EUROPEAN CENTRES OF EXCELLENCE IN HIGH PERFORMANCE COMPUTING

SUPPORTING EUROPEAN SCIENCE, INDUSTRY & SOCIETY

OVERVIEW

A NEW ERA OF HIGH PERFORMANCE COMPUTING IN EUROPE

Why do we need High Performance Computing?

Today, our society is facing massive challenges: Climate change and the biodiversity crisis are forcing a switch to a de-carbonised and, more generally, circular and sustainable economy; growing threats to public health due to aging populations, pandemics and pollution; the need to adapt to geopolitical and economic shifts. Solutions have many faces: switching to renewable forms of energy, developing novel materials reducing dependency on scarce resources, understanding and mitigating climate change, fast development of new drugs, personalized medicine, establishing more efficient forms of transport, understanding and anticipating social and economic phenomena – the list could go on and on. At the same time, the European Union (EU) has clearly identified the strategic importance of reducing its dependency on the import of resources (both technological and energy resources) from other parts of the world. Therefore, an important long-term strategy for the European Union is to use its technological potential to gain more independence and create well-being for its citizens through the solution of the key societal challenges.

The solution of many of the core technical challenges can be facilitated by the use of advanced computing: starting from the small cases that can be run on a notebook or a workstation to extreme-scale computational analysis of the largest (entire climate system) to the tiniest (atomic structure of materials) structures, ranging from technical structures (like buildings, turbines, or airplanes) to geological and earth-scale (volcanoes, climate system) to societal (migration, economics) and biological (virology, biomechanics) systems. This involves many computational techniques, ranging from physics-based simulations to analysis of massive amounts of data, and increasingly a mix of these and other approaches.

Shaping a truly European Supercomputing Ecosystem

The recently established EuroHPC Joint Undertaking is a joint initiative between the EU, European countries and private partners to develop a world class super-computing ecosystem in Europe. One goal is to install several extreme scale supercomputing systems, the fastest attaining Exascale performance by 2024 (see Exascale section). In parallel, the European Union supports large research programmes to develop genuinely European computing technology as a part of its strategy towards sovereignty in critical technologies.

Building computers fast enough to run the afore-mentioned high performance computations is necessary, but evidently not sufficient: it is the software (building itself on the theoretical models of the underlying phenomena) providing the bridge between theory and physical reality which makes computing so powerful. The European *High Performance Computing Centres of Excellence* (HPC CoEs) are building exactly these bridges between the world's top-notch computers, currently scratching at the Exascale level (see below), and the big societal challenges: they make sure efficient codes are ready for many different disciplines: the fastest computer is of no use if nobody can use it or no software can exploit its capabilities, and thus applications and their support through the HPC CoEs are a core part of the EU and EuroHPC strategy.

While many of these challenges need tremendous levels of computing power and data processing capabilities – in some cases, well beyond current capabilities, and with the potential to open up whole new areas for computing – the impact is not limited to the high end but goes the whole way down to smaller scale applications and single workstations (which are parallel computers as well these days). An important aspect of the CoEs is community building around the application ecosystem they develop, by making these technologies useable for a broad class of users in academia and industry beyond the narrow circle of specialists, providing training and building bridges between disciplines.

Figure 1: The LUMI supercomputer installed in Finland is one of the 3 pre-Exascale systems owned by the EuroHPC JU. It will provide in its final stage a peak performance of about 500 PetaFLOP/s and 17 Petabyte of storage. Its power consumption is about 200 MW, and its waste heat will provide 20% of the district heat in the area.

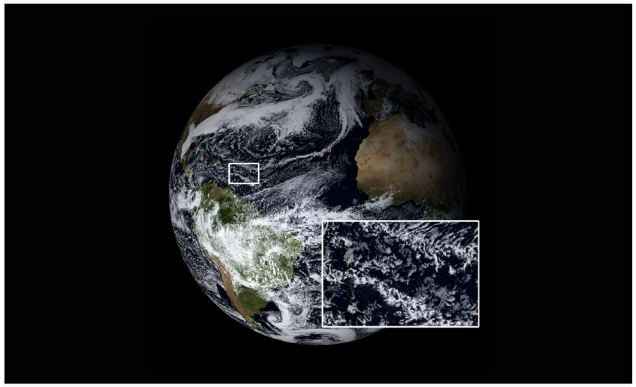


Figure 2: Global climate simulations are among the most demanding applications in terms of computational power. This is a result from the ESiWACE CoE, showing that with the emerging Exascale supercomputers, storm-resolving global climate models are within reach, establishing a direct link between global warming and the frequency of extreme wheather.

BACKGROUND ON HIGH PERFORMANCE COMPUTING

Those of us who have been using computers for longer than a decade remember the days when normal off-the-shelf desktop computers (or more precisely, their processors) got faster every 1-2 years, just by running each operation faster (i.e. increasing the processor clock rate). However, this trend has come to an end, essentially for power consumption reasons. And while it was a welcome bounty for the desktop users, speeding up single operations never was a viable design principle for building really fast supercomputers. These computers were – and still are – built exploiting the principle of parallelism: combining a large number of individual processing units to work together. Due to the limits for increasing clock rates, this principle has also entered the low-end market segment: virtually all processors sold nowadays are multicore processors, i.e. integrate several computing units.

Parallel processing to get high performance

Now, exploiting this parallelism is relatively easy if we have several applications running at once: Web browser, mail client, video call, document processor etc. However, if we want to use it for a single computation, we need to think about parallelisation of the software – something which is typically much harder than writing a program for a single computing unit performing operations sequentially. We must divide the computational work into smaller pieces, distribute those to the single processing units and combine the results at the end. A typical approach for that is domain decomposition: in simulations of continuous effects such as fluids or solid matter, the domains of interest (for instance the region around an airplane) are portioned into smaller chunks which are assigned to individual processors; these however must collaborate to exchange parts of the solutions where the individual chunks have a common border. This leads to problems with data transport (as the processors could be physically far away, and data transport is a very costly operation compared to mere computing). Another problem to be solved is that the computational load of these chunks may change over time (for instance if the problem has moving boundaries), and thus processors get a different amount of work to process. This may slow down the whole process, if one processor gets overloaded and the others have to wait. Therefore, load balancing techniques need to be employed that shift work from overloaded to underloaded processors.

Building an Exascale system (while affording the power bill)

While today's standard processors are extremely fast by human standards, they are not fast enough for many applications, or not power-efficient enough. The current goal is to build computers in the Exascale range: Performing a billion billion (no typo!) operations per second (see Exascale section on the right). Would we build that from standard CPUs alone, the energy consumption of such a computer would be prohibitive, and connecting them with standard networks would not get us anywhere near the desired performance for virtually any application. The good news is that large quantities of exactly the same computations can be handled much more efficiently by specialised hardware, say adding 1000 numbers to some other 1000 numbers yielding 1000 individual numbers as sums. Such hardware is present at small scale in many off-the-shelf processors, and is available in more specialised hardware like graphics cards (GPUs used for General Purpose computing - GPGPUs), and also even more specialised accelerator devices. Modern supercomputers thus support parallel execution in many different ways and with many different devices.

Applications: reaping the performance

Applications, on the other hand, typically consist of different components and computational steps, for each of which a different type of hardware is optimal. The art (and hard work) of parallel programming is thus to cope with both the heterogeneity of the hardware and the software to make best use of the machines. One speaks of application "scalability" if it runs faster at the same rate as the hardware gets faster. When talking about Exascale systems, we are talking about theoretical peak performance. Reaching practical performance numbers close to that is very hard for a realistic application (i.e. designed to solve a problem of interest and not to demonstrate maximal performance). Also, the amount of data used by an application plays a decisive role in the design or choice of an optimal hardware system, as moving data is very expensive both in terms of time needed and in energy use.

Exascale

The speed of a computer (its peak performance) is typically measured by the number of arithmetic operations per second it can perform at maximum (floating point operations per second or FLOP/s). As of summer 2021, the fastest computers have reached a speed of 447 PetaFLOP/s (on a standard benchmark), which are 447×10^{15} (or 447,000,000,000,000,000) operations per second (see below). A system delivering one ExaFLOP/s would be about 2× faster (with at least 10^{18} operations per second). In comparison, a standard desktop computer reaches a computing performance in the 100s GigaFLOP/s range (one GigaFLOP/s is 10^9 operations per second). An Exascale computer would be roughly more powerful by a factor of 10 millon. In practice, it is very hard to get close to the peak performance for an application.

Giga, Peta, Exa & Co

As large numbers are painstaking to write and read, a system of prefixes is used to denote the number of zeros: Well known is kilo denoting a factor of 1000 (= 10^3 = 3 zeros), followed by mega (= 10^6 = 6 zeros), giga (10^9), tera (10^{12}), peta (10^{15}) and exa (10^{18}).

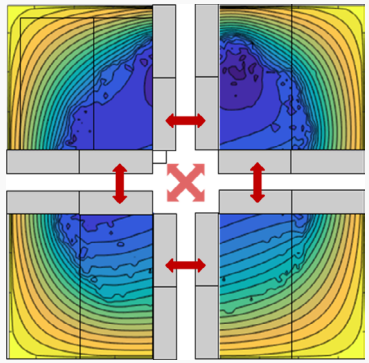


Figure 3: Solutions to a physical field equation can be approximated by decomposing the domain of interest into a mesh (tiling) of small volumes or elements (e.g. cubes – not shown). For parallel processing, this mesh needs furthermore to be partitioned into one chunk per processor (4 in this example), and computed data near the boundary (grey zones) must be exchanged between processors (red arrows). Image © Barcelona Supercomputing Centre.

THE COES AND THEIR APPLICATIONS

The HPC Centres of Excellence presented in this brochure address an extensive set of challenges. However, instead of reporting exhaustively on everything that each CoE works on, we chose to highlight some core missions, activities and results by concentrating on examples – there is much more to discover when following the links to the CoEs' web pages! A good place to start is the HPC CoE hub at hpccoe.eu.

One group of CoEs tackles the simulation of global and social systems: ESiWACE set out to bring prediction of extreme whether events driven by climate change on a global level into reach, while ChEESE tackles the question of how to predict natural catastrophes and presents intriguing successes on helping the local administration in setting up measures during the recent volcanic outbreak on La Palma. HiDALGO shows how urban air pollution can be predicted to implement countermeasures.

Engineering has been a user of advanced simulation for a long time, but is still not saturated. Aircraft need to become more efficient and use less fuel – EXCELLERAT demonstrates how to perform a simulation of an entire aircraft in flight to optimize its design. While EoCoE shows how to optimally plan wind parks in complex terrains (such as mountains), RAISE demonstrates how a clever combination of simulation and artificial intelligence can reduce the computing power needed in the case of wind parks on more level terrains, and also how to use such tools to identify good locations for geothermal energy. While we will need synthetic fuels in a future decarbonised economy, little is known about their characteristics for use in combustion engines and this is where CoEC comes in.

A field which holds great promise is computational material science: Could one design materials with desired properties in the computer instead of performing painstaking lab experiments? The promise is so big and the potential implications so revolutionary, that several CoEs tackle this question from different angles. NOMAD follows approaches using atomic-level simulation and AI-based material data base search to find new catalysts for a kind of artificial photosynthesis. TREX concentrates on a different class of highly parallelisable applications in the same field, while MaX uses quantum-level techniques to analyse material interfaces which are relevant in materials design from nano- and optoelectronics to batteries and manufacturing of the future.

Health has been a focus of human interest since pre-historic times. Now, also medicine and life science see inroads for computational approaches, which support pharmacologists, epidemiologists and medics. BioExcel operates on the molecular level, guiding the search for new drugs, or tracing the probable evolution of the COVID-19 virus. PerMedCoE works on the cell level and shows how genetic information can be used to predict the severity of a COVID-19 infection. CompBioMed takes the computational approach to the whole organ level and for instance addresses the question of how to assess the probability of bone failures for elderly people, based on individual measurements.

How the CoEs work

All CoEs address cross-cutting aspects: they support community building around their codes, provide trainings for users, and do a great job of putting the pieces together, transforming bunches of – at times - research-oriented codes used by a few highly specialised scientists on disparate platforms into something that runs efficiently on the most powerful computers, is easy to use, accesses data in an efficient way and bundles entire workflows, such that the whole technology is usable by a far broader community, both in industry and academia.

We can state that for many of the big challenges, interdisciplinary approaches are needed. Two examples for such challenges are the COVID-19 pandemic, where several CoEs have contributed to solutions, and the know-how accumulated is hopefully of use also for a future pandemic. Another example is climate change mitigation, arguably the most pressing problem of our time.

At the core of all this is the need to run code efficiently. While all CoEs are tackling this challenge for their codes (as they will explain), POP offers this parallelisation and optimisation as cross-cutting services to the HPC community at large. And this CoE is where our journey starts.

The European HPC CoE Council (HPC3)

As we have seen, each CoE has its own story to tell. Yet, while CoEs are acting in quite diverse application domains and scientific communities, they have a lot in common in terms of their overall approach: building and strengthening the HPC ecosystem in their respective domains, by increasing the maturity level of key applications and training the users. In May 2019, the 10 then active CoE established the European HPC CoE Council, supported by the CSA FocusCoE, to bundle forces for joint activities and create a common voice for the high-end HPC application community. This brochure is a joint work of the CoEs brought to you by HPC3, in the hope that you will enjoy reading and gain some insights in the often very technical world of HPC applications.

Edouard Audit
HPC3 Chair

Elisa Molinari
HPC3 Vice-Chair

Erwin Laure
HPC3 General Secretary

GET IN TOUCH

EuroHPC Joint Undertaking

POP - PERFORMANCE OPTIMISATION AND PRODUCTIVITY

THE MISSION

HPC is an essential tool for today's science and industry. But getting good performance on a supercomputer is hard and requires much specialised know how and tools. Entering POP, The Performance Optimization and Productivity CoE: It supports developers of HPC computer simulation codes with services, methods, and computer tools that check whether the code performs its calculations in the most efficient and effective way, and if not, provide hints on what needs to be improved.

BACKGROUND: PROGRAMMING SUPERCOMPUTERS

Systems used for HPC consist of hundreds to hundreds of thousands of CPU cores and other more specialised hardware. Many use different layers of parallelism: Vector units (doing the same operation to a fixed array of values in parallel), shared-memory nodes (like the multicore CPUs now commonly found in laptops), and clusters of nodes with their own memory connected by a fast network (distributed memory machines). All of these layers require a different set of programming approaches: Shared-memory programming is based on shared access to variables (like provided by OpenMP), while distributed memory machines require explicit communication as supported by MPI (message passing interface). POP partners developed specialised open-source tools that measure the application performance and help to identify bottlenecks and inefficiencies on all layers.

THE CHALLENGE

To make best the use of HPC systems, the computational work of an application has to be distributed evenly across the individual cores and in a way that minimises the necessary synchronisation and data movements, while of course always ensuring correct results. As supercomputers and their operation are very expensive, all processor cores should ideally be busy all the time and the calculations should be done in the fastest and most efficient way. Many things can go wrong here, and special expertise is needed to find performance bottlenecks, identify their root cause, and come up with solutions to resolve them. The difficulty of this task increases significantly with size and complexity of the application code and the system architecture.

CORE ACTIVITIES

POP experts assess the performance and efficiency of HPC applications in close cooperation with their code developers, and provide recommendations for improving non-satisfactory performance. POP also provides training on methods and tools for performance assessment, teaching application developers to assess and optimize their own application codes. Typical problems are load imbalance, when some cores have much more work to do than others (worst case: all are waiting for a single core doing all the work), communication or synchronisation issues (cores need to wait for distant input data), and insufficient work (there is not enough to do for one or more cores).

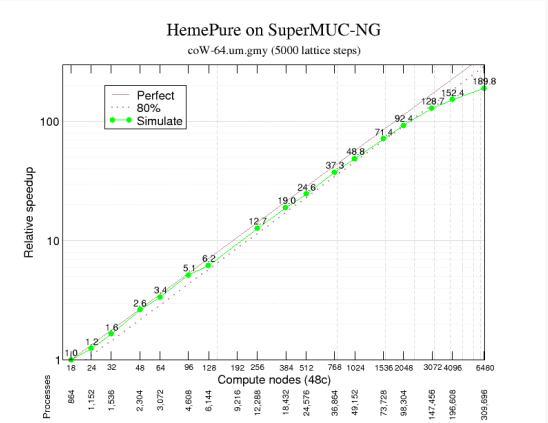
ACHIEVEMENTS

By improving and optimising HPC applications of academic and research institutions and large and small enterprises, POP allows them to either get results faster than their competitors do or to perform more calculation runs at the same time, thereby increasing the impact and value of their science or products. In the first five years of its operation, POP has assessed the efficiency of 350 HPC applications and their performance could often be doubled or tripled. In some cases, the improvements were ten-fold or even larger.

Some examples of successful services for other CoEs are:

- Together with the ChEESE CoE we analysed and optimized their Probabilistic Volcanic Hazard Assessment Work Flow package (PVHA_WF). The results were exceptional: 500× faster execution enables ChEESE experts to receive results much more quickly and to better handle much more complex scenarios.
- HemeLB is the flagship code of the CompBioMed CoE for simulation of large-scale three-dimensional fluid flows in complex sparse geometries such as those found in vascular networks. The performance of the code was analyzed and optimized on up to 309,696 CPU cores (using almost the complete LRZ SuperMUC-NG supercomputer, except parts that the analysis identified as unsuitable for this application) (see figure below).

More success stories can be found on the POP website.



The green curve shows the speedup (ratio of runtime on a single core vs. on n cores) obtained for the HemeLB code, compared to the ideal linear speedup. The code achieves good speedups but the POP analysis identified some issues with core counts over 50,000.

GET IN TOUCH WITH US

THE MISSION
CoEC supports the EU decarbonisation goals in the energy and transportation sectors applying Exascale computing technologies to advanced combustion simulation software to gain insight into the behaviour of synthetic, sustainable fuels.

THE CHALLENGE
The integration of sustainable fuels into current engine architectures and new concepts requires a detailed understanding of their combustion characteristics in real applications. Coordinating research and innovation efforts can make low and zero-carbon solutions economically viable at high Technology Readiness Levels. Although some sustainable fuels are of simpler chemical composition (like H_2), the combustion dynamics and their performance in practical applications still must be demonstrated; more complex sustainable fuels add significant difficulties. A specific CoEC challenge is the effective handling of massive data volumes and extreme-scale computing to harness the outstanding capabilities of the upcoming Exascale systems.

CORE ACTIVITY 1
CoEC includes 11 flagship codes that represent the state-of-the-art in simulation technology for combustion applications at the European level, from academic codes to proprietary industrial software. These codes cover all physical aspects involved in combustion, including detailed chemistry, turbulence, instabilities, solid-gas and liquid-gas combustion systems, pollutant formation and advanced propulsion systems based on plasma-assisted combustion and combustion of metal particles.

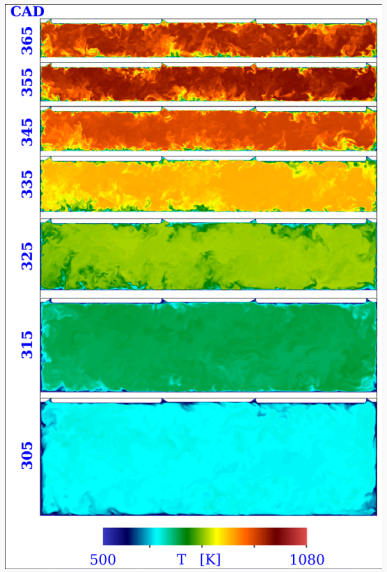
These codes use different algorithms and discretisation schemes and are of different maturity; thus the development roadmaps need to be individually adjusted. CoEC includes developers and users of these codes which pursue the progress on key scientific and technological areas that is required to accommodate these fuels in practical applications. Concrete activities include performance analysis and optimisation, and development of highly parallel algorithms and advanced simulation methodologies.

CORE ACTIVITY 2
CoEC's 13 Exascale Challenge Demonstrators (ECD) are targeted simulation studies which evaluate and improve the accuracy, reliability, and performance of the combustion codes. Focus is placed on the ability to simulate realistic engine conditions and computational performance.
A notable CoEC ECD is the use of alternative fuels, like H_2 and H_2 blends in practical systems, shedding light on the potential of alternative fuels to substitute conventional fuels for the current engine technology, providing further understanding on burner operability, flashback and lean-blow out.

ACHIEVEMENT / KEY RESULTS
One of CoEC´s success stories is the implementation of a moving mesh capability for high-order methods (spectral elements) in the code NekRS, which targets energy-efficient GPGPUs. This validates the previously mentioned ECD on the Use of alternative fuels, H_2 and H_2 blends in practical systems, as well as detailed chemistry DNS calculation of turbulent hydrogen and hydrogen-blends combustion and Flame-wall interactions.

NekRS is an Exascale ready code which achieves high accuracy with minimal data movement, while advanced algorithms, scalable iterative solvers and high-order discretizations enable the highly efficient simulation of non-reactive and reactive turbulent flows on the leading high performance computing systems. This development will permit the simulation of realistic engine conditions using the next generation of supercomputers and advance the knowledge on the application of sustainable fuels in the marine sector.

The extension of NekRS capabilities towards the simulation of time-varying geometries is essential for the direct numerical simulation of flow and combustion in engines on the upcoming Exascale systems, and was validated by simulating compression in the TCC-III engine, a four-stroke internal combustion engine with a transparent quartz cylinder and a large quartz window in the piston, enabling direct observation of the combustion processes.



Temperature snapshots at different crank angle degrees (CAD) of the middle plane of the in-cylinder flow of the TCC-III engine during the compression stage. Copyright: Christos Frouzakis (CoEC)

GET IN TOUCH WITH US

THE MISSION

As Europe moves towards a decarbonized energy ecosystem, we need precise and efficient methods to produce, store and manage clean energy. Wind power is the renewable source with the most successful deployment over the past decade; if you were setting up a windfarm, wouldn't you want to ensure your turbines are optimally placed? Simulating your entire windfarm would give you a way to, ultimately, maximize your installation's energy output. But doing so requires the use of cutting-edge computational methods to run HPC simulations of a wind farm in complex terrain and up to 100 turbines. Such a modelization is one of EoCoE's core use cases, which also include simulations applied to perovskite solar cells, hydrology, meteorology and tokamak fusion reactors, all crucial domains a successful the European energy transition.

COMPLEX TERRAIN

The increased understanding of turbulent flow in the wind farms has played a key role in the spectacular advance in wind power. To fully exploit these advances and move from a single wind turbine to the simulation of a large windfarm and its terrain requires increases in the size of the problems to be solved and the amount of computational resources of at least two orders of magnitude. EoCoE used the Alya code for performing these simulations (a flagship code for showcasing the power of Europe's next generation supercomputer).

To reach its performance objectives for complex terrain simulations, the EoCoE team ported the Alya code on GPUs, using only OpenACC. Tremendous results were achieved; for instance the team got Alya running 68 times faster for a core part (element assembly) compared to a CPU. Porting Alya meant ensuring the code would fully exploit modern GPUs, and EoCoE reached 50% of the maximum floating-point performance on an NVIDIA A100 GPU – a very good ratio. In a cooperation between Barcelona Supercomputing Centre (BSC) and the National Research Council of Italy, a fast linear solver (AMG4PSBLAS) was integrated, running 3 times faster than the standard solver used before. Thanks to work on dynamic load balancing taking full advantage of the GPU computing power, Alya GPU-only simulations now run 4.3 times faster than the previous CPU-only simulations.

In addition to GPU porting, the EoCoE team tackled the co-execution of CPU/GPU fuel simulations. The joint task force between BSC and Friedrich-Alexander University achieved a 20% improvement on CPUs - a significant result that validates the consortium's collaborative code improvement strategy. EoCoE believes that there is still more potential in the CPU/GPU combination for this type of application; a study ("mini-app") built using CUDA, an alternate framework for GPU programming, obtained a massive speed-up, as it ran 11 times faster than the OpenACC GPU-only runs.

FULL ROTOR SIMULATION

The simulation of a full rotor, with rotor blades passing very closely to the mast, causes enormous difficulty for generating a computational mesh, which was solved by the EoCoE team. Also, the turbulent flow past a rotor is very complex and a challenge for numerical solvers. The EoCoE team improved the linear system solver with new algorithmic components. In order to ensure wide uptake and provide standardized feedback loops to developers, a portable version of the Alya code with all its dependencies, designed to run on any HPC platform, was created.

Thanks to this work, a simulation of a (mid-to-large scale) turbine with 70m rotor diameter and wind velocity of 16 m/s was run on the MareNostrum IV supercomputer. The turbine's solid mesh consists of over 4.3 million elements, and its fluid mesh of close to 18 million elements.

CONCLUSION

As the second phase of the EoCoE project comes to a close, its teams can run full rotor simulations and air flow over complex terrain, and plan to keep strengthening their collaborations and improving their codes in the future, making exascale HPC simulations a fundamental tool for improving the European energy ecosystem as a whole.

Beyond the windfarm challenge, and as mentioned above, the EoCoE project applies simulations to several crucial domains for the European energy transition. In all of them, significant progress was achieved by the consortium, in line with what was detailed here for the wind simulations.

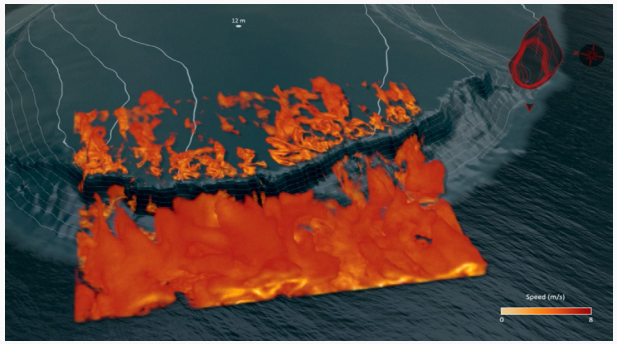


Figure 1: Detailed simulation of irregular wind velocity over complex terrain (using a mesh with 358 million nodes and 2 billion volume elements).

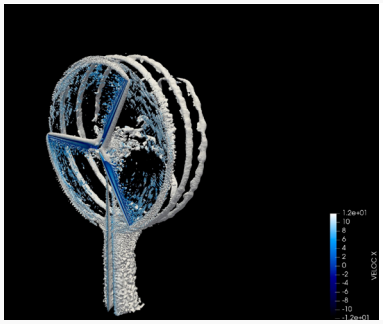


Figure 2: A challenging goal: modeling a complete wind turbine, with full rotor and wind flow (NUMAD wind turbine, using 168 million volume elements) .

GET IN TOUCH WITH US

THE MISSION

Safer cars, aircraft with lower carbon-footprints and noise pollution, more energy efficient turbines for transport and energy production – these are just a few examples of the societal benefits from engineering design with large-scale computer simulation. The engineering design aspirations require the ability to model much larger, more complex systems and to execute those models on the future Exascale supercomputers. EXCELLERAT aims to deliver on that challenge, enabling the evolution, optimisation and deployment of engineering application software on the next generation supercomputers and in doing so increasing Europe's competitiveness in the engineering sector.
EXCELLERAT is considering a broad spectrum of use cases to get state-of-the-art solutions for typical engineering problems and to show the power of the HPC resources in a real-world context.

BACKGROUND: THE ALYA SOLVER

Alya is a high-performance computational mechanics code for the solution of complex coupled multi-physics / multi-scale / multi-domain problems, arising primarily from the engineering domain. Alya includes a number of "physics" options including: incompressible/compressible flows, non-linear solid mechanics, chemistry, particle transport, multiphase problems, heat transfer, turbulence modelling, electrical propagation.

THE CHALLENGE

Preparing for engineering applications in the Exascale era requires the flexibility and adaptability of the leading-edge simulation codes (one speaks of code malleability) to match the increasing variability in the design of parallel computing devices in future Exascale systems. On the other hand, the complexity of application workflows and the expertise needed for using supercomputers raise usability issues, a particular issue for SMEs.

FULL AIRPLANE SIMULATIONS ON HETEROGENEOUS ARCHITECTURES

Achieving highly efficient full aircraft simulations on heterogeneous architectures requires attention to detail regarding the parallelisation of the numerical schemes. Minimising idle times of underloaded devices at algorithmic synchronisation points can be achieved using dynamic load balancing (DLB). The Alya Computational Fluid Dynamics (CFD) code has been provisioned with a distributed memory DLB mechanism, adapting the mesh partitions based on runtime measurements. The core parts of the approach are an efficient in-house Space Filling Curve based mesh partitioner and a runtime redistribution module to migrate the simulation between two different partitions.

This way, Alya can partition a 250 Million elements mesh of an airplane within 0.08 seconds using 128 nodes (6,144 CPU-cores) of the MareNostrum V supercomputer, i.e. without significant overhead. This approach was applied to perform full airplane simulations on the heterogeneous POWER9 cluster installed at the Barcelona Supercomputing Center, demonstrating well-balanced co-execution using both CPUs and GPUs and reducing compute time by 23% over a GPU-only exection.

ENABLING HIGH-PERFORMANCE COMPUTING FOR INDUSTRY THROUGH A DATA EXCHANGE & WORKFLOW PORTAL

The EXCELLERAT Data Exchange and Workflow Portal provides a supercomputer work environment targeting industrial users and SMEs in particular, supporting application workflow scheduling and execution as well as providing direct data analytics support (i.e numerical model and data processing in one platform). The platform enables a safe and traceable, bidirectional data transfer between the data generators and EXCELLERAT's six HPC centres, avoiding duplication and leading to time and cost reductions.

A modal decomposition toolkit has been developed for the analysis of the large amounts of data that arise from Exascale simulations. This scalable and efficient parallel toolkit is MPI-based and employs the optimised LAPACK/ScaLAPACK libraries. Highly efficient parallel input-output approaches for handling time-series data arising, for example, in transient simulations have been realised. Since different solution data formats are utilised within the engineering CFD community, a flexible modular interface has been developed to facilitate the support for simulation codes not yet covered.

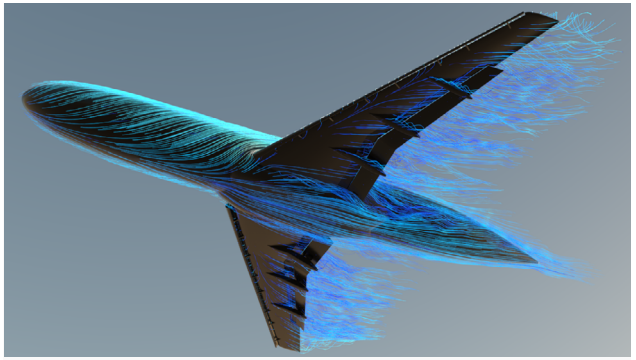


Figure 1: Simulation of aeronautical flow.

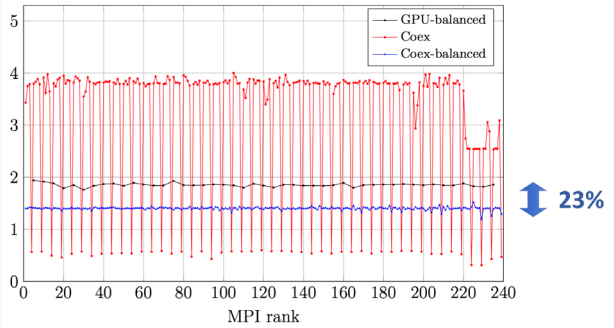


Figure 2: Comparison of time per core ("MPI rank") for unbalanced CPU/GPU co-execution ("coex") vs. balanced coex vs. pure GPU execution. Note that the execution times in the unbalanced case vary by a factor of 8.

GET IN TOUCH WITH US

THE MISSION

In the CoE RAISE, researchers from science and industry develop novel, scalable Artificial Intelligence (AI) technologies for application on next-generation supercomputers. These methods are developed in conjunction with representative use cases from Engineering and Natural Sciences, which address important societal challenges such as exploring subsurface regions suitable for geothermal energy capture or carbon storage, or optimizing wind farm layouts and their operation.

BACKGROUND: COMPLEX HYBRID WORKFLOWS

AI tasks run very efficiently on graphics processing units (GPUs) while traditional HPC methods (like simulations) usually use clusters of standard server CPUs. Nowadays, many workflows run AI and simulation tasks concurrently on different hardware and enable them to regularly exchange required data.

THE CHALLENGE

The main challenge is to find scalable, generalizable AI solutions that interoperate with traditional simulation and data processing methods and at the same time efficiently utilise heterogeneous supercomputing hardware. This includes orchestration methods using all hardware components of current, heterogeneous supercomputers, where each component is best suited for a specific set of tasks, and supporting complex hybrid workflows. CoE RAISE develops such methods and workflows to benefit important use cases, for instance for „Seismic imaging with remote sensing for energy applications“ or „AI for wind farm layout optimization“.

SEISMIC IMAGING WITH REMOTE SENSING FOR ENERGY APPLICATIONS

Climate change requires a rapid move from (hydro)carbon energy to sustainable energy sources such as geothermal sources. The primary method for discovering promising geothermal sites is seismic imaging, which relies on sound waves. Seismic imaging is applied over large geographical areas and extends deep into the Earth's subsurface. In RAISE, scientists employ such techniques to reveal the underground structure of the Earth. This information is combined with exact knowledge of the surface conditions extracted from satellite imaging data (see Figure 1). AI methods combine both techniques and extract the maximum information from the available data to facilitate the exploration of sustainable energy resources, uncover regions suitable for carbon storage, and provide corresponding means for monitoring safe and efficient operation.

AI FOR WIND FARM LAYOUT OPTIMISATION

The wind energy industry relies on advanced simulations for the design of wind turbines and wind farms, and for wind and power predictions. At this stage, the objective is to assess the potential of a wind farm and to optimise the placement of wind turbines on the terrain. Short- and long-term forecasts are necessary for gauging the wind resources available in the coming days or months/years. Simulating the flow over a full wind farm is extremely difficult and consumes a huge amount of time and computing resources. To save costs and energy, reduced-order models were developed that bridge the spatial and time scales between the detailed dynamics of the turbine blades and the wind flow across the full farm. Replacing the individual turbines featuring many geometric details by simple "sinks" for the wind velocity can yield very similar flows (see Figure 2) and is at the same time much cheaper. RAISE develops deep learning methods based on detailed turbine simulation results to model such cheap solutions.

BACKGROUND: LARGE-EDDY SIMULATION (LES)

LES is a method for simulating fluid flows that filters fine-grained physical phenomena (such as turbulence) and models them by sub-grid-scale models. LES is computationally cheaper than resolving all physical scales yet delivers sufficiently accurate results for many applications.

ACHIEVEMENTS / KEY RESULTS

In its first months of operation, RAISE has already demonstrated that the combination of traditional, simulation-based HPC methods with AI techniques produces better results with less costs, energy and time.

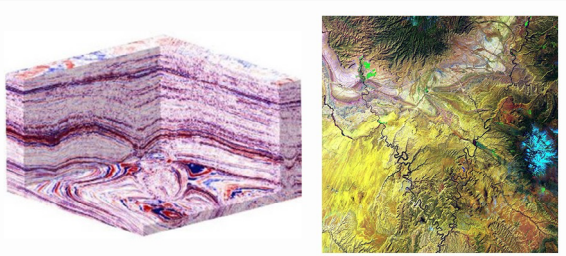


Figure 1: 3D seismic imaging volume showing the subsurface layering structures (left) and satellite image highlighting the different properties of the Earth's surface (right).

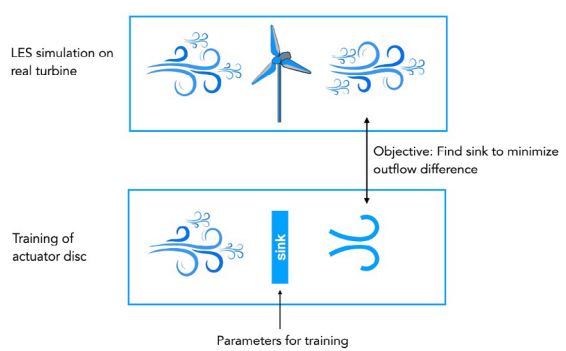


Figure 2: Training process to create a wind turbine model.

GET IN TOUCH WITH US

BIOEXCEL - COE FOR BIOMOLECULAR RESEARCH

THE MISSION

How quickly can we find a cure for COVID-19, or develop a vaccine, and adapt it to an evolving virus? Pandemics show us that speed can save literally millions of lives.

Computational modelling and simulation techniques for Life Sciences can offer that speed. Powerful and sophisticated software packages are critically important for fundamental and applied research, including industrial development, in particularly in the areas of drug design, biotechnology, food, and chemical industries. BioExcel's mission is to provide Life Sciences researchers with high-quality, user-friendly software, increase their expertise and skills, and strengthen the community as a whole.

BACKGROUND: MOLECULAR MECHANICS

Drug development in its basis is finding the most effective molecules (i.e. medicine) which can affect certain undesired processes in the body (i.e. sickness). Computers allow us make models of molecules and simulate their structure and dynamics subject to the fundamental Newton equations of motion. Analysis of this molecular motion then allows us to understand not only how the molecules interact with each other but suggest ways to control their interactions. On method is to use molecular dynamics (MD) simulations and free energy calculations of molecular interactions in order to predict which objects ("ligands") bind best to target proteins.

THE CHALLENGE

Many available applications for MD need to be adapted to make use of the tremendous power offered by the latest generations of supercomputers. Also, these tools are often too complex or fragmented for many potential users to adapt for their needs.

CORE ACTIVITY

BioExcel improves the performance, efficiency, scalability, and usability of software packages of high importance for biomolecular Life Science research. The CoE optimizes these codes for the EuroHPC systems including new European accelerator processors. Through advanced technologies for memory and storage use we design workflows solutions which are able to execute thousands of related calculations (called ensembles), a particularly effective strategy e.g. for the "screening" of large number of potential drug candidates. By making workflows user-friendly, BioExcel supports their adaptation to specific problems, making efficient use of large supercomputers. Notably, BioExcel hosts the development of several of the most widely used European HPC codes, for instance GROMACS for molecular dynamics simulations of biomolecular systems and HADDOCK for integrative modelling of biomolecular complexes.

HIGH-THROUGHPUT DRUG SCREENING AT PRE-EXASCALE LEVEL

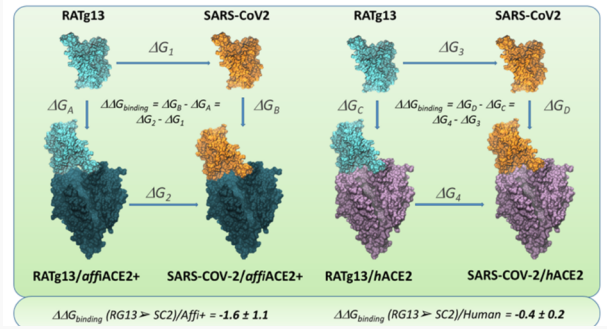
BioExcel demonstrated that workflows built on open source software packages can efficiently leverage the most advanced European supercomputers to screen hundreds of pharmaceutically relevant compounds in a matter of days instead of weeks. A showcase running 72h on 50,000 cores on the supercomputer "Raven" (Max Planck Computing and Data Facility) reached a stunning 93% of the theoretical peak performance (3.5 PFlop/s) – an extraordinary high value proving the significant potential of the toolset to massively accelerate drug discovery.

LARGE-SCALE SARS-COV2 MUTATION ANALYSIS USING BIOEXCEL HPC WORKFLOWS

BioExcel has created a set of workflows designed to answer questions around the SARS-CoV-2 spike protein and the way it invades human cells:

a) Understanding the mechanism of virus entrance into the cell and the adaptation of the virus to different host species.
b) Understanding the different sensitivity to the virus (beyond the effects caused by age) of individuals.
c) Predicting the next mutations of the virus and how it might adapt to be even more infectious.
d) Understanding how the virus has evolved by comparing its structure/genome to other coronavirus strains in different species.
e) The most remarkable results outline a unique evolutionary process in which SARS-CoV-2 gained the ability to directly infect humans after having developed an unusually high affinity for the horseshoe bat, its natural host receptor (see Figure): i.e. there was no need for the virus to adapt itself to other animal species, rather it was able to transfer directly from bats to humans.

Visit our COVID-19 page for more information.



Quantifying differences in binding to bat (affiACE2*) and human (hACE2) receptors of the SARS-CoV2 virus and its variant found in bats (RATg13), using thermodynamic cycles.

GET IN TOUCH WITH US

THE MISSION

Computational biomedicine refers to the use of computer-based tools and approaches (in silico) to simulate and model the human body in health and disease. In the European Union, this new science has become synonymous with the concept of the digital twin, a computer model of a given person. The focus of the CompBioMed Coe is on a methodological and technological framework that, once established, will enable researchers to study the human body as a single computer simulated complex system (or digital twin). The leading edge of Computational Biomedicine harnesses computer simulations that are conducted on massively powerful supercomputers. The tremendous power of these machines allows larger and more complex biological systems to be simulated, yielding better, more accurate, and more practically useful output.

THE CHALLENGE

Whilst progress has been made on various individual applications which are capable of modelling aspects of the human body (heart, bloodflow, bones etc.), combining these models into something that will model a whole human requires incredibly large computational power and expertise. In addition, the computational language for these models may be different, making it difficult to combine models to create something meaningful. CompBioMed is working with software developers (those creating the applications) and supercomputer centres and hardware developers (those creating the computers) to ensure that our applications have the best performance possible on the current and next generation of supercomputers.

BLOOD FLOW MODELS FOR PERSONALISED TREATMENTS

Blood flow models have improved dramatically in the last 5 years within the CompBioMed Centre of Excellence. With the arrival of a new generation of supercomputers, CompBioMed has adapted their models to capitalise on the performance they offer. Larger and more complex models can now be studied and the amount of data that is processed has increased. With this, we have developed high-resolution models that can simultaneously capture 3D flow in both the arteries and veins of human-scale vasculatures. The ongoing advance in this work, is to join this blood flow model to a heart model enabling the full circulation of blood to be modelled in one computational run, enabling a better understanding of the system, and increasing the possibility of personalised treatment regimes.

IMPROVED DRUG SCREENING

The drug discovery process can typically take decades and billions of dollars to identify new and effective drugs using laboratory methods and animal plus clinical trials. The process that CompBioMed has developed uses a combination of machine learning (ML) and traditional computational methods to reduce the time needed to find a drug which interacts effectively with the target theoretically by a factor of 10. By screening many prospective known drugs against a target within the virtual human body using faster ML techniques, the traditional methods of laboratory and animal trials can then work with a smaller and more promising subset of drug candidates and undertake more refined analysis of the drug's ability to affect the target. These combined approaches allow for many more drugs to be tested and ruled out computationally with good accuracy, improving the chances of finding new and effective drugs and saving the time and cost of running every candidate through laboratory and animal testing.

IMPROVED TREATMENTS BY BONE STRENGTH PREDICTION

Finally, CompBioMed are working on bone strength and fracture analysis capable of determining the individual risk of fracture and the possible medical interventions (drugs, surgery etc) that could reduce it. From hip CT scans, for example, they can analyse the forces acting upon the hip joint during walking and in case of a fall. From this, they can predict if a given patient's thigh bone can withstand those forces, or if it would fracture. Thanks to HPC, bone strength evolution can also be estimated over several years. This work is now being tested in various scenarios to determine its efficacy for predicting treatment success in patients. In the future, this work could help doctors to give patients a better-informed prognosis, with more accuracy and certainty.

KEY RESULTS

In comparison to other fields of study within the computational landscape (weather, energy, etc.), biomedicine is a relatively new field in HPC. CompBioMed has improved current applications in biomedicine as well as advanced the visibility and trust with which the applications are held within medicine. The bone and blood flow modelling simulation is already being used in collaboration with hospitals local to the research teams developing them and CompBioMed has worked with medical students to introduce them early to this technology.

Image from Virtual Humans film courtesy of CompBioMed and Barcelona Supercomputing Center.

GET IN TOUCH WITH US

PERMEDCOE - HPC TO TREAT DISEASE AT INDIVIDUAL LEVEL

THE MISSION

Imagine knowing how your body would respond to therapy even before it started. That would undoubtedly save a lot of time finding the proper treatment. Also, it would prevent the adverse effects that inadequate treatment might have on you. Isn't this what we all understand by personalised medicine?

PerMedCoE tackles this question by simulating the fate of cells and tissues configured to represent patients and their disease conditions, aiming to faithfully predict adequately how viable treatment options will affect them.

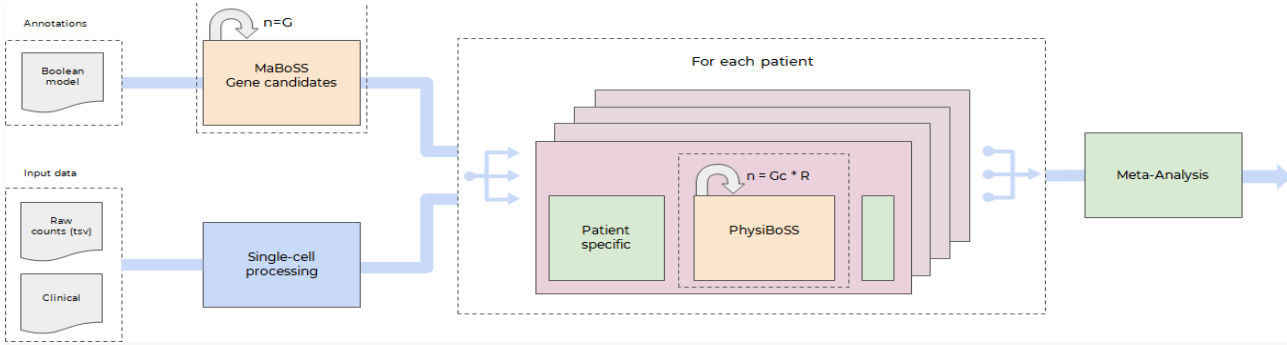
Today, simulations of this nature are just beyond our grasp, but not for long. Exascale capabilities are just about to disrupt our top-tier computing sites, and with them comes the opportunity to break through and realize this new paradigm in healthcare. PerMedCoE is preparing the tools and infrastructures to face this challenge.

BACKGROUND: DISEASE MODELING

To understand a complex phenomenon like human disease, one has to look at the systemic interactions of several layers and scales (environment, organ, tissue, cell, pathways and molecules). To deal with these sorts of problems, we need to combine data (such as genomics) with what we know of the elaborate internal workings of cellular biology (think of cells as computer chips full of inputs and output ports).

THE CHALLENGE

Today, many tools can perform simulations at individual scales (e.g. cells or organs), but they are generally not ready to fully exploit future Exascale systems. They need to be optimized, re-programmed using specialised high-performance tools, and scrutinised to find and remove bottlenecks so that they can run thousands of times faster. They also need to be put together into workflows that are simple enough to be deployed by researchers, reliable enough to be used in practice, and general enough to be repurposed as the field advances.



To handle pipelines involving heterogeneous workloads (data analysis, personalisation, multi-scale simulation), PerMedCoE uses software containers to create building blocks portable to a range of HPC resources, permitting future rapid responses to pandemics.

COVID-19 WORKFLOW

The COVID-19 severity pipeline is a good example: It takes the genetic background of a patient and uses simulations of large numbers of cells to determine the severity of a COVID-19 infection (Figure above). The workflow starts with two groups of patients (moderate vs. severe symptoms), and for each individual performs multiscale modelling from the inner workings of cells (describing death of infected lung cells, based on individual genetic information) to the complex interactions between them and with the environment, to determine whether these in-silico experiments are able to predict the disease subtype.

PREPARING FOR EXASCALE

The CoE has identified a series of tools central to its use cases, which are scaled up and prepared for operation in Exascale environments. One example is Physicell-X, an MPI powered version of the seminal cellular simulation application. Here techniques like domain partition have been applied to distribute the load across different computing nodes, which is complicated by issues like the irregular shape that the cell masses may have, or the ability of cells to migrate from one domain to the next.

A number of fundamental tools in this field have already been optimized and extended (besides PhysiCell, also COBREXA, and MaBoSS). PerMedCoE is preparing another workflow addressing tumor response to targeted therapies. The CoE has reached out to end-users and key industrial players to understand their needs and their perceived value of the offering, offers a variety of trainings on its results, and organises community benchmarking efforts, which should ensure the relevance of our core tools as well as help stimulate breakthroughs also in tools beyond the current scope of the CoE.

GET IN TOUCH WITH US

CHEESE - COE FOR EXASCALE IN SOLID EARTH

THE MISSION

Earthquakes, landslides, tsunamis and volcanic eruptions ("geohazards") pose a threat of widespread damage or loss of property and life. What if we could predict where a coast needs to be evacuated, before the Tsumani reaches the beaches? One key aspect is to use "urgent computing" approaches in order to deliver the planning support in time for effective mitigation actions to be implemented.
The aim of Center of Excellence in Solid Earth (ChEESE) is to prepare key applications for geohazard predictions to get ready for that job. One of the core ChEESE examples is the prediction of volcanic ash dispersal, which helped local authorities during the outbreak on La Palma 2021.

THE CHALLENGE

In order to be ready for urgent computing, application codes and the related workflows have to be able to effectively exploit Europe's current and future leading supercomputers – pre-Exascale and later Exascale systems. In order to keep the longer-term application code maintenance and development manageable, performance portability needs to be considered: the expectation is that future supercomputing systems will employ a variety of heterogeneous architectures and the Geohazard flagship codes must run effectively on as many of these as possible. This entails the solution of many technical challenges, including the adaptation of the software (and potentially also the algorithms implemented) to ensure effective use of the hierarchical and in many cases heterogeneous computer architectures and extending the software stacks required to deliver the workflows needed by Geohazard analysis and prediction services. A key aspect to enable this is collaboration with the other European HPC ecosystem actions in the sense of co-design encompassing application capabilities and needs.

CORE ACTIVITIES

ChEESE works with 10 key geohazard codes forming the core of 12 pilot demonstrators for several geohazard challenges. Multiple layers of parallelism are exploited (thread parallelism combined with task parallelism in conjunction with vectorization and fine-grained parallelism for use of accelerators) and at the same time, the possibilities or constraints of memory hier-archies, interconnects and input-output features are taken into consideration. Generally, the maturity of the demonstrators is increased, to be applicable for production use. Pilot Exascale Demonstrators are tested on pre-Exascale hardware prototypes.

ACHIEVEMENT / KEY RESULTS

ChEESE´s service for urgent computing was used during the Cumbre Vieja volcano eruption on La Palma Island in the Canary Islands (September to December 2021).
Using the MareNostrum 4 supercomputer based at the Barcelona Supercomputing Center, ChEESE was able to cover a number of possible eruption scenarios.
With the resulting forecast information, local authorities could make informed decisions such as rerouting flights that might be affected by volcanic ash, as well as implementing confinement orders based on the air quality.
Another success story is ChEESE´s live demo of its pilot demonstrator on Faster Than Real-Time (FTRT) Tsunami Simulations, which was coordinated between Instituto Geográfico Nacional (IGN), the Spanish Tsunami Warning Center and ChEESE partners from the University of Málaga.
Using ChEESE flagship code Tsunami-HySEA, the FTRT workflow was performed in the framework of a real Tsunami Early Warning System using the computational resources of the MARCONI100 supercomputer. This, together with post-processing, also in real time, made it possible to provide alert levels in coastal segments and automatic message generation, representing a great advance that National Tsunami Warning Centers could incorporate in order to make better and earlier decisions.

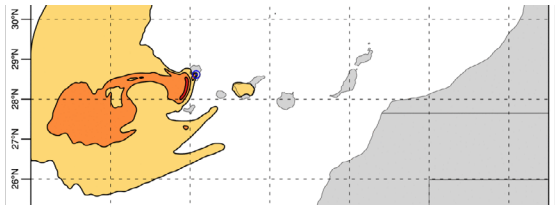


Figure 1: The ChEESE volcanic ash forecast running on Mare-Nostrum 4 computes the circulation of ash in the atmosphere and the gases emitted by the volcano in the hours and days after the eruption.

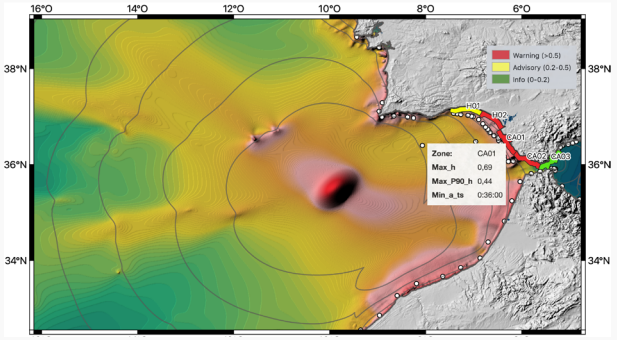


Figure 2: Tsunami warning map obtained from the results of the 135 simulations performed by including variability in the parameters of the seismic source model from the reference scenario. Forecast points and target areas are colored by alert level.

GET IN TOUCH WITH US

THE MISSION

The Earth's atmosphere is about 1.1° C warmer today than in preindustrial times, and the trend is likely going to continue. How does this translate into the local weather, and especially extreme events, in a future climate?

Predicting, mitigating and adapting to global climate change down to the regional and local scales is one of the biggest challenges of this century. We need a new generation of efficient and optimised storm-resolving weather and climate models to find reliable answers, and to predict extreme weather events triggered by global climate change.

BACKGROUND: GLOBAL STORM-RESOLVING MODELS

Global storm-resolving weather and climate models represent the atmosphere as a very fine mesh with grid spaces of only a few km and use a simulation time step short enough to resolve vertical energy transfers. This allows an explicit calculation of small-scale processes such as cloud formation and dynamics during heavy precipitation events.

THE CHALLENGE

For practically useful predictions, global storm-resolving weather and climate simulations need a throughput of at least one simulated model year per day of computing (SYPD). Moreover, these simulations produce massive amounts of data that need to be processed, analysed and stored. To leverage the available performance of future (pre-)exascale systems we need a concerted effort to

(1) apply new technologies for code optimisation to handle diversity and heterogeneity of computing architectures and improve data handling and storage solutions, making prediction simulations as efficient and user-friendly as possible, and
(2) train scientists in the weather and climate research communities in the use of those technologies.

PERFORMANCE AND SCALABILITY

One of the key activities in ESiWACE2 is pushing Europe's leading weather and climate model codes to higher throughputs in their global high-resolution configurations by improving their performance and scalability.
ESiWACE targets the European flagship model codes: ICON, IFS, EC-Earth and IPSL-CM. These models are able to compute not only atmosphere processes but also ocean processes and the interactions of both. The aim is a maximum possible spatial resolution that still yields a throughput of 1 SYPD.

COMMUNITY SUPPORT

ESiWACE is building tools to mitigate the effects of the high-resolution data deluge: within simulations, in the storage system and in the post-processing phase, and closely collaborating with the HPC industry. For model developers, ESiWACE is offering a number of services: special software packages, public HPC services and training programmes support the weather and climate community on the path to Exascale computing. ESiWACE evaluates and establishes new approaches to programming such models through a separation of the physical simulation code from parallelisation and optimisation techniques and evaluates the use of Artificial Intelligence and Machine Learning to improve models.

ACHIEVEMENT / KEY RESULTS

Within the DYAMOND Initiative (a framework for the intercomparison of global storm-resolving models), ESiWACE has demonstrated that the European weather and climate model codes are in the top league of storm-resolving models - simulation output can barely be distinguished from satellite images ("Palmer-Turing" test see linked paper).

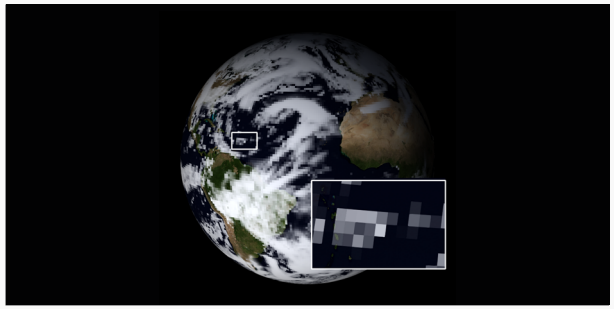


Figure 1: Clouds simulated for a day in February with an atmospheric model at 80 km horizontal resolution, a grid spacing commonly used in current climate simulations for the popular Coupled Model Intercomparison Project (CMIP).

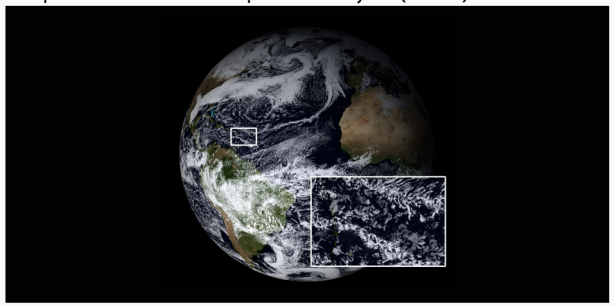


Figure 2: Clouds on a day in February simulated by a global model at 2.5 km resolution as facilitated by ESiWACE. In this "DYAMOND simulation" the atmospheric circulation is represented in far greater detail, allowing for the distinction of different cloud structures and cloud types. The weather situation differs from Figure 1 as the two simulations were initialised with different data sets.

GET IN TOUCH WITH US

DYAMOND Initiative

Stevens, B., Satoh, M., Auger, L. et al. DYAMOND: the DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains. Prog Earth Planet Sci 6, 61 (2019).

THE MISSION

Today, society faces a growing number of urgent problems of global scale: how to reduce the pollution in cities or, at least, mitigate the problems it is causing? How to address conflicts and support displaced people? Are we able to understand what happens in social networks? HiDALGO implements solutions that address Global Challenges by simulating what is expected to happen, and what effects specific mitigation measures would have.

THE CHALLENGE

Addressing Global Challenges requires a multidisciplinary approach for creating complex models (e.g. combining sociological aspects with physical ones) and implementing them on supercomputers. Agent-based systems, physical simulations (e.g. CFD simulations), real-time data streams, machine learning/AI and data analytics are elements of such models leading to complex workflows combining these disciplines with coupled models, in innovative and efficient ways. The large areas involved, high number of people affected and the required fine resolution require highly scalable codes on large supercomputer systems.

SCALABLE TOOLS FOR AGENTS-BASED MODELS

HiDALGO did focus on four global challenges: FLEE (for migration), FACS (for COVID-19 spread, Figure 2), EigHist (for social networks) and UAP (for air pollution, Figure 1). The CoE developed applications and software tools for these scenarios, and performed validation and benchmarking on leading supercomputers. In the case of FACS and EigHist, HiDALGO started from scratch or from basic codes, and produced complex, well parallelised applications. The CoE improved scalability and performance (e.g. by fitting the CFD solver in the UAP to run efficiently on GPGPUs), and increased their accuracy (e.g. by using detailed meshes for urban air pollution).

HYBRID WORKFLOWS FOR COUPLED MODELS

Capturing the mechanisms behind global challenges and creating models for actual use by decision makers or responders requires the combination of different approaches, and the integration of expert software into complex workflows with user-friendly interfaces. Simulation steps (like CFD or weather forecasts) are combined with data analytics for pre- and post-processing, visualisation, and machine learning/AI models. These workflows use an orchestrator which supports supercomputers and Cloud resources as well as data management. Coupling mechanisms support the combination of different data sources and simulations (like CFD and pollutant chemistry). Finally, the Fabsim tool allows running large ensembles for different input data, in effect enabling to calculate the probability of different outcomes.

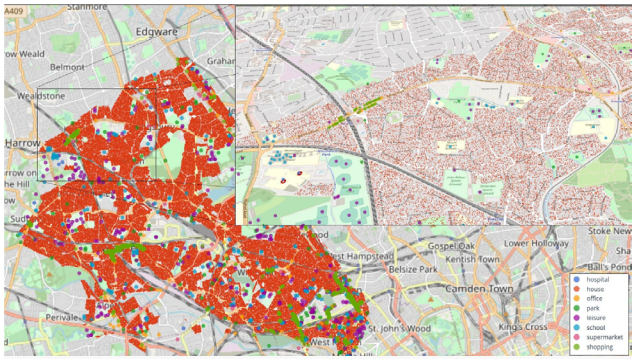


Figure 1: This simulation of COVID-19 spread in a part of London was created by the HiDALGO Flu and Coronavirus Simulator FACS, an agent-based model which incorporates SEIRDI (Susceptible-Exposed-Infectious-Recovered-Dead-Immunized) states. It approximates viral spread on the individual building level, and incorporates geospatial data sources from OpenStreetMap. COVID-19 spread is modelled at local level, providing estimations of infections and hospital arrivals, given a range of public health interventions. FACS also supports the modelling of vaccination policies, as well as the introduction of new viral strains and changes in vaccine efficacy.

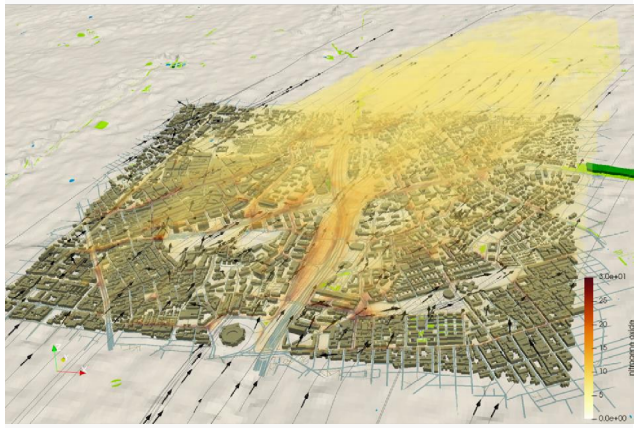


Figure 2: 3D visualisation of air pollution for the city of Madrid, depicting the concentration of nitrogene oxide (NO_x) and the prevailing wind directions. Data was produced by a HiDALGO UAP workflow, computing the detailed air flow through the urban 3D topography based on macroscopic weather data, and a coupled pollutant model predicting the creation (by a traffic model) and dispersion of NO_x. Reduced-order modelling techniques were used to avoid the need for a continuously running a huge CFD simulation; this key innovation makes 24×7 city-scale pollutant forecasts affordable. It is planned to couple the forecast results to a traffic control system, preventing peaks of pollution.

GET IN TOUCH WITH US

THE MISSION

How to reduce CO_2 emissions and best harvest solar energy? How to design next generation batteries or semiconductor chips? How to enable superconductor-based devices, or solid-state quantum bits? Materials are crucial to science and technology, to industry, and to our society. The combination of powerful supercomputers with highly accurate quantum mechanical simulations drives a new paradigm, whereby researchers can predict and screen complex materials and properties "in silico". MaX is enabling world-leading electronic structure codes for Exascale computing, with the goal of greatly accelerating design, characterisation and optimisation of materials for the above uses and for manufacturing the technologies of the future.

THE CHALLENGE

Materials simulation codes must deliver the accuracy needed to sustain or streamline experiments; to address the complexity of real-life conditions; to exploit and drive the hardware evolution; and to leverage all data produced in computations and experiments. They must be verified and validated. Their capabilities must be integrated in workflows that provide highly curated data, and are easy to use for scientists and engineers. To keep pace with future HPC systems and "post-Moore" technologies, MaX codes must become modular and be ported, scaled, and optimised for the upcoming Exascale architectures, while keeping their scientific communities engaged and active.

CORE ACTIVITY

MaX focuses on a set of complementary flagship codes for the simulation of materials at the quantum level: Quantum ESPRESSO, SIESTA, Yambo, FLEUR, cp2k, and BigDFT. These are open-source community-developed software for electronic structure calculations, with a large world-wide user base (see Figure 1). In addition, MaX contributes to a materials informatics framework to manage, store, share, and disseminate complex scientific workflows and curated datasets. Tools like AiiDA are essential to further grow the use of advanced quantum simulations in science and in particular in industry.

As discussed in the introduction, the emergence and rapid evolution of accelerators (like GPGPUs today) and heterogeneous systems, complemented by disruptive new approaches to computing, requires significant application development. MaX focused on modularization and restructuring of its flagship codes, separating software layers close to the scientific developers from layers close to the hardware, thus minimising the changes required to support different hardware. Available programming models were carefully analysed to understand which solutions fit best and how to combine them. A domain specific library dedicated to performance portability (deviceXlib) was designed and developed, in contact with vendors, to abstract common CPU and accelerator operations. This library now targets AMD and Intel CPUs and GPGPUs, NVIDIA GPGPUs and ARM plus IBM CPUs.

All MaX flagship applications now run production-ready for GPGPU-accelerated machines. As an example, Yambo (the MaX code focused on excited state calculations), was able to exploit a large fraction of the Marconi100 machine to compute quasi-particle energies for a graphene-based interface which is relevant for innovative opto-electronics and battery applications.

Figure 1: Global impact of the MaX flagship codes: bubbles represent the number of indexed scientific publications that use and cite the codes in the period 2019-2021 (over 10,800 in 110 countries worldwide, over 3,200 in Europe).

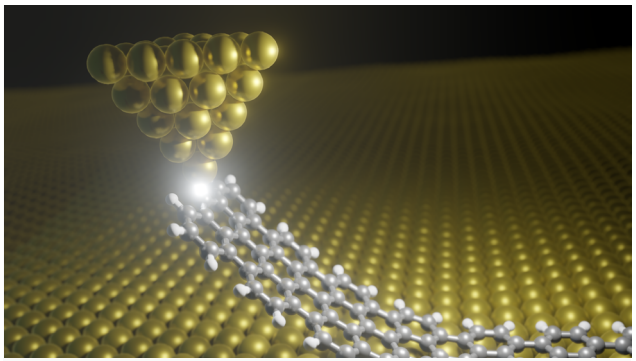


Figure 2: Graphene nanoribbons (GNRs) are strips of graphene less than 50 nanometres wide. They have been known to possess very interesting semiconductor properties (for experts: a highly tuneable band gap), and have the potential to lead to more efficient and colour-tuneable light sources. The picture depicts a 7-atom wide GNR suspended by the tip of a tunnelling microscope over a gold substrate; when a suitable voltage is applied, the ribbon emits light at a high intensity, and colour can be controlled by voltage variations. MaX codes played a significant role in enabling the experimental proof, and will no doubt guide the further development of GNR-based technologies.

GET IN TOUCH WITH US

NOMAD - NOVEL MATERIALS DISCOVERY

THE MISSION

What if we found an efficient way to split water into hydrogen and oxygen, just using sunlight, in a sort of "artificial photosynthesis" process? That would enable sustainable large-scale production of green H_2 as a renewable and clean energy carrier, a major step towards a carbon-neutral economy. Providing computational tools to find good catalyst materials for photo-catalytic water splitting is one of the key application examples of the NOMAD CoE, which develops numerical and artificial intelligence tools to model complex and realistic materials for industrially-relevant applications (in particular related to energy and environment). Another one is the direct transformation of waste heat into electricity, through novel thermoelectric materials.

WHAT ARE CATALYSTS?

Catalysts are substances which increase the rate of a chemical reaction without being consumed by it; the catalyst remains unchanged. Depending on the rate of reaction, very small amounts of catalysts can facilitate the reaction of large amounts of substances.

THE CHALLENGE

The properties of candidate materials can be simulated based on their atomistic structure in silico, and suitable candidates for a particular purpose (for instance, catalysis for hydrogen production) can thus be identified. In practice, the "search space" (number of possible materials) is huge, and these simulations require lots of time and compute power. One of the mathematical/computational methods used is DFT (density functional theory), and the computing effort for calculating a solution is proportional to the third power (N_3) of the number of atoms involved. That is, if we double the number N of atoms, we need 8× more time (or a 8× faster computer if we want to do it in the same time), and if we have 10×N atoms, we need 1000× the time. This behaviour severely limits the complexity of substances which can be simulated.

ATOMIC SIMULATION TO PREDICT MATERIAL PROPERTIES

The so-called Eigensolver (see below), a core part of the DFT method, dominates the computing effort, and NOMAD is making this step as efficient as possible and able to run on leading edge supercomputers. The open-source Eigensolver ELPA, originally developed in a collaboration between the Max Planck Computing and Data Facility and NOMAD, is the most efficient and best scaling implementation, delivering immense computational savings in DFT calculations. The 2021 release of ELPA works well on the fastest current supercomputers, including those using GPUs from NVIDIA, Intel, and AMD as well as the ARM-based Fugaku system, one of the fastest supercomputers available to date.

DFT AND EIGENSOLVERS

The interaction of atoms in a crystalline solid is governed by the so-called many-body Schrödinger Equations. As these equations are hard to solve, the idea of Density Functional Theory is to use an approximation based on electron density, while retaining sufficient accuracy to gain insight into the properties and phenomena exhibited by real solid matter. The resulting systems can be solved numerically by so-called Eigensolvers.

NOMAD AI TOOLKIT TO FIND INFORMATION ON EXISTING MATERIALS

In practice, it is simply not tractable to simulate all possible materials and thus for sure identify the best one for a given purpose. However, the properties of many different materials are already known by experiment or simulation. Therefore, how can we exploit this immense corpus of knowledge in our search? Here is where the NOMAD AI Toolkit comes into play: it uses the latest artificial-intelligence techniques (including machine learning, deep learning, and compressed sensing) to analyse all available material data, in order to identify correlations, patterns and structures in the datasets themselves, and to consequently detect trends and anomalies. Thus, the NOMAD AI Toolkit – which can be operated via a Web interface – enables scientists and engineers to decide which materials are useful for specific applications, or which new materials should be the focus of further, detailed study.

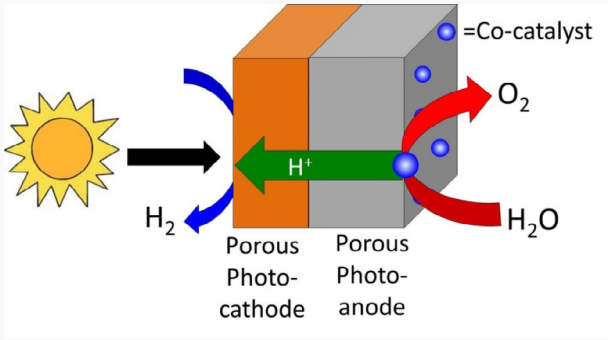


Figure 1: Identification of a better catalyst for water splitting has enormous potential in carbon-dioxide management, in hydrogen-based energy production, and many more industrial applications.

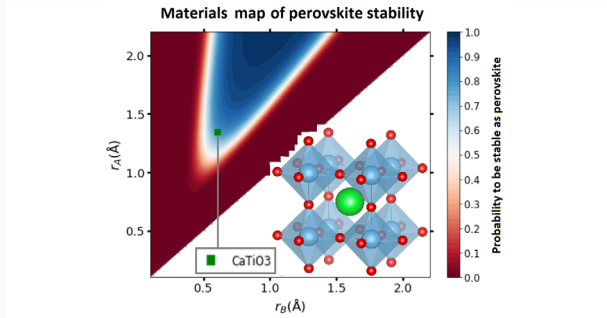


Figure 2: Perovskites are an important class of crystallic materials used e.g. for photovoltaic cells. The novel AI method SISSO permits to predict with high accuracy whether a given material combination crystallizes as perovskite.

GET IN TOUCH WITH US

TREX - TARGETING REAL CHEMICAL ACCURACY AT THE EXASCALE

THE MISSION

Computer simulations of quantum systems have become an indispensable tool in chemistry, physics, biology and materials science and the arrival of Exascale computers in the coming years has the potential to increase that utility even further.

The TREX Center of Excellence aims at developing, promoting, and maintaining open source high-performance software solutions in the field of quantum chemistry and condensed matter physics, ready to take advantage of upcoming exascale architectures.

BACKGROUND: QUANTUM MONTE CARLO TECHNOLOGY

In materials science, the ability to optimize materials properties often requires a deep understanding of the relationship between chemical-atomistic structures and the physical properties of the material itself, both in its solid and fluid phases. Quantum Monte Carlo (QMC) methods allow for reliable calculations to better predict materials properties at the chemical accuracy. This is useful not only for materials science but also for a fundamental microscopic description of the underlying physical processes. At the same time, QMC approaches are among the few methods in the field of quantum simulations that have the potential to fully exploit the massive parallelism of the upcoming Exascale architectures.

THE CHALLENGE

In order to realize the potential of computational quantum chemistry and condensed matter physics, existing software needs to be radically redesigned so that it can fully exploit the massively parallel architectures on which Exascale computers are based. The QMC family of methods is likely to play a major role in this redesign effort, but leading QMC developers must be brought together with mainstream computational scientists and high-performance computing (HPC) experts to deliver such advances to the research community.

TREX LIBRARIES FOR SOFTWARE DEVELOPERS AND HPC EXPERTS

TREX gathers European scientists, HPC stakeholders, and companies working on quantum chemistry and condensed matter simulations in the framework of stochastic QMC methods to develop and disseminate these unique computational tools for new materials design and the understanding of the fundamental properties of matter.

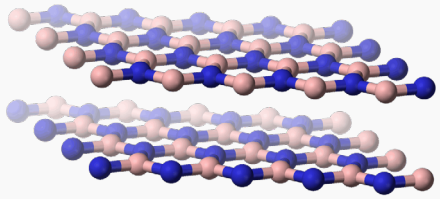
Planned software development targets eight different quantum chemistry codes. Key QMC algorithms are implemented in open-source libraries and optimised for upcoming Exascale systems, ready to be integrated into quantum chemical codes which thus could benefit from the Exascale transition.

PRE-EXASCALE DEMONSTRATOR APPLICATIONS

Applications developed by TREX target materials for energy conversion, quantum magnetism, high-temperature superconductor and functional materials.

Very recently, high (close to room) temperature super-conductors have been found in hydrogen-rich materials at very high pressures. At ambient pressure, that could revolutionize our electronic-based technology. In these materials, the quantum nature of hydrogen and its influence on their vibrational properties are the essential "glue" for superconductivity, yet its quantitative prediction is a challenge. TREX has been able to devise new algorithms and software to include these effects into the material description, and demonstrated this by computing the phase diagram of high-pressure pristine hydrogen, a paradigmatic system for hydrogen-rich materials, with an unprecedented accuracy, thanks to a combination of quantum Monte Carlo, advanced sampling of quantum nuclear trajectories, and machine learning approaches.

In the domain of functional materials, a scientific challenge undertaken in TREX project is the investigation of van der Waals (vdW) interactions in out-of-equilibrium, strained or electronically excited, materials. These interactions are key in binding molecules to form larger structures and play a decisive role in stabilising layered nanostructures or biomolecules such as DNA or enzymes. Yet, they are of pure quantum nature, and conventional quantum chemistry methods are typically not adequate to describe them in out-of-equilibrium systems. TREX has developed computational tools ideally suited for investigating vdW interactions in these „difficult" systems, and calculations on molecular systems have revealed non-trivial mechanisms and changes in vdW interactions when molecules are kicked out of equilibrium. This finding may aid designing nanostructures with high phosphorescence quantum yields or novel optoelectronic devices.



Predicting the structure and the functionality of layered materials, such as boron nitride (shown above), or graphene, requires an accurate treatment of van der Waals forces, enabled by TREX' key computational components.

GET IN TOUCH WITH US

Project Coordinator Contact: scapos AG | Schloss Birlinghoven 1, 53757 Sankt Augustin | Germany

Layout: High-Performance Computing Center Stuttgart (HLRS) | Nobelstraße 19, 70569 Stuttgart | Germany | Tel.: +49 (0)711 68587269

FocusCoE has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement N° 823964.