, Peter	peter.grossoehme@megware.com	MEGWARE Computer Vertrieb und Service GmbH, Germany
Haase, Gundolf	gundolf.haase@uni-graz.at	University of Graz, Austria
Harisch, Damjan	damjan.harisch@arnes.si	Arnes, Slovenia
Haunschmid, Ernst	not disclosed	Technische Universität Wien, Austria
Herček, Ariela	ariela.hercek@arnes.si	Arnes, Slovenia
Hoef nger, Siegfried	siegfried.hoef ger@tuwien.ac.at	Technische Universität Wien, Austria
Hunold, Sascha	not disclosed	Technische Universität Wien, Austria
Ilc, Nejc	nejc.ilc@fri.uni-lj.si	University of Ljubljana, Slovenia
Ilijas, Tomi	info@arctur.si	Arctur, Slovenia
Javoršek, Jan Jona	jona.javorsek@ijs.si	Jožef Stefan Institute, Slovenia

Jenko Bizjan, Barbara	not disclosed	University Medical Center Ljubljana, Slovenia
Jukič, Marko	marko.jukic@um.si	University of Maribor, Slovenia
Karasek, Tomas	tomas.karasek@vsb.cz	VSB - Technical University of Ostrava, Czech Republic
Kero, Stefano	stefan.kero@eviden.com	Eviden, Austria
Koniorczyk, Mátyás	koniorczyk.matyas@wigner.hu	Wigner Research Centre for Physics, Hungary
Kovač, Mario	mario.kovac@fer.hr	University of Zagreb, Croatia (Local Name: Hrvatska)
Kozubek, Tomas	tomas.kozubek@vsb.cz	VSB - Technical University of Ostrava, Czech Republic
Kralj, Marko	marko.kralj@izum.si	Institute of Information Science (IZUM), Slovenia
Kreidl, Sabine	not disclosed	Universität Innsbruck, Austria
Lebar Bajec, Iztok	ilb@fri.uni-lj.si	University of Ljubljana, Slovenia
Lesjak, Dejan	dejan.lesjak@ijs.si	Jožef Stefan Institute, Slovenia
Lorenčič, Samo	samo.lorencic@izum.si	Institute of Information Science (IZUM), Slovenia
Lotrič, Uroš	not disclosed	University of Ljubljana, Slovenia
Magrys, Marek	not disclosed	ACK Cyfronet AGH, Poland
Maj, Marta	not disclosed	ACK Cyfronet AGH, Poland
Mantegazza, Francesco Carlo	francesco.mantegazza@uni-graz.at	University of Graz, Austria
Mazur, Hubert	not disclosed	Wrocław University of Science and Technology, Poland
McKevitt, James	jamesm20@univie.ac.at	University of Vienna, Austria
Melic, Alena	alena.melic@tuwien.ac.at	Technische Universität Wien, Austria
Muck, Katrin	katrin.muck@tuwien.ac.at	Technische Universität Wien, Austria
Neumayer, Michael	michael.neumayer@univie.ac.at	University of Vienna, Austria
Noga, Klemens	not disclosed	ACK Cyfronet AGH, Poland
Novotný, Michal	michal.novotny@vsb.cz	VSB - Technical University of Ostrava, Czech Republic
Oppel, Markus	markus.oppel@univie.ac.at	University of Vienna, Austria
Ostermann, Alexander	alexander.ostermann@uibk.ac.at	Universität Innsbruck, Austria
Pančur, Matjaž	not disclosed	University of Ljubljana, Slovenia
Pešatová, Karina	not disclosed	VSB - Technical University of Ostrava, Czech Republic
Pfennig, Tobias	tobias.pfennig@megware.com	MEGWARE Computer Vertrieb und Service GmbH, Germany
Pilipović, Ratko	ratko.pilipovic@fri.uni-lj.si	University of Ljubljana, Slovenia
Povh, Janez	janez.povh@rudolfovo.eu	Rudolfovo - Science and technology center Noov mesto, Slovenia
Prah, Alja	not disclosed	Jožef Stefan Institute, Slovenia

Prica, Teo	teo.prica@izum.si	Institute of Information Science (IZUM), Slovenia
Rauber, Andreas	rauber@ifs.tuwien.ac.at	Technische Universität Wien, Austria
Ravnik, Vid	vid.ravnik2@um.si	University of Maribor, Slovenia
Reichl, Irene	irene.reichl@cadfem.at	CADFEM (Austria) GmbH, Austria
Reiter, Eduard	eduard.reiter@uibk.ac.at	Universität Innsbruck, Austria
Riha, Lubomir	not disclosed	VSB - Technical University of Ostrava, Czech Republic
Roedig, Constanze	constanze.roedig@tuwien.ac.at	Technische Universität Wien, Austria
Romaner, Lorenz	lorenz.romaner@unileoben.ac.at	Montanuniversität Leoben, Austria
Sattari , Sanaz	sanaz.sattari@tuwien.ac.at	Technische Universität Wien, Austria
Schlögl, Alois	alois.schloegl@ist.ac.at	Institute of Science and Technology Austria (ISTA), Austria
Seren, Ümit	uemit.seren@gmi.oeaw.ac.at	Gregor Mendel Institute of Molecular Plant Biology (GMI), Austria
Šikaleska, Violeta	violeta.sikaleska@izum.si	Institute of Information Science (IZUM), Slovenia
Sluga, Davor	davor.sluga@fri.uni-lj.si	University of Ljubljana, Slovenia
Steiner, Soner	soner.steiner@tuwien.ac.at	Technische Universität Wien, Austria
Störi, Herbert	herbert.stoeri@tuwien.ac.at	Technische Universität Wien, Austria
Swartvagher, Philippe	swartvagher@par.tuwien.ac.at	Technische Universität Wien, Austria
Tabernik, Domen	domen.tabernik@fri.uni-lj.si	University of Ljubljana, Slovenia
Thaler, Martin	martin.thaler@uibk.ac.at	Universität Innsbruck, Austria
Tomić, Draško	drasko@irb.hr	Ruđer Bošković Institute, Croatia (Local Name: Hrvatska)
Tomšič, Pavel	pavel.tomsic@fs.uni-lj.si	University of Ljubljana, Slovenia
Tykierko, Mateusz	mateusz.tykierko@pwr.edu.pl	Wrocław University of Science and Technology, Poland
Valh, Dejan	dejan.valh@izum.si	Institute of Information Science (IZUM), Slovenia
Vardas, Ioannis	vardas@par.tuwien.ac.at	Technische Universität Wien, Austria
Vasileska, Ivona	ivona.vasileska@fs.uni-lj.si	University of Ljubljana, Slovenia
Vialov, Ivan	ivan.vialov@tuwien.ac.at	Technische Universität Wien, Austria
Vianello, Marco	marcov@math.unipd.it	University of Padova , Italy
Vodopivec, Urša	ursa.vodopivec@arnes.si	Arnes, Slovenia
Vrhovšek, Blaž	blaz.vrhovsek@kclj.si	University Medical Center Ljubljana, Slovenia
Wang, Yin	Yin.Wang@uibk.ac.at	Universität Innsbruck, Austria
Wimmer, Iuliia	julia.wimmer@tuwien.ac.at	University of Vienna, Austria
Zabloudil, Jan	not disclosed	University of Natural Resources and Life Sciences (BOKU Wien), Austria

Zaharijas, Gabrijela	not disclosed	University of Nova Gorica , Slovenia
Zamuda, Aleš	ales.zamuda@um.si	University of Maribor, Slovenia
Zechner, Jürgen	juergen.zechner@tailsit.com	TAILSIT GmbH, Austria
Zemljak, Aleš	ales.zemljak@izum.si	Institute of Information Science (IZUM), Slovenia
Żyta, Kamil	not disclosed	Wrocław University of Science and Technology, Poland

One participant preferred not to be listed.

DOI: https://doi.org/10.25365/phaidra.423

ISBN: 978-3-200-09311-9

Published by:
EuroCC Austria
c/o Universität Wien
Universitätsring 1
1010 Vienna, Austria
https://eurocc-austria.at/

Edited by:
Eduard Reiter, Research Area Scientific Computing, University of Innsbruck, Austria

Layout:
Irene Reichl and Claudia Blaas-Schenner, VSC Research Center, TU Wien, 2016