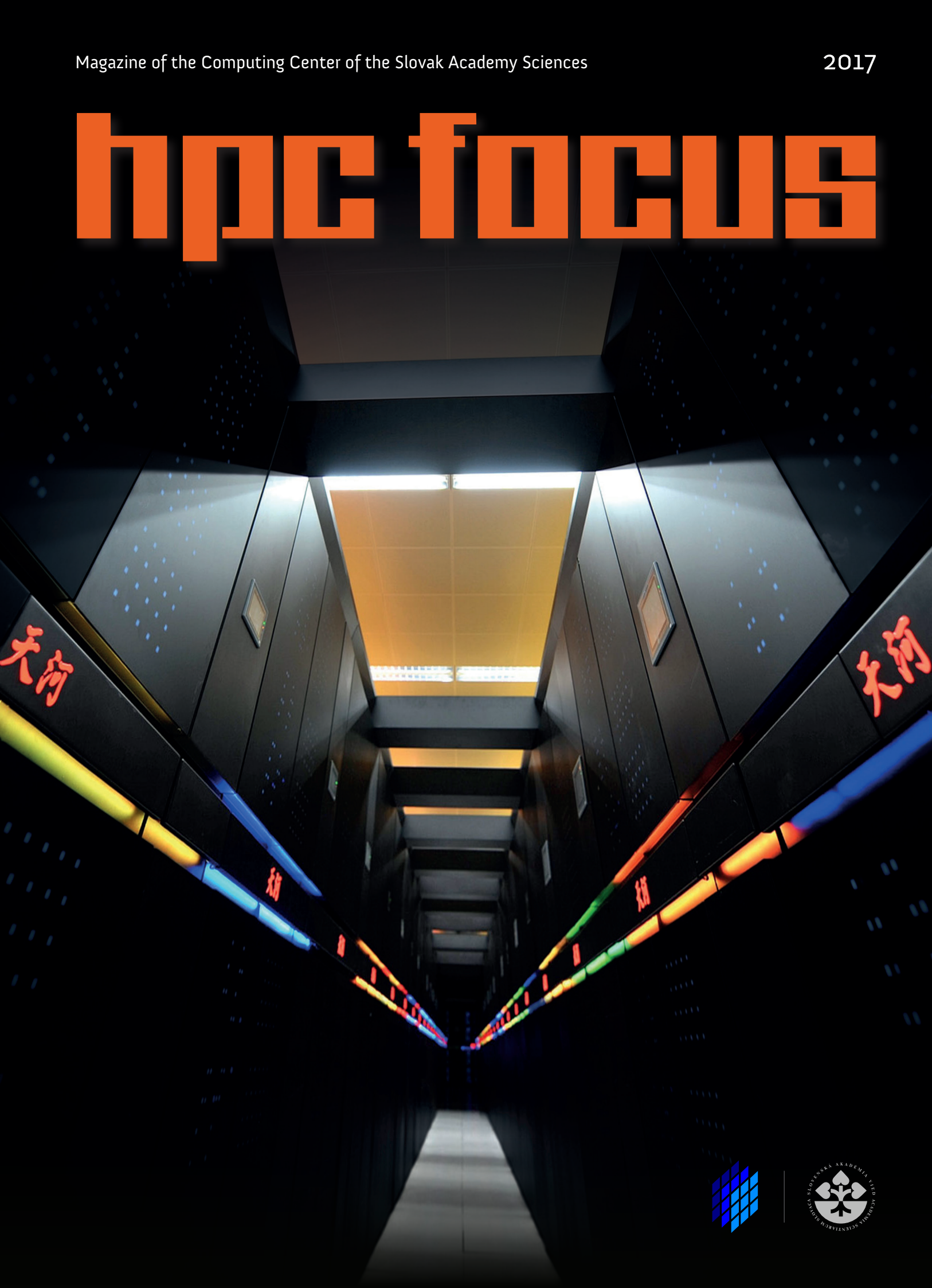
Magazine of the Computing Center of the Slovak Academy Sciences 2017

hpc focus



5th ANNIVERSARY OF THE SUPERCOMPUTER AUREL

Mgr. Lukáš Demovič, PhD.
Director of CC SAS

The introduction of the director

Dear readers,

On October 19 2017 it has been exactly *five years*, since the most powerful Slovak supercomputer Aurel became accessible to its users. A lot has happened during this time - we tackled the "childhood" problems with the help of our contractors, IBM and Datalan, and organised a number of trainings for our users, taught directly by the supercomputer's developers from Poughkeepsie. We successfully established ourselves in the European HPC community by joining the PRACE organisation, which allows our users to access the most powerful HPC devices in Europe and take part in numerous activities and events organized by this institution.

As we all know, we also ran into issues, mostly due to the lack of finances for Aurel's production, which resulted in a partial downtime of some parts of the system - however Aurel as whole has never been completely shut down and our users could always use it under the same terms. The differing opinions of the Computing Center's management and the SAS presidium on Aurel's operation were also a long lasting complication, which resulted in the withdrawal of the previous Computing Center's director. Considering the sizeable scientific output of the project, it becomes clear that it was an obvious success - more than 170 active slovak scientists have shown that they can not only fully utilize the supercomputer, but most importantly use it for purposeful work. This is confirmed by the more than eighty five percent utilization and almost ninety publications in international scientific journals, but also the constant interest of new users.

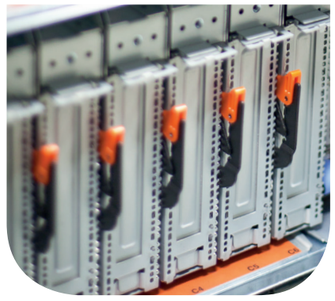
You may think that five years is not much, but in the world of IT (and especially HPC) this is basically ages and as every machine becomes obsolete, the supercomputer does twice as fast. Progress is moving forward unbelievably quickly and new devices are not only more powerful than Aurel, but much more energy efficient. User's needs evolve as well and new scientific disciplines move in with different requirements than Aurel's original specifications. It is therefore time to think about future of the SIVVP project and to look for support and funding of the next generation supercomputer. As the director of the SAS Computing Center, I plan to invest maximum effort into this goal. I would also dare to claim that we have the support of the SAS leadership and the highest authorities have also expressed agreement. I hope that in the next edition of HPC Focus I will be able to introduce the framework of the next stage of the project.

I would like to use this opportunity to thank the late director of CC SAS Tomáš Lacko for a lot of hard work that went into the creation and implementation of the SIVVP project, but also the partners from the SAS Institute of Informatics, SAS Institute of Experimental Physics, Slovak University of Technology in Bratislava, Žilina University, Matej Bel University in Banská Bystrica and Technical University of Košice. Next, I would like to thank my colleagues from the HPC department, who are in charge of Aurel's operation and user support, and also the contractor companies Datalan and IBM, who continue to help us with various technical issues. A great deal of gratitude goes to the SAS management for the support of the project as well as fair and constructive discussion about its future. Finally, we are most grateful for the interest and support of our users from SAS institutes and slovak universities, who give meaning to all of our efforts.

To you, our readers, I wish pleasant reading of this year's edition of HPC Focus and if you are interested in more information about HPC, feel free to contact us. We are here for you!

Lukáš Demovič

CONTENT



Európska únia

Agentúra Ministerstva školstva, vedy, výskumu a športu SR pre štrukturálne fondy EÚ

We support research activities in Slovakia.
The project is co-financed from EC sources.

01

HPC

IVAN PLANDER

Massive Parallelism and High Performance Computing

Parallel high performance computing is tightly connected to supercomputers. Supercomputers are very powerful computers built on parallel architectures. According to Flynn taxonomy two computer architectures belong to parallel architectures: SIMD (Single Instruction- Multiple Data) and MIMD (Multiple Instruction-Multiple Data). The SPMD architecture (Single Program-Multiple Data) represents the prevalent style of parallel computing where all processors use the same program and every processor has its own data. Computer architecture SIMD, oftentimes implemented in supercomputers, allows great computing speeds, thanks to massive parallelism. The methods of solving scientific and technological problems are a part of computational science. Mathematical models of these problems are usually described by systems of linear and nonlinear equations or systems of ordinary or partial differential equations and stochastic relations. Solving these models in closed form is usually not possible, numerical methods usually have to be used with the help of supercomputers. Parallel architectures of computers and computing algorithms belong to the theory of high performance computing.

Computers and information science develop in three main directions: the development of new components and technologies, the developments of computer architectures and applied artificial intelligence.

Parallel computer architectures are represented by three types of parallel systems:

- Supercomputers
- Computer clusters
- Computer networks

Supercomputers based on high parallelism of closely tied processors are the most important for scientific and technical computations. Computer architectures are highly developed and diversified. In the past most programs ran on conventional single-core processors, capable of processing a single thread at a time. Lately, these simple processors are being replaced by multi-core CPUs containing accelerators capable of high performance universal computing. In the past few years a new class of HPC (High Performance Computing). systems appeared. These systems use unconventional processor architectures such as Cell processors or programmable arrays i.e. GPU (Graphics Processing Units) – for "demanding" tasks and conventional CPUs for non-computationally-intensive tasks such as input/output, communications etc.

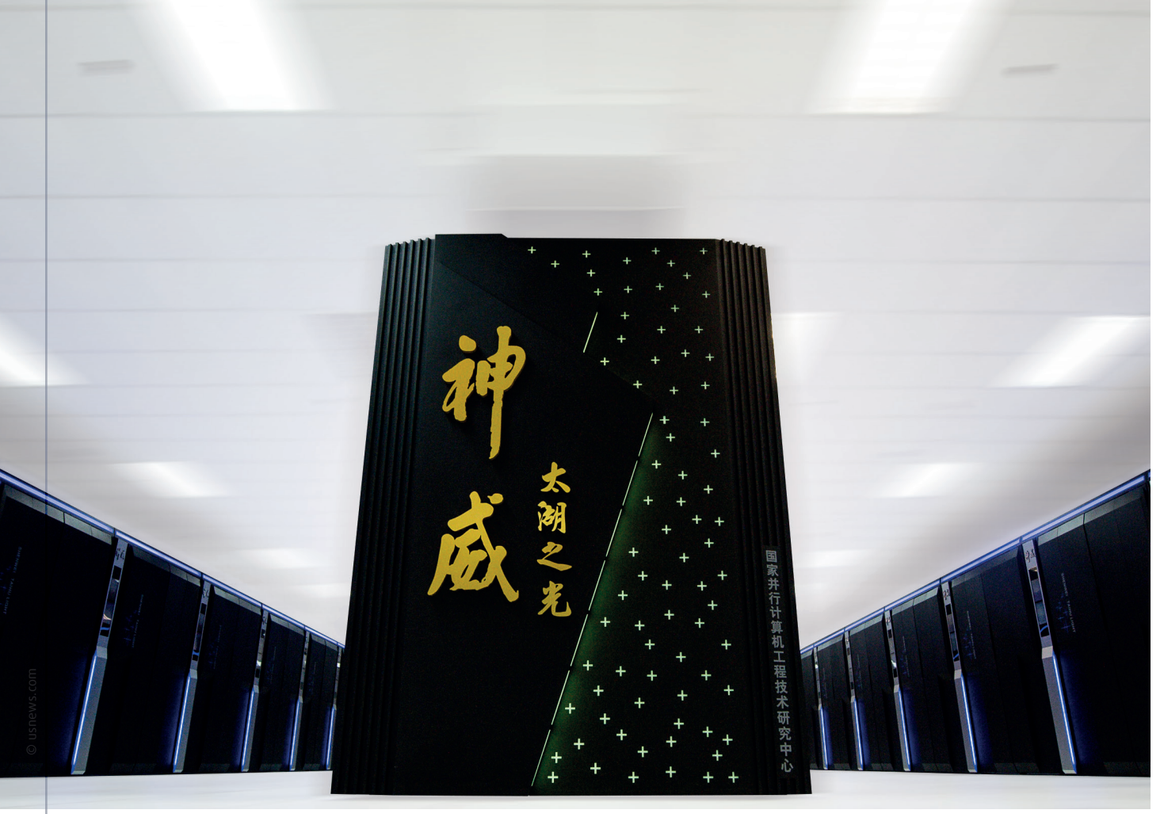
Computer clusters offer comparable computing

performance as supercomputers with much lower production costs. The development of PVM (Parallel Virtual Machine). software played an important role during the development of computer clusters along with the Unix operating system. This open source software allowed for the creation of a virtual supercomputer – basically high performance computing cluster, created from personal computers connected together. Heterogeneous clusters standing on the top of this model reached overall performance comparable to the throughput of the most expensive supercomputers at a much lower price. Beowulf type cluster is capable of performing tightly fitting parallel high performance computations on the same level as a supercomputer. Grid structure and grid paradigm is often compared to an energy network. Grid allows seamless resource virtualization in a way that allows users to access practically limitless computing cycles and data memory. The infrastructure assigning physical CPUs and memory is hidden from the users. Similarly, individuals connected to an electrical grid do not usually know which power station, generator or transformer supplies them with electricity.

Grid structures and efforts to develop their applications have always been connected to the development of middleware, especially through multidisciplinary teams. One of the fastest growing areas of Grid computing is biology, computational biology, bioinformatics and computational neurosciences and other areas containing grid technologies, like the process of accessing, gathering and mining data. Grid provides an important platform for the realization of resource-intense scientific and technical applications with higher cost efficiency. Physics is another field of science with quickly developing grid applications. For example the Large Hadron Collider in Cern (Geneva) generates a large amount of data, with every physical event resulting in 1018 ways parallelism. Commercial applications are equally important.

Contemporary scientific achievements in computation and communication hardware and software are causing dramatic changes and acceleration in all areas of computing. Computational complexity in science, technology and economics is quickly growing. The complexity of simulations executed on supercomputing systems requires high speed of processing and high level of parallelism (systems with thousands and millions of processing elements).

Dr. h.c. prof. Ing. Ivan Plander, DrSc., Is one of the founders of the Institute of Technical Cybernetics SAS (1956), author of the first analog computer in Slovakia (SAS, 1957), author and principal coordinator of the RPP-16 project and initiator of the research on artificial intelligence and robotics in SAS.



Numerous scientific and economics applications require linear algebra optimization to solve large simulation problems, like multiplication and inversion of large matrices (100 000 x 100 000 etc.).

A new class of high performance supercomputing systems appeared in the recent years. These systems use non conventional processor architectures like cell accelerators or graphical processing units (GPU) – for complex calculations and conventional central processing units – usually for less intense jobs, such as input/output, communications etc. General purpose graphical processing units (GPGPU) were developed for scientific calculations. Graphical processing units (GPU) were an integral part of most personal computers and gaming consoles. Greater visual realism of games required the development of GPUs. They evolved from simple accelerators of 2D designed applications based on graphics into powerful units required by 3D games. The computational power of modern GPUs created effort to use them not only in graphics, but also in numerically intensive computing. This interest was demonstrated by the creation of general purpose GPU coprocessors or GPGPU by such manufacturers as NVIDIA (Tesla 2011) and AMD (Radeon 2011). The GPGPU processors were quickly accepted by the community as can be demonstrated by the fact that four out of ten in the list of Top500 supercomputers from June 2017 use coprocessors GPGPU (Figure 1).

Globally we can say that a graphical processing unit GPU is and accelerator, sometimes called a coprocessor, originally designed to perform specialized graphical operations faster than a standard PC. It contains one or more microchips containing a finite amount of vector or raster algorithms.

Graphical processing units were developed during the 1970s and 1980s, but started to be taken seriously in the 1990s. This was the time when these accelerators first appeared in gaming consoles, where they allowed for more realistic games using the 3D GPUs. Modern graphic cards allow stream processing (chaining or pipeline). A number of degrees of visualization give the possibility to create data parallelism. This means that every geometrical node

TOP 10 OF MOST POWERFUL SUPERCOMPUTERS

Number	Institution	System	Number of Cores	R_{MAX} PETAFLOP/S	R_{PEAK} PETAFLOP/S	El.Input MW
1	National Super Computing Centre in Wuxi ČÍNA	Sunway TaihuLight Sunway MPP	10 649 600	93,0	125,4	15,4
2	National Super Computer Centre in Guangzhou ČÍNA	Tianhe-2 (Milky Way-2) Intel Xeon E5	3 120 000	**33,9**	54,9	17,8
3	Swiss National Supercomputing Centre (CSCS) SWITZERLAND	PizDaint Cray XC50 NVIDIA P100 Cray Inc.	361 760	**19,6**	25,9	2.3
4	DOE/SC/Oak Ridge National Laboratory USA	Titan-Cray XK7 NVIDIA K20x Cray Inc.	560 840	**17, 6**	27,1	8,2
5	DOE/NNSA/LLNL USA	Sequoia BlueGene/Q IBM	1 572 864	**17,2**	20,1	7,9
6	DOE/SC/LBLN/NERSC USA	Cori-CrayXC40 Intel Xeon Cray Inc.	622,3	**14,0**	27,9	3,9
7	Joint Center for Advanced HPC JAPAN	Oakfoest PACS-PRIMARY CX 1640 M1 Fujitsu	556,1	**13,6**	24,9	2,7
8	RIKEN Advanced Institute for Computational Science JAPAN	K computer SPARC64 VIIIfx Fujitsu	705,0	**10,5**	11,3	12,7
9	Argone National Laboratory USA	Mira BlueGene/Q IBM	786,4 185 000	**8,5** **5,6**	10	3,9
10	DOE/NNSA/SNL USA	Trinity Cray XC40 Cray Inc.	301,1	**8,1**	11,1	4,2

FIGURE 1
List of most powerful supercomputers Top500 from June 2017.
(Ranking of Supercomputers according to the LINPACK benchmarks).

or image element can be calculated independently from others, but using the same algorithms, in other words using the SIMD approach. Parallel data approach SIMD later evolved into a more complicated sophisticated model SPMD.

Modern GPUs contain tens or hundreds of processing units, while each contains more arithmetic-logical units ALU, allowing effective usage of SIMD architecture. The fact that modern GPUs contain hundreds of ALU gives them computing power in the order of Tflop/s (NVIDIA GPU GeForce GTX580M – NVIDIA 2011). This level of performance naturally resulted in utilising the GPUs to solve computationally intense problems outside the traditional graphical applications.

Numerous scientific applications require the acceleration of linear algebraic operations. For example: to accelerate the multiplication of matrices, we can use the segmentation algorithm. Consider the multiplication of matrices

$$C = A \,.\, B,$$

where A is a matrix $(m \times k)$, B is a matrix $(k \times n)$. Matrix A can be split into column vectors in the form of $(r+1)$ matrices

$$A = \begin{pmatrix} A_o \\ A_1 \\ . \\ . \\ A_r \end{pmatrix}$$

Every imput A_i is a matrix $(p_i \times k)$

$$A = \sum_i p_i = m.$$

In practice, all p_i are the same.

Similarly, we can split matrix B into row vectors that are matrices $(s+1)$, so

$$B = (B_o, B_1, \ldots, B_s),$$

where every B_j is a matrix $(k \times q_j)$ and

$$\sum_j q_j = n.$$

Here, all q_j will be the same. Then vector multiplication of these two vectors can be written as $C = A \,.\, B$

$$C = \begin{pmatrix} A_o \\ A_1 \\ . \\ . \\ A_r \end{pmatrix} . (B_o, B_1, \ldots, B_s) = \begin{pmatrix} A_oB_o & A_oB_1 \ldots A_oB_s \\ A_1B_o & A_1B_1 \ldots A_1B_s \\ & . \\ & . \\ A_rB_o & A_rB_1 \ldots A_rB_s \end{pmatrix}$$

Every element of $C_{ij} = A_i \,.\, B_j$ is a matrix $(p_i \times q_j)$ and can be computed independently. Generalizing this approach by full implementation on a supercomputer yields SIMD matrix multiplication. The values pi a qj can be chosen such that every sub-multiplications fits the size of the supercomputer's memory. Every multiplication is performed on one GPU processor and the additions on the CPU processors. This means that every node of calculation uses the SPMD model. All matrix operations can be performed on an SPMD model supercomputer analogically.

By means of the SIVVP project (Slovak Infrastructure for High Performance Computing) have access to a high performance supercomputer IBM 775 AUREL with peak performance of 130 Tflop/s, currently operated by the SAS Computing Center. Aurel supercomputer has already produced numerous scientific results in the fields of quantum chemistry, seismological wave fields, fire and smoke spreading simulations, electron structure theory and a number of others.

Using this system requires experts capable of programming and securing the realization of complex numerical computations. They can be trained at the workplaces of Slovak universities and Slovak Academy of Sciences, if they can gain access to this supercomputer environment freely. Experience suggests that only a sufficiently large number of programmers and application specialists can ensure effective utilization of supercomputer facilities. A specialized institution should be established in Slovakia for this reason, preferably a national supercomputing center (NSC). Such an institution would have sufficient technical and application equipment along with professional personnel to ensure the realisation of supercomputer numerical calculations on its own computing environment and the future development of HPC in Slovakia.

There are numerous similar supercomputing centers throughout the member states of the European Union. The best known one is EPCC (Edinburgh Parallel Computing Centre) in Edinburgh. Supercomputing centers provide services free of charge for universities and their research, development and training, which sustains the potential of their users and employees.

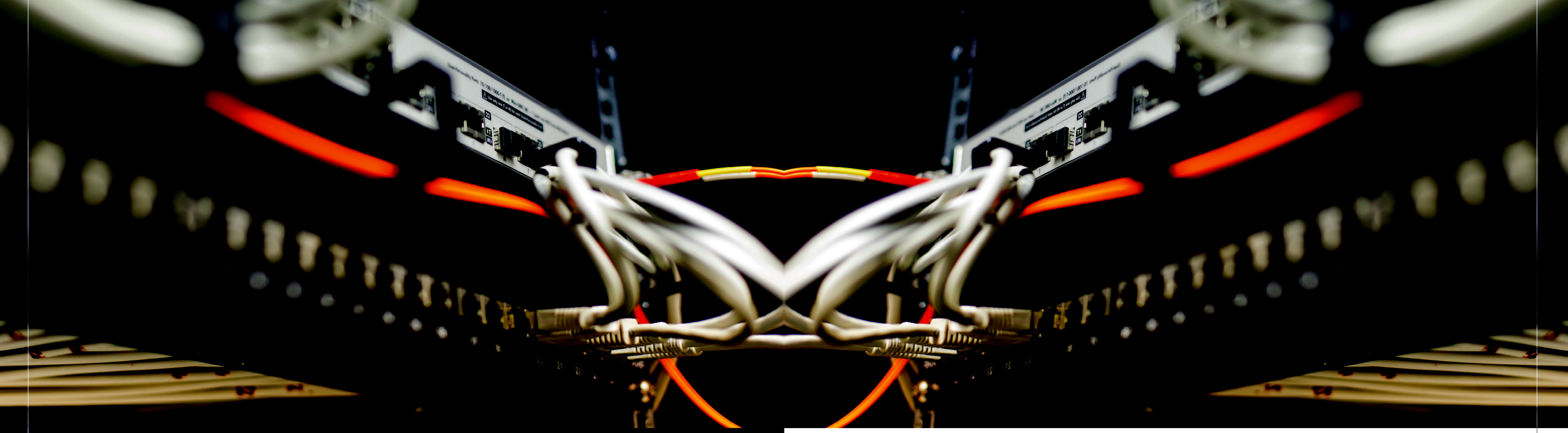
The creation of a Slovak supercomputing center would incorporate Slovak republic into the family supercomputer users, which already contains Slovenia, Czech republic and Poland emancipating Slovak scientists and their work in the international environment. Slovakia could catch up with its neighboring countries in the field of high performance computing and perform competitive research on par with the rest of Europe.

DR. H.C. PROF. ING. IVAN PLANDER, DRSC., IS THE FOUNDER AND THE FIRST RECTOR ALEXANDER DUBCEK UNIVERSITY OF TRENCIN (1997) AND CURRENTLY A PROFESSOR AND SCIENTIST IN THE AREA OF APPLIED INFORMATICS AND PARALLEL COMPUTER SYSTEMS.

02

APPLICATION OF THE HPC

hpc focus



High-performance computing at the Department of mathematics and descriptive geometry of the FCE SUT in Bratislava

PROF. RNDR. KAROL MIKULA, DRSC.
ING. ROBERT ČUNDERLÍK, PHD.
ING. MAREK MACÁK, PHD.
ING. ZUZANA MINARECHOVÁ, PHD.
ING. ROBERT ŠPIR, PHD.

Department of mathematics and descriptive geometry
Faculty of Civil Engineering,
Slovak University of Technology in Bratislava
11 Radlinského Street
810 05 Bratislava

In recent years, a HPC team under the supervision of prof. RNDr. Karola Mikulu, PhD. has been formed at our department. Thanks to his contacts and hard work, the team has become a member of the world elite in the field.

1 Processing of massive amounts of biomedicine data

In our applications using HPC cluster, we focus on processing long 3D biomedicine data series. This data is obtained from our collaborators on the project, biologists from CNRS - CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE in France. Data consists of three-dimensional microscopic scans of early stage of cell evolution from various organisms, such as zebrafish (Danio rerio), sea urchin or rabbit. These images reflect early evolution stage, about two hours after the fertilization of egg, up to fifteen hours, in about one minute interval. Aggregation of these images results in something like 4D video. We use image processing algorithms to extract valuable information from the data series. The data is filtered on input to eliminate noise using geodesic mean curvature flow, and consequently processed to extract coordinates of cell nuclei using the level-set center detection method. Despite the fact that both of these methods work on separate sets of 3D data, we use parallel processing to speed-up the process. The most computationally demanding application is the automatic tracking and reconstructing of the cell motion in the whole data set. To do so, we create a 4D sample out of series of 3D data, in which we construct the 4D segmentation using the cell nuclei coordinates, representing the motion on cells in space and time. Next, we get the so-called potential field by calculation of the distance function from initial cells inside the segmentations, which approximates cell motion in the data set, so that we finally obtain trajectories of individual cells. The algorithms are parallelized using several techniques, such as Open-MP and MPI, allowing for optimal distribution of computations across nodes of the computer cluster, better RAM utilization and overall calculation speed-up from days to a few hours. Results of our join research with French co-workers were published in 2016 in prestigious Nature Communications journal (member of Nature journals group).

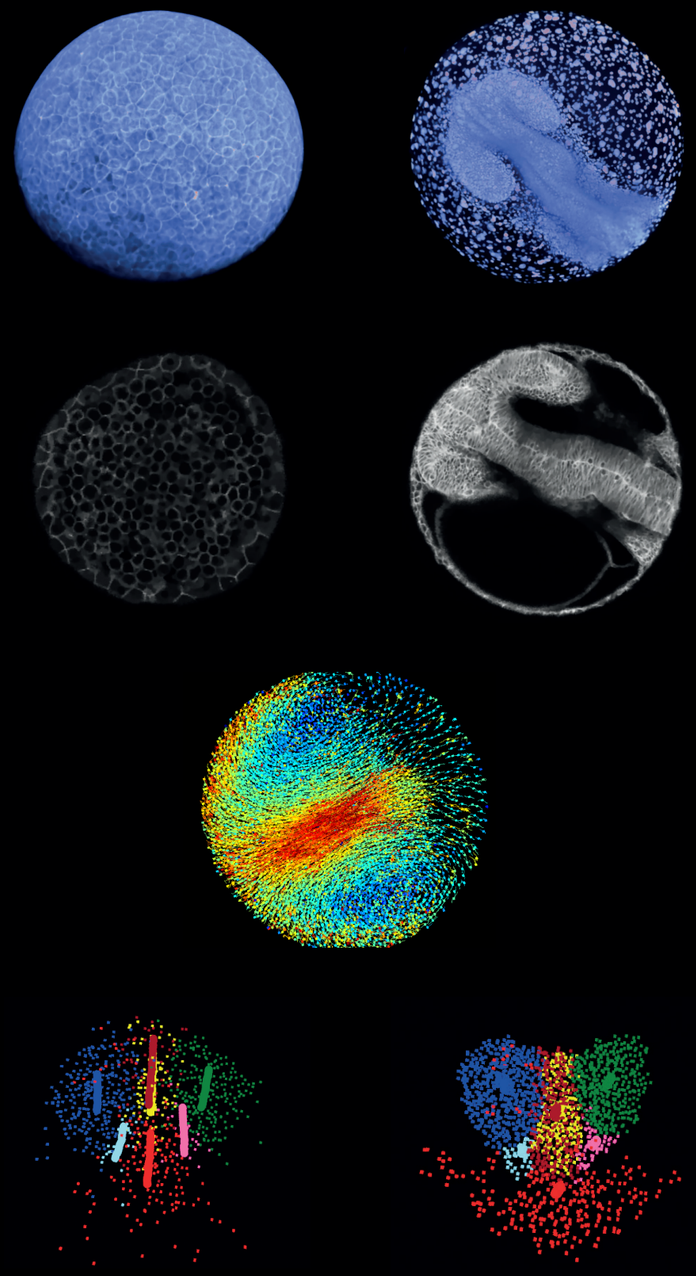


FIGURE 1:

First four figures show microscopic pictures of early cell evolution of zebrafish. These pictures show separately cell nuclei and cell membranes.

The fifth figure depicts stained cell nuclei, where the color reflects its velocity (red are fast, blue are slow cells).

Figures in the fourth row depict visualization of formation of individual, color-separated organs of a species, along with trajectories of the cell populations.

Ing. Marek Macák, PhD. is a university lecturer at the Faculty of Civil Engineering of the Slovak Technical University in Bratislava.

2 Solution of a geodetic boundary problem

One of the major tasks of geodesy is to determine the size of Earth, its shape and gravity field. The Earth gravity models can be calculated in various ways. At our department, we apply methods based on numerical solution of partial differential equations using boundary and finite elements and finite volumes.

Since the boundary of our computational domain is the Earth surface, that is known to be rough, we use triangles to approximate it. In order to achieve the most detailed results, we have to create a detailed grid with resolution of one minute (about 110 km on equatorial). Such a discretization corresponds to 21600x10800 unknowns on top of Earth surface. Our calculations use input data generated by satellites and surface measurements.

One of the applications of our research is determination of the Wo constant – reference value of gravity potential, which is fundamental in definition of World Height System. This problem is related to the phenomenon of different sea water temperature, salinity and presence of other substances. In consequence, the individual local height systems derived from the sea level are shifted against each other. For example, in Europe there are twelve valid height systems at the moment and the difference among them is more than several decimeters. Similar discrepancies are found in other continents and islands as well. For instance, an island country of New Zealand has thirteen local height systems differing by several centimeters with respect to each other. In order to unify the height systems and define the World Height System, it is necessary to choose the fundamental reference surface. The most suitable surface is the one that corresponds to the steady sea level and can be defined using the aforementioned constant gravity potential Wo and is call the geoid.

In 2011, the International Association of Geodesy (IAG) has decided to create a team of scientists, that are assigned to participate on calculation of new value of Wo constant under the supervision of Dr. Laura Sánchez from Technical University in Münich.

One of the four groups in the team in our laboratory.

The new value of the **W0** constant is **62 636 853,4 m^2/s^2**, and was officially accepted on the 26th general assembly of World union of geodesy and geophysics in Prague, in the summer of 2015.

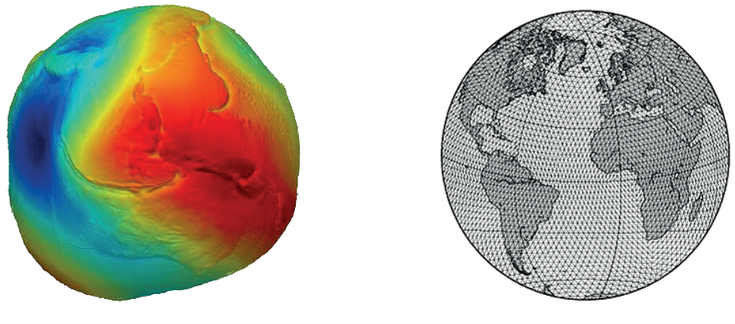


FIGURE 2:

Left: Physical shape of Earth (geoid).
Right: Triangulation – illustration of the computational grid.

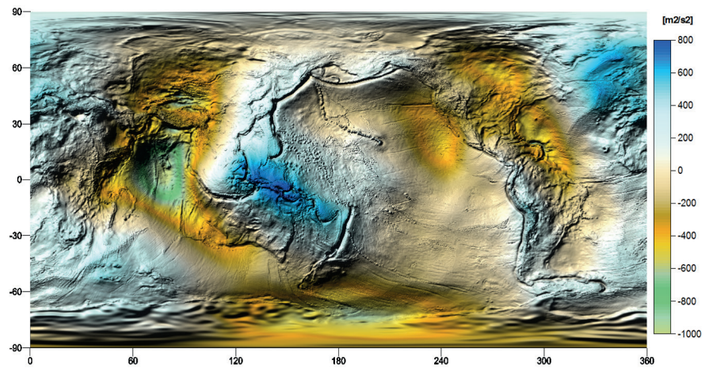


FIGURE 3:

The disturbing potential is the direct output of our computations. It is defined as the difference between the real earth and normal gravity potential. If compared with the real Earth potential, it increases with the deviations from normal body, ideal sphere. The disturbing potential is a physical quantity that enables us to determine geometric characteristics of anomalous, thus real gravity field, i.e. height of geoid or height anomaly.

3 Computational cluster

In order to facilitate very demanding calculations, a parallel computer cluster was assembled in 2015 on the premises of our department. The cluster runs using GNU/Linux operating system, has 8 GB RAM per CPU core, an optimal choice for our calculations, and has access to 40 TB of shared storage space. Fast data connections is achieved using 1 Gbit/s ethernet connection. Currently, our cluster has 160 CPU cores and was designed to be expandable (if funding is found).

> When Aurel supercomputer was made available, we gained a great new tool for our calculations that would hardly be feasible on our in-house cluster.

When Aurel supercomputer was made available, we gain a great new tool for our calculations that would hardly be feasible on our in-house cluster. Working on Aurel is beneficial for us, regardless of some time spent waiting in queue with jobs of other users. Aurel has become an effective tool for us thanks to its big size, pushing our calculations a step further.

Magdaléna Májeková
Yoel Rodriguez

Using HPC to model Membrane Proteins – Sarco/Endoplasmic Reticulum Ca^{2+}-ATPase (Case Study)

MAGDALÉNA MÁJEKOVÁ
Institute of Experimental Pharmacology & Toxicology SAS, Dúbravská cesta 9, 841 04 Bratislava, Slovak Republic

YOEL RODRÍGUEZ
Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, USA; Department of Natural Sciences, Eugenio María de Hostos Community College, 500 Grand Concourse, Bronx, New York 10451, USA

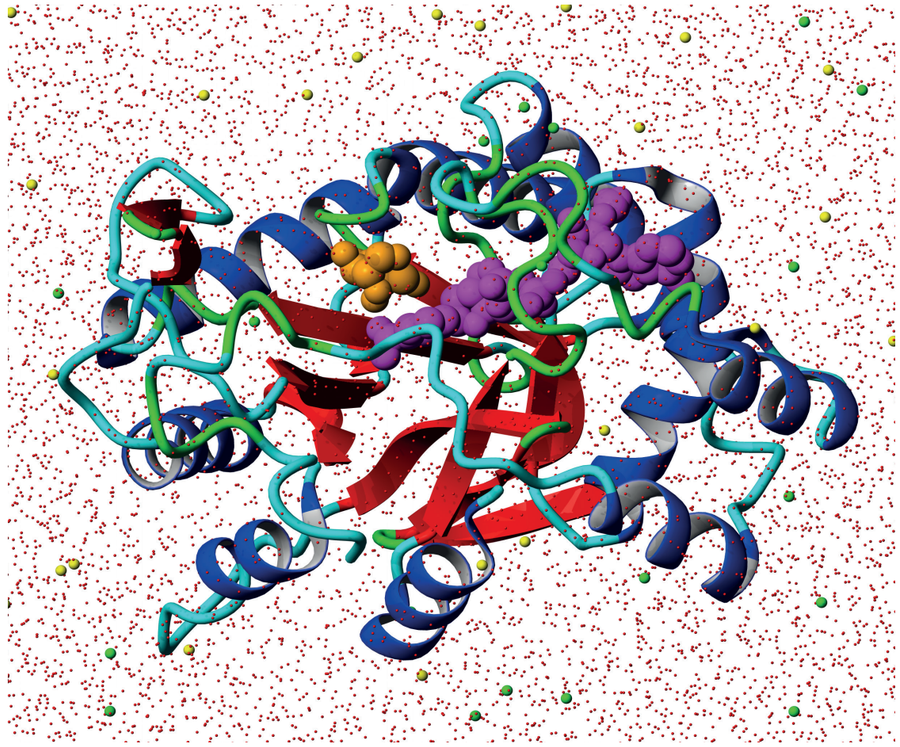


Figure 1: Enzyme aldose reductase with cofactor NADPH (Nicotinamide Adenine Dinucleotide Phosphate, magenta) and substrate D-glucose (orange) in water molecules, Na^+ and Cl^- ions environment.

Research of physiological and pathological processes in organism often requires simulation of proteins - the important components of living matter. The more complex the biological problem of study is, the more accurate the simulation should be set up. The length of the proteins in eukaryotes and, by implication, also in humans, is several hundreds of residues on average; it could even reach more than 20,000 amino acids (Brocchieri et al., 2005). Taking into account the proteins environment, they could be divided into two groups:

- cytoplasmic proteins, surrounded by a water solution of ions and other small molecules
- membrane proteins, anchored in biological membranes.

Molecular dynamics (MD) simulations of membrane proteins represent sophisticated tasks. The first reason is due to the system size. The systems comprising proteins embedded into membranes are obviously larger than those representing cytoplasmic proteins, as the ones embedded into membranes should generally be separated into three different phases (e.g., the cytoplasm, phospholipid bilayer and the lumen). While the computational modeling of a cytoplasmic protein needs only to be surrounded by waters, ions and eventually small molecules (see Figure 1), the computational modeling of membrane proteins requires to incorporate the phospholipidic bilayer, which creates the basis of a biological membrane (see Figure 2). In addition, if keeping in mind a physically realistic picture of such system, a membrane protein system simulation becomes more complex than the simulation of a cytoplasmic protein in just water environment. The basic building blocks of membranes are much larger and dynamically more complex than those systems that are just built with water molecules. Moreover, membranes can be comprised of several kinds of lipids. Thus, this can result in additional difficulties for conducting computational simulations of membrane proteins - the level of complexity of our modeling system (see below membrane-SERCA-ligand models).

The second reason is the real-time simulation, which should be used to perform the computational modeling study. An average relaxation time of water molecule necessary for its interaction with solute molecules with hydrophobic surfaces (which occur partially in proteins) is in the range of picoseconds (Comez et al. 2014; Muntean et al. 2012). Values of relaxation time linking to proteins structural changes go from nanoseconds to microseconds depending on the size and type of conformational change. The longest time interval (i.e., microseconds) is necessary for protein backbone conformational changes, such as protein folding (Shaw et al. 2010; Khodadadi et al. 2015). In the case of protein complexes (e.g., ryanodine or glutamate receptor), there could be sequences of conformational changes occurring chronologically. This may increase the modeling simulation time.

Researchers from the Department of Biochemical Pharmacology of IEPT SAS Ľ. Horáková, J. Lomenová a P. Rezbáriková

RNDr. Magdaléna Májeková, PhD. is a leading scientist in the Institute of Experimental Pharmacology and Toxicology SAS in Bratislava.

have been performing experimental studies of the structure and function of SERCA (i.e., Sarco/Endoplasmic Reticulum Ca^{2+}-ATPase) – calcium pump of sarcoplasmic reticulum (SR) for a long period of time. Several years ago this experimental approach was complemented with a theoretical study of isolated protein (M. Májeková). Last year, this study has been brought up to the level of more realistic models encompassing proteins-membrane immersed in a water environment and ions (Y. Rodríguez).

It has been shown that the physiological role of SERCA is linked to the pumping calcium ions from the cytoplasm (inner cell medium) to the lumen (inner SR medium), which maintains calcium homeostasis in the cell. Concentration of calcium plays an important role in many biological processes, such as muscle contraction, gene expression, insulin secretion, etc. SERCA dysfunction could promote the onset and progress of serious chronic diseases such as cardiovascular, neurodegenerative and muscle diseases, inflammation, cancer and diabetes.

Novel compounds yhat are able to accelerate, slow down or stop the activity of calcium pump have been discovered experimentally. Understanding the molecular mechanisms of these processes may help us in designing small molecules – drug candidates, able to modulate the activity of SERCA and consequently to restrict the progression of chronic diseases. Thus, aimed at explaining key mechanisms of inhibition and activation effects of these discovered ligands towards SERCA, Prof. Rodríguez, who visited our department last year thanks to a Fulbright fellowship, has developed different membrane-SERCA-ligand models. These models are currently being studied by means of computational biophysical approaches using molecular dynamics simulations, conformational analysis and free energy calculations. Altogether, the knowledge gained from this study will help us to advance the development of new drugs potentially useful in human diseases impacted by SERCA activity.

Khodadadia S, Sokolov P. 2015. *Protein dynamics: from rattling in a cage to structural relaxation.* Soft Matter 11: 4984-4998.

Comez L, Paolantoni M et al., 2014. *Hydrophobic Hydration in Water–tert-Butyl Alcohol Solutions by Extended Depolarized Light Scattering.* Phys. Chem. B, 119 (29): 9236–9243.

Shaw DE, et al. 2010. *Atomic-Level Characterization of the Structural Dynamics of Proteins.* Science 330: 341-346.

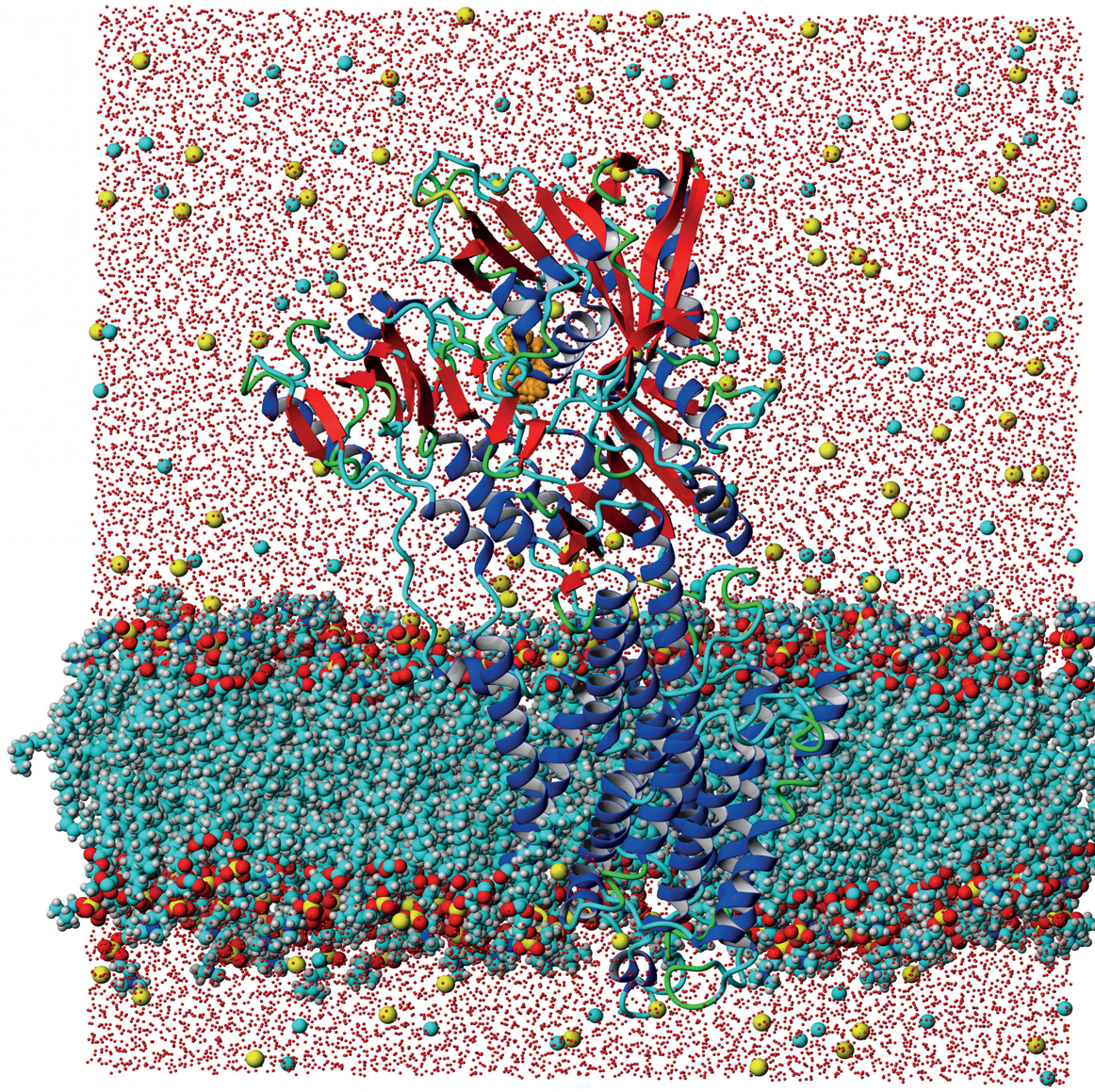
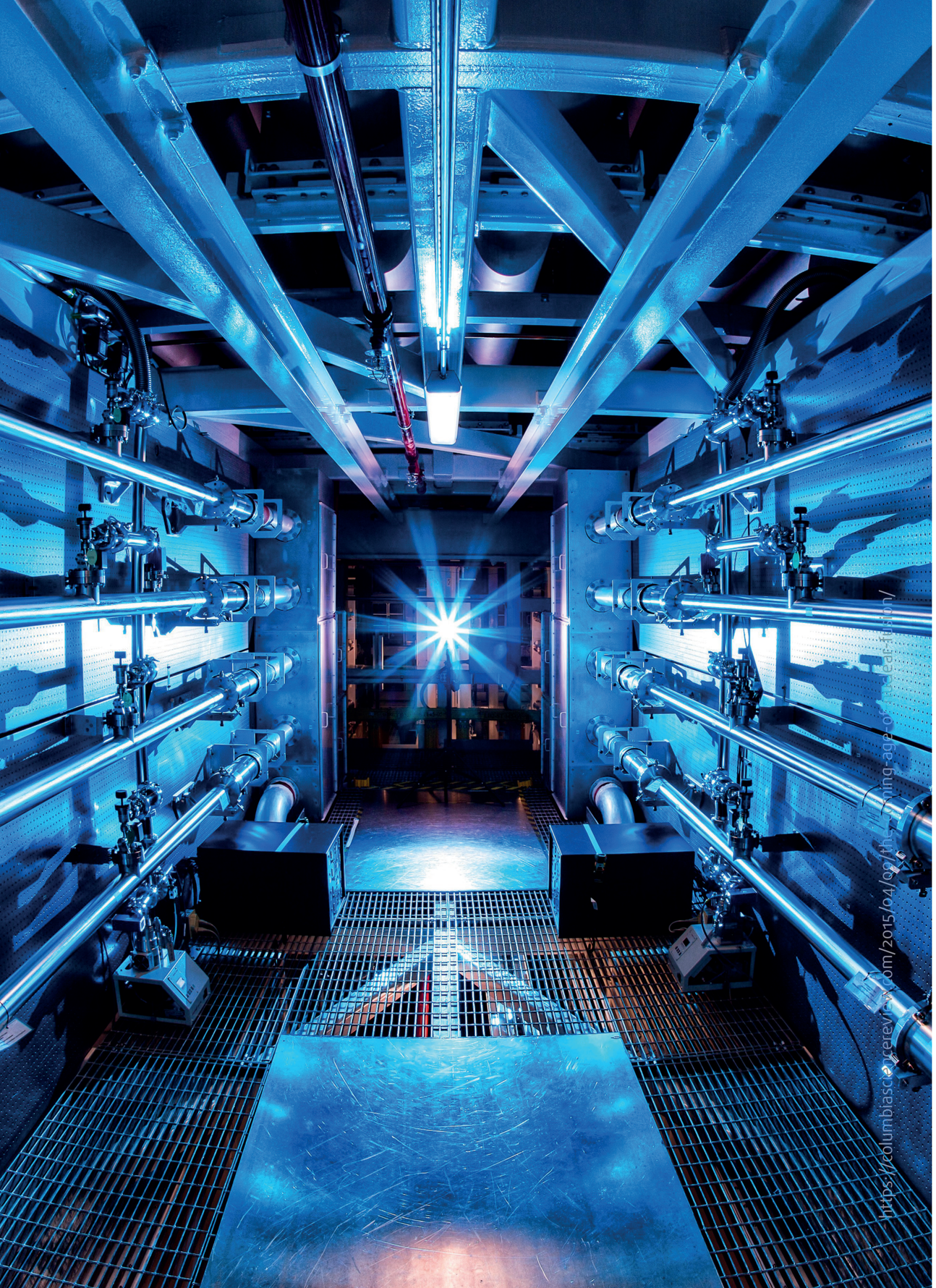


Figure 2:

A cross-section of the SERCA1 model embedded in the biological membrane and surrounded water molecules, Na^+ and Cl^- ions environment. In the cytosolic part of the protein, there are weak intermolecular interactions bound routine (orange). Status after 100 ns simulation.



https://columbiasc.wcrevi.com/2015/04/09/the-coming-age-of-nuclear-fusion/

Michal Novotný

Molybdenum based materials for application in nuclear fusion.

The ever present expansion of industry and development of new technologies has been gaining more and more momentum in the recent years. The time it takes from discovery of a new phenomenon to its practical utilization has been severely reduced from decades to a few years. But for one such discovery with a significant potential for humankind this is not the case. The phenomenon in question is the nuclear fusion. Since its theoretical formulation in 1930s until today it still requires a lot of research in several fields for it to reach its industrial application in the forthcoming years. The research fields range from nuclear physics and study of nuclear reactions, plasma physics, reactor design and engineering to development of material able to withstand the harsh conditions in the reactor. One of the limiting factors in the research are the aforementioned materials, since most of those employed in similar industries either do not meet the required attributes, or those that do are too expensive to manufacture. Using conventional methods of new materials development takes often long time and is financially challenging, since one must first synthesize the material and subsequently test it, while the desired properties are not guarantied. A viable way of eliminating both of these complications is to employ computational material design and the tools and methods of theoretical and computational chemistry.

What was unthinkable twenty years ago is becoming almost ubiquitous now. Thanks to advances of HPC and computational chemistry we can simulate realistic systems (dozens to hundreds of atoms) and thereby in a short amount of time and with reduced expenses provide a reliable prediction of material with the desired properties. Computational chemistry is inevitably dependent on the HPC infrastructure it has available, since its potential to use it is limitless.

Nuclear fusion takes place at temperatures of $15x10^7$ K. To reach this temperature a combination of three heating

Mgr. Michal Novotný is student of postgraduate studies at the Department of Physical and Theoretical Chemistry at the University of Natural Sciences of the UK in Bratislava.

mechanisms is involved. The first one heats up the plasma using electromagnetic radiation with frequencies of 40-55 MHz (Ion Cyclotron Resonance Heating, ICRH). The second one uses frequencies around 170 GHZ (Electron C clotron Resonance Heating ECRH). The third technique uses accelerated ions with kinetic energies of 1 MeV which are directly injected into the plasma (Neutral Beam Injector, NBI). Together these tree systems provide the heating power of 50 MW which is required for the plasma ignition. The focus of our group lies in the materials used to produce negatively charged ions for the NBI system. The materials typically incorporate molybdenum, tungsten or europium with a thin layer of cesium on their surface. In the current phase of our research we are interested in the changes in the work function caused by cesium on the surface or incorporated into the surface layers. The work function is the energy needed to remove an electron from the solid into vacuum. The value of the work function defines the effectiveness with which the conversion of the neutral hydrogen atoms to negatively charged ones will take place. This conversion is the main limiting factor of the NBI systems. Through experiments conducted on tungsten covered with cesium it is known that there is a specific coverage at which the minimum of the work function occurs. This minimum is often lower than the work function of pure cesium (work function tungsten 4.55 eV, cesium 2.14 eV and the minimum 1.82 eV). However the experiment can not explain why this minimum occurs. To understand this effect and to be able to exploit it in other materials our group is trying to reproduce this experiment in our simulations.

For our models to resemble the real system as close as possible, it is necessary to account for a large number of atoms in the simulations. Unfortunately here we are limited by the hardware that we have available. In solid state electrons form so called electronic bands which are similar to the orbitals in atoms and molecules. In contrast to simulations of molecules, where the Schrödinger equation is solved for the individual orbitals and electrons, in solid state simulations it is solved for the electronic bands which are subsequently filled (populated) with electrons based on their thermal energy distribution. Another characteristic of solid state simulations is the numerical sampling of space with the use of k-points. The nature of solid state simulations allows us to take advantage of two methods of parallelization one for the electronic bands and the other for the k-points.

Our calculations were done in three phases. The first one was the structural relaxationof the crystal cells and the surfaces of tungsten molybdenum. These systems were usuallycomprised of 55 atoms. 15 k-points and 464 bands. The simulations took place on the Aurel supercomputer on 8 nodes (256 cores) and took about 79 hours each (wall clock time). In the second phase we used the relaxed structures added varying coverages of cesium and let the structures relax again. The simulation were more demanding since the number of atoms increased to about 200 and the number of bands to 1700. The number of k-pints was reduced to 9. since the surface was pre-relaxed the simulations took about 50 hours on 8 nodes. The final phase were the calculations of the work functions of the prepared structures which required only a relaxation of the electronic structure. These calculations were done on 4 nodes which took about 24 hours each. Overall we have tested 20 different surfaces in this manner.

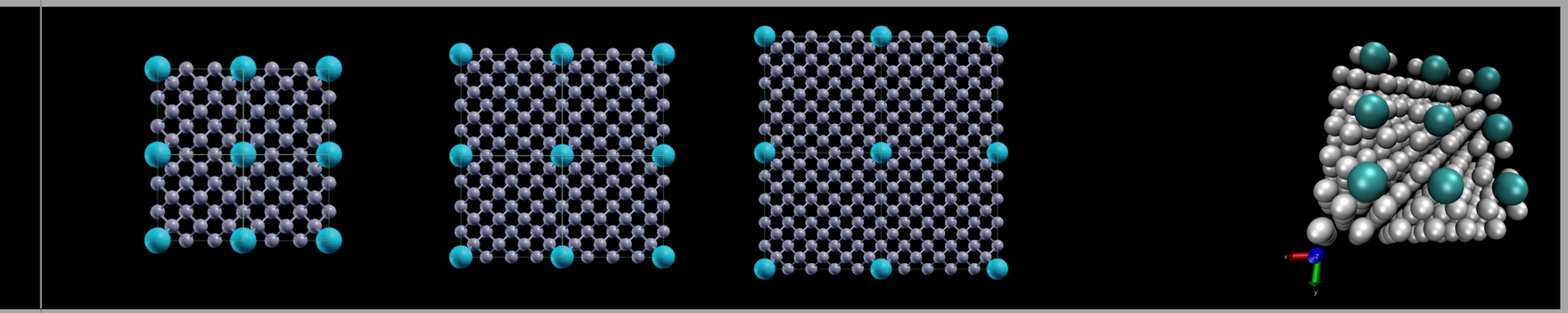


Figure 1:
Various coverages of cesium on a molybdenum surface.

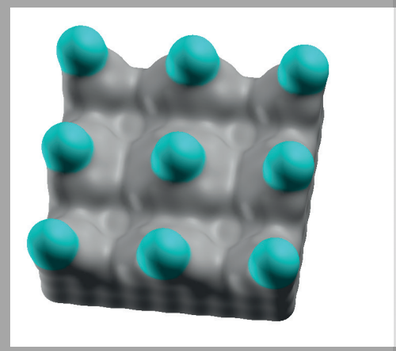


Figure 2:
Molybdenum surface covered with cesium.

It is evident that even a relatively smaller project such as ours requires significant computational power which is for many sites impossible to either acquire or maintain. The Aurel supercomputer, the computational cluster in Žilina and other resources acquired from the SIVVP projects are essential tools for our group and enable us to continually produce high quality scientific results.

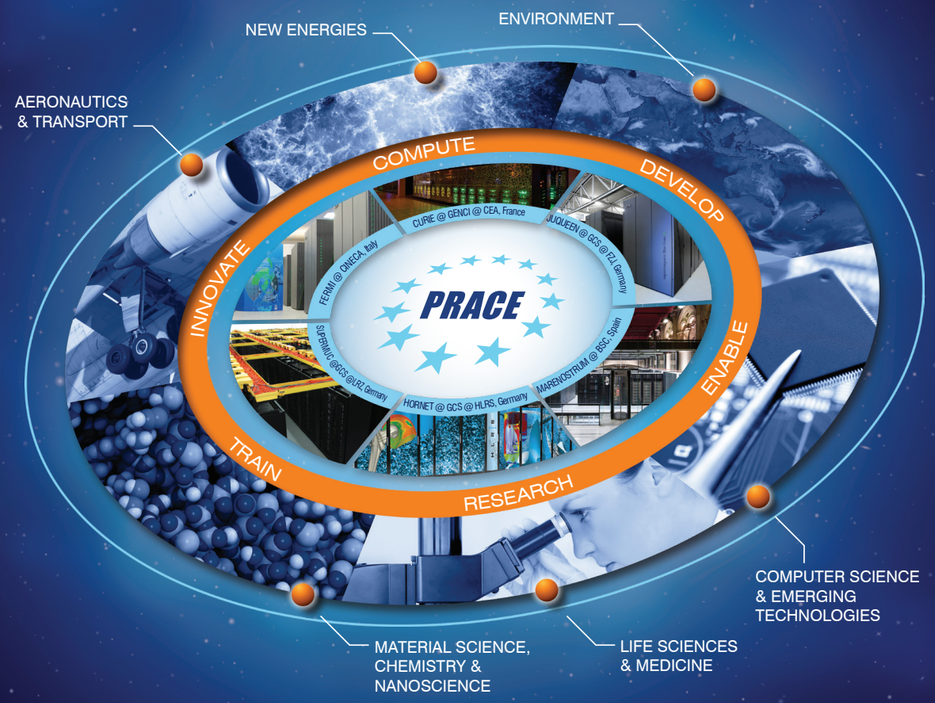
03

PRACE

hpc focus

PRACE SUMMER OF HPC at the COMPUTING CENTRE of SAS

Michal Pitoňák



SAS Computing Centre is a partner of the PRACE (Partnership for Advanced Computing in Europe) H2020 program, which from the beginning of its existence organizes an international, two-months internships called the Summer of HPC (High-performance Computing), *http://summerofhpc.prace-ri.eu*. In each year of the programme (about) twenty selected students are distributed among the top-rank supercomputing institutions in Europe – partner organizations of PRACE, to participate on HPC related research.

Students interested in participation on the programme provide their curriculum vitae, demonstrate their motivation towards the participation, appropriate programming skills as well as select three projects of their preference from the list of available projects for the annual, they would like to work on. Selected participants obtain a stipend covering all their expenses during the internship. The Summer of HPC programme begins with a training week organized on the premises of one of the PRACE partners. During this week, students are trained in parallel programming, work in HPC environment, (big) data visualization and more. After the training week, students travel to destination institutions, where they work on the assigned projects. At the end of the summer, students elaborate a final report, briefly summarizing the goals and achievements of their project as well as a short, five-minute video presenting their efforts to wide audience (*https://youtube.com*, search for e.g. "PRACE Summer of HPC 2017 presentations"). Participants of the Summer of HPC are motivated to continuously document and share progress of their projects of the official blog of the Summer of HPC as well as on social networks, in order to catch attention of their peers towards the HPC. Their effort is rewarded by financially honored "Best HPC ambassador" price, that is awarded along with the "Best visualization award" each year to the winner in the category.

During the history of the Summer of HPC programme only a single student from Slovakia participated, Mgr. Ján Hreha, a PhD student of physics from the Faculty of mathematics, physics and informatics of the Comenius University in Bratislava. Ján spent the summer of 2015 at the NIIF (National Information Infrastructure Development Programme) institute in Budapest working on project titled "Bringing Hybrid Architecture's power for Atomistic Simulations".

The Computing Centre of SAS actively participates on the Summer of HPC programme from the beginning of it's partnership with PRACE, and hosts two students each year. In 2016 it was Oisín Benson from Ireland and Katerina Gelata from Greece. Oisín worked under supervision of doc. Mgr. Michal Pitoňák, PhD. on a project titled "Quantum Chemistry in Spark", while Katerina worked under supervision of prof. RNDr. Jozef Noga, DrSc. on a project titled "Calculating Nanotubes". This year, 2017, it was Adrián Rodríguez-Bazaga from Spain, working under supervision of doc. Pitoňák on a project titled "Apache Spark: Are Big Data tools applicable in HPC?", and Andreas Neophytou from Greece, working under supervision of prof. Noga, working on a project titled "Calculation of nanotubes by utilizing the helical symmetry properties". Interviews with the aforementioned participants are published separately in this issue of HPC Focus.

In the beginning of 2018 a new call for participation on the Summer of HPC project on the PRACE web pages (*http://www.prace-ri.eu*) will be announced. We would like to encourage all potential candidates, be it university or PhD students, to participate and take advantage of this unique opportunity to acquire new informations and experience related to HPC on prestigious European supercomputing institutions.

SUMMER of HPC 2016

2-month intership at the CC of SAS

Katerina Galata
Oisín Benson

Tell us a little bit about yourselves. Where are you from? What's your academic background?

KG: My name is Katerina Galata and I am from Athens, Greece. I am a chemical engineer in the national technical university of Athens and right now I am the final year of my studies, pursuing a master. I am currently also working on my diploma under the supervision of Prof. Doros Theodorou.

OB: I'm from Dublin, Ireland. I'm a chess player and ex-captain of the Trinity Armstrong (Leinster Division 1) team. I also enjoy grand strategy games and sci-fi particularly Greg Egans Diaspora. I just finished a BA in theoretical physics at Trinity College Dublin.

How did you find out about PRACE Summer of HPC? Why did you decided to participate?

KG: My professor suggested that I should apply to this program, as it would be a great opportunity for me. I would be able to learn a lot of new things and I would gain experience in the field I am interested to work afterwards. So I went through the subjects of the projects and I realised that these are indeed very interesting. So I decided to apply for the projects that I thought were the most interesting ones and in the meantime were close to my field (of course I also had to check if I had the prerequisites needed).

OB: I heard about SoHPC from a few places. I knew several students from other years in my course who had completed it, including my chess teammate Anthony who had throughly enjoyed his time in the Czech Republic. It had been also been recommended to me by researchers at ICHEC & the professors who taught programming modules in my course. Out of my degree the modules I enjoyed the most were the coding ones. Implementing algorithms such as Ising Model and seeing the results of the simulation was very satisfying. PRACE sohpc seemed like a fantastic opportunity to get to grips with problems and tools on the cutting edge of research alongside the chance to live in another country for while.

Why did you decided to come to our organization, what were your other picks?

KG: I decided to come to SAS, because I really liked the subject I applied for. Its title was " Calculation of nanotubes by utilizing the helical symmetry properties". I had a strong theoretical background in that field and I had the prerequisites needed for it. I had also applied for "write the spanish" but the program in SAS was my first choice and I was hoping I would be selected for this.

OB: My top two picks were both big data style projects. I had spent the previous summers at the tech company Workday working on data science projects using Hive & R as well as gaining some experience with scala. During my time at the company there was a lot of excitement about the new tools coming online such as Spark. Michals project was more or less a perfect fit for me as it would allow me to build on my previous experience with Scala and learn about the new data processing engine Spark. On top of this I had already covered a significant amount of quantum chemistry through the condensed matter theory modules of my degree meaning that I could spend more of my time on learning Spark. The other choice was a link algorithm project using Apache Fink in Slovenia. This sparked my interest because it involved algorithms finding links between different research papers, allowing computers to process and review more research papers than anyone could hope to in a lifetime.

What were your assignments during your stay with us? Did you receive any help/training from our colleagues?

Katerina Galata is a chemical engineer in the national technical university of Athens and right now she is the final year of her studies, pursuing a master.

Oisín Benson finished a BA in theoretical physics at Trinity College Dublin.

KG: In this project I had to parallelize some routines of an existing code. This was a bit tough at first since I was not familiar with the code, and as everybody knows it is always hard to understand what another programmer is doing and why he is doing it. Now consider that this code was made up by several programmers through the years. Fortunately my duty was depending only on some parts of the code. Prof. Jozef Noga provided me with the necessary directions and gave me some suggestions in order to help me do what was needed in the best and easiest way possible. And of course, Lukas Demovic helped me and guided me every time I faced difficulties while working on the project. Thanks to him I developed my skills and knowledge as far as it concerns MPI, as I hadn't worked in such a huge project before. He also taught me how to work on a supercomputer such as Zilina. So I am most grateful to both of them.

OB: My assignment was to adapt the libint C library for the Hartree-Fock Method in Quantum Chemistry to Spark and then to run performance tests & optimise it. Before this however we would implement the code in the cross-paradigm language Scala which heavily utilises many of the functional methods and transformations that Spark relies on. Master sys admin Lukáš helped set up the necessary tools for the project and saved me a lot of headaches. Once the tools were set up and I was done doing some initial reading and tests on Spark, Michal laid out a rough plan of what we needed to do and directed me to some useful resources from other implementations of the Hartree-Fock method. The downside of using the modern tools was that I could only really direct any questions about Scala or Spark to Michal. Still despite being unfamiliar with the tools, Lukáš was able to provide some help bug bashing, even if only as a sanity check.

What results did you achieve? Were you satisfied with them? What are future applications of your results?

KG: The aim of the project was to calculate the properties of nanotubes by using the helical symmetry. Implementation is based on a formulation in two-dimensional reciprocal space where one dimension is continuous whereas the second one is discrete. Independent particle quantum chemistry methods, such as Hartee-Fock and/or DFT or simple post Hartree-Fock MP2 are used to calculate the band structures. MPI parallelization is now implemented to enable highly accurate calculations for the band structures of nanotubes on distributed nodes, with distributed memory. New possibilities for tractable highly accurate calculations of energies and band structures for nanotubes with predictive power and with facilitated band topology whose interpretation is much more transparent than in the conventionally used one-dimensional approach are now available. The results we got can help turn this project into a standard tool for in silico design. The code needs further parallelization to be optimized of course.

OB: Unfortunately we did not achieve as much as we'd planned to but we did make significant progress into the problem. I initially rewrote the C implementation of libint in Scala with the help of it's numerical library breeze. Breeze was the obvious numerical tool to use since it's operations are the basis of Sparks machine learning library MlLib. Once we had this running for small basis sets we began adapting it for Spark. This meant the code needed to more heavily utilise the functional side of the Scala language which natively contains many of the data transformations that are part of Spark. After some headaches we managed to get a version working for small basis sets of H_2O in the final week. However we did not have time to overcome the remaining issue that was causing the code to give diverging energies for larger basis sets. Michal & his Phd student Miro will continue to work on it and are optimistic that they can get the code to scale correctly. If this is done it will be possible to run the algorithm on a Spark cluster with Gb basis set files. While I would have liked to have finished this off before I left I am happy with the progress I made and the new tools I learnt which I will definitely use in my masters and future career.

Did you like Bratislava? Did you like working at the Computing Center of the Slovak Academy of Sciences?

KG: Well.. Slovakia, from what I have seen, has a great natural beauty. And life in Bratislava is not tense at all, which really helped since for the first time, I had to work during the summer. Another advantage is that you can walk through the city and get anywhere you want within 40 minutes (maximum). Bratislava is also just a couple of hours away from other European capitals which is great as you can travel around during the weekend. As far as it concerns your question on whether I liked working at the Computing Center of the Slovak Academy of Sciences, my belief is that what defines your working experience at a place is your colleagues, not your office. So yes. I had a great time there. Everybody was so friendly and cheerful. They were always making sure that I had a great time there. (Lukas even offered me his bicycle to use it while I was staying in Bratislava, and I really appreciated this gesture.) They even showed us around the city. So I can't be anything less than completely satisfied and happy. I couldn't have asked for anything better, since they made sure we had the best experience while we were staying in Bratislava and working in SAS. And I was really relieved because being a foreigner in a new city can be tough, especially when you don't speak the native language.

OB: I enjoyed my time in Bratislava. The city was very vibrant and seemed to have quite a young population. I did not miss the grey Irish summer and it was nice to be able to go out without spending an arm and a leg. Bratislava provided another bonus since it neighbours other major European cities. I was effectively able to squeeze in a mini inter-railing trip, with day trips to Vienna & a weekend in Budapest. Everyone at the office made Katerina and I feel very welcome and the summer barbecue was a great way to finish our research placement. Having another strong chess player, Jozef, at the office was great to & I had quite a few blitz chess games with him over the summer.

Overall I've had a great summer here, I've learnt a lot, made new friends and have got to visit some great cities.

THANK YOU for your interview and we wish you a lot of success at work and in your personal life.

SUMMER of HPC 2017

Adrián Rodríguez-Bazaga
Andreas Neophytou

Firstly, could you please introduce yourself a little. What field did/do you study and where? What is your specialization?

ARB: My name is Adrián Rodríguez-Bazaga, and I'm a Spanish Computer Scientist specialized in Algorithms, Artificial Intelligence and the Theory of Computation, currently studying a Master's Degree in Innovation and Research, specializing in Machine Learning and Data Science.

I'm currently working as a Research Scientist in the Department of Life Science, at the Barcelona Supercomputing Center, involved in a research project in the field of Data Science and Natural Language Processing, trying to analyse the evolution of bioinformatics in the last century, using some novel approaches that require the use of Machine Learning techniques, such as the Latent Dirichlet Allocation, and clustering techniques such as Spectral Clustering.

AN: My name is Andreas Neophytou and I hail from Birmingham, England where I studied Natural Sciences. I majored in Chemistry and Biochemistry, but I also minored in Computer Science which gave me a window into HPC and Scientific Computing. Currently I'm a Masters student at the University of Birmingham working on the Exploration of Crystal Energy Landscapes. Essentially I apply optimisation methods to semi-empirical energy models for small organic molecules in order to find their lowest energy structures.

How did you find out about the Summer of HPC organized by PRACE? Why did you decide to participate? What made you choose us – the Computing Centre of the Slovak Academy of Sciences?

ARB: I found about Summer of HPC because a teacher of my former university told me about it, and I decided that it was a good idea to participate because it was an excellent opportunity to get a serious hands-on experience in High Performance Computing. I chose the Slovak Academy of Sciences because one of the projects required the usage and knowledge of some tools, which I was interested, such as Apache Spark.

AN: A doctoral student at my University sent me a link to the programme, my experience with HPC at the time was minimal and so I thought it to be a perfect opportunity to gain hands-on experience in HPC and parallel programming. I chose CCSAS mainly due to the project, it was the only project offered that involved applying HPC to a Chemistry related problem.

Can you describe the progress of your project?

ARB: My project progress was quite good, we achieve to complete some goals and discover new ways to research on, that lead us to cooperate even after the SoHPC period, with the ultimate objective of writing a research paper about our achievements.

AN: The mission statement of my project was "implement MPI parallelization to enable highly accurate calculations for the band structures of nanotubes on distributed nodes, with distributed memory". The final project did not deviate much from this, due to the size and complexity of the software I was working with the project ended up being more difficult than I first expected it to be.

What was your objective? Was the assignment clear/easy to understand? Did you reached the goal of the project?

ARB: The objective was to combine the benefits of native MPI-like approaches with Big Data tools advantages to bridge the gap between High Performance Computing and Big Data, in order to ascertain if it is useful to use typical Big Data tools to solve High Performance Computing problems by exploring and comparing two distributed computing frameworks implemented on commodity cluster architectures. We compared Apache Spark Big Data framework written in Scala with ‚traditional' approaches, such as using distributed memory model with MPI on a distributed file system, HDFS (Hadoop Distributed File System), and native C libraries that create interface to encapsulate this file system functionalities. To be more precise, we've chosen the K-means clustering algorithm

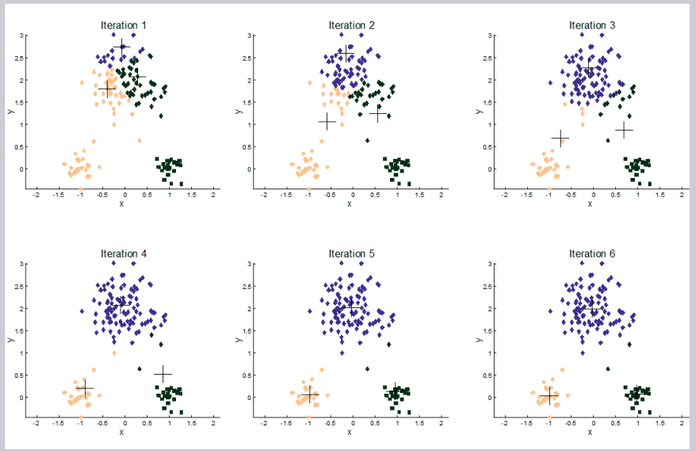
Adrián Rodríguez-Bazaga is currently studying a Master's Degree in Innovation and Research, specializing in Machine Learning and Data Science. He specialize in Algorithms, Artificial Intelligence and the Theory of Computation.

that will be ran on variable size datasets and will be compared in terms of computational time and failure resilience for both approaches.

AN: My instructor was Professor Jozef Noga, he was extremely helpful and I don't think I would've been able to finish the project without his input. However, everyone at CCSAS was also a great help, not just with the project but also in making my stay over the summer extremely enjoyable.

FIGURE:

Representation of individual iterative K-means of the algorithm to search for cluster centers using model data.



AN: The object was clear from the beginning, but that didn't make it any easier! I did manage to complete the project however (although I was skeptical at times).

Who was your instructor? Was he helpful?

ARB: My instructor was Dr. Michal Pitonak, who was very helpful in the development of the project, always providing new resources, trying to answer all my doubts and making sure that the project went correctly, he's not only a great researcher, but a great person, with whom I'm currently working in extending the results achieved during the SoHPC period.

Are you going to use the knowledge you obtained during the SoHPC in the future?

ARB: Of course, the knowledge acquired during the SoHPC is helping me in my daily work, because I need to deal with very complex problems which need to be deployed in HPC clusters, so the fact that I have already worked, thanks to the SoHPC programme, in a HPC environment, is a quite good advantage.

AN: Of course, I've already applied MPI to code used by the research group I'm currently embedded in. Needless to say I'm quite a popular figure around here now!

How would you rate your stay at the Computing Centre SAS during the two month internship.

ARB: In general, my stay was very good, of course there is no place like home but I cannot complain, the people was very kind and I feel very satisfied with the choice of my destination.

AN: I genuinely enjoyed my stay at CC SAS, if I had to quantify it my stay was an 8/10 (two points were lost due to the rainy weather!). Truthfully I'm not sure what could be done to improve the experience.

Was this your first time in Bratislava, in Slovakia?

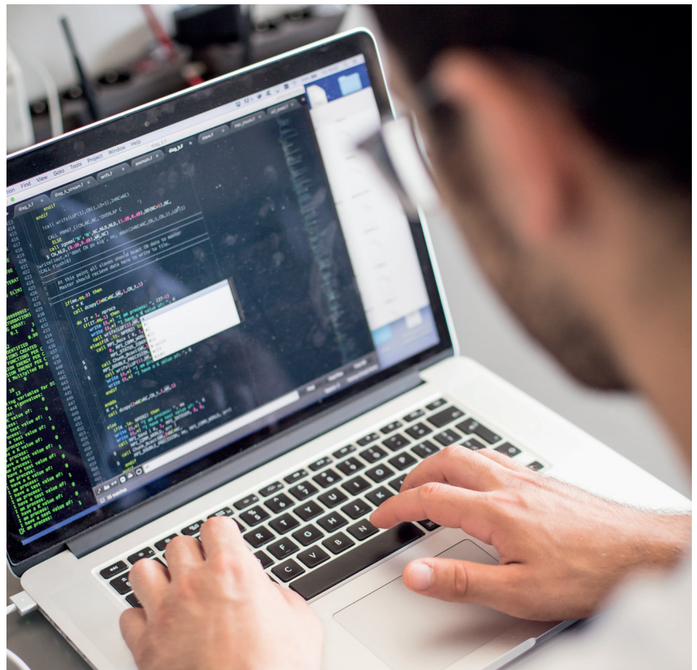
ARB: Yes, it was my first time in Bratislava and in Slovakia, even in Eastern Europe. The city was very nice and it has very interesting historical monuments and places.

AN: This was my first time in Slovakia, I enjoyed my time there. The people were very friendly, I had many interesting conversations with strangers on the bus to work.

Did you visit any interesting places? Would you ever come back here as a tourist?

ARB: We visited some interesting places, like the Bratislava Castle, the Presidential Palace and it's garden, I also remember that "Man at Work" monument in Old Town, it was very funny. There are some other interesting things to see in Bratislava so of course I would come back as a tourist and see them all.

AN: I did manage to do the sightseeing around Bratislava, as well as visit other nearby countries (Vienna and Budapest). However, the highlight was canoeing (and capsizing) down the Danube with everyone from CCSAS. I do plan on returning in the future, next time I plan to visit the High Tatras Mountains as well.



Andreas Neophytou studied Natural Sciences. He majored in Chemistry and Biochemistry, but he also minored in Computer Science. Currently he is a Masters student at the University of Birmingham working on the Exploration of Crystal Energy Landscapes.

hpc focus

DIRECTOR of CC SAS
Lukáš Demovič

RESPONSIBLE EDITOR
Gabriela Obadalová

GRAPHIC DESIGN AND DTP
Gabriela Obadalová

TRANSLATION
Michal Pitoňák
Miloslav Valčo

PHOTOGRAPHER
Pavol Novák

PRINT
FOART, s.r.o.
Prešovská 45
821 02 Bratislava
Slovak Republic
Tel.: +421 (0)2/207 821 18
E-mail: foart@foart.sk
www.foart.sk

PUBLISHER
Computing Centre of the
Slovak Academy Sciences
Dúbravská cesta 9
845 35 Bratislava
Slovak Republic
Tel.: +421 (0)2/3229 3111
Fax: +421 (0)2/3229 3103
E-mail: vssav@savba.sk
www.vs.sav.sk

The texts have not undergone proofreading.

Computing Centre of the
Slovak Academy of Science

Dúbravská cesta 9
845 35 Bratislava
Slovak Republic

www.vs.sav.sk