Magazine of the Computing Center of the Slovak Academy of Sciences 2023

hpc focus

EURO SLOVAKIA

ISBN 978-80-89871-20-9
ISSN 2729-9090

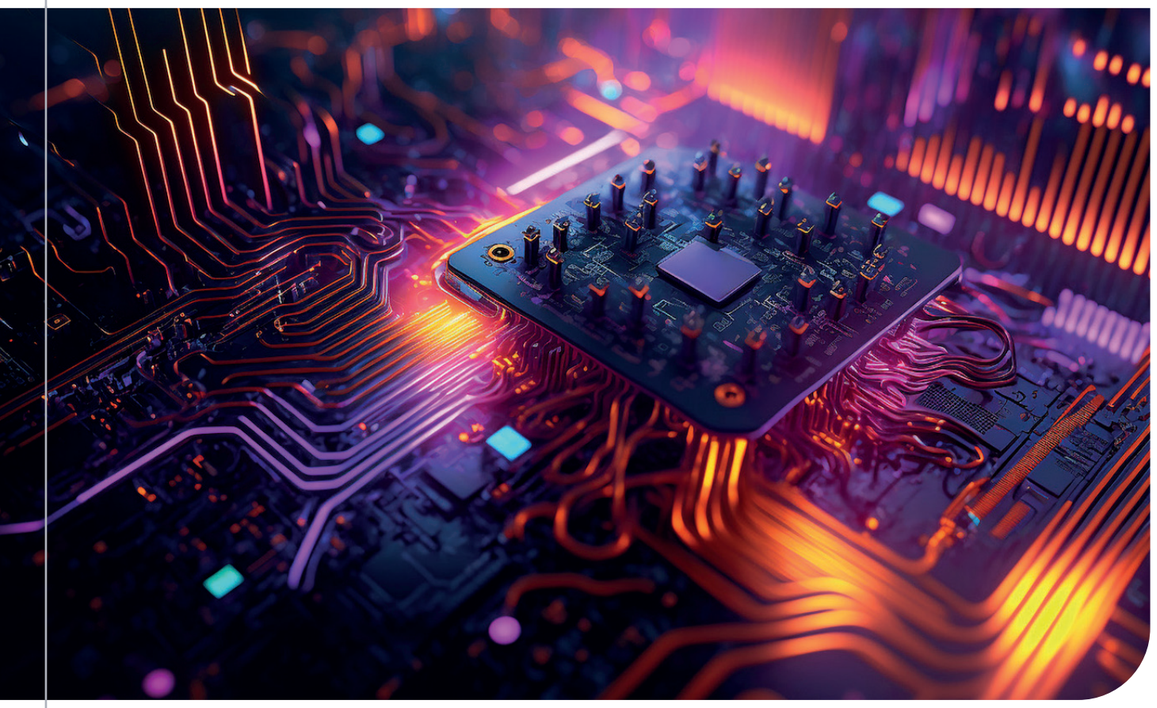
Editorial

Dear HPC users and supporters,

just as the EuroHPC JU (The European High Performance Computing Joint Undertaking) announced the second European exascale supercomputer to be built in France, following the first one in Jülich, Germany, here in Slovakia we are eagerly anticipating the launch of the Devana HPC system. Various unforeseen circumstances, including repeated public procurements, delays in component delivery, and an extended period of preparation and testing, have slightly postponed the availability of the new HPC system to Slovak users. Nevertheless, we are confident in the influx of quality project proposals and believe that Devana's impressive performance of 800 TFlop/s will be continuously and fully utilized.

In addition to Devana, we are also in the advanced stages of preparing for another Slovak supercomputer, tentatively named Perun. Funded by the recovery and resilience facility, Perun is expected to deliver performance of approximately 50 PFlop/s and should be operational by the end of 2025. We are also committed to developing the ecosystem and supporting research and development based on HPC technology.

However, these infrastructure investments, while essential, represent only one facet of the broader picture of HPC development. Another critical, yet often overlooked aspect, is securing funding for operation and services, which, in the context of HPC technologies, entail significant annual financial commitments. The amount of funds the hosting institution is prepared to invest over the sys-



tem's lifetime should be a crucial factor in the initial hardware investment. The current funding situation in Slovakia is not entirely clear, although it is THE determining factor in access policies and conditions. The launch of Devana and the preparation of Perun present an opportunity to establish a national HPC development strategy created and supported broadly by user representatives. Of course, the cooperation of relevant public administration bodies during its implementation is crutial. This initiative would position Slovakia as an equal partner among European nations and enable Slovak scientists and researchers to collaborate with their European peers in tackling complex issues across various fields.

We invite you to learn more about the current HPC access policy on our website. Additionally, we encourage current and prospective users, including those from small and medium enterprises, to leverage Devana for their research and development projects. We look forward to collaborating with you!

Lucia Demovičová

OBSAH

Európska únia
Európsky fond regionálneho rozvoja
Operačný program VÝSKUM a VÝVOJ
Agentúra Ministerstva školstva, vedy, výskumu a športu SR pre štrukturálne fondy EÚ
Podporujeme výskumné aktivity na Slovensku
Projekt je spolufinancovaný zo zdrojov EÚ

01

NSCC
NCC for HPC
EUROHPC

hpc focus

DEVANA

NEW SLOVAK SUPERCOMPUTER AVAILABLE FOR RESEARCH AND INNOVATION

01 Access & Services

A new era of HPC (High-Performance Computing) services has begun for the Slovak scientific and research community, thanks to the recently launched supercomputer Devana. Devana was procured within the framework of a project called National Competence Centre for High-Performance Computing (project number in the ITMS2014+ system: 311071AKF2), which is funded by the European Regional Development Fund, EU Structural Funds for Society Informatics, specifically from the Integrated Infrastructure operational program 2014-2020. However, its use is not limited to the academic community; rather, the system is accessible to private companies and public administration. In this article, we will provide practical information for current and future users.

First and foremost, it is necessary to emphasize that the mentioned new era is not only about the new supercomputer - in an effort to reach the European level, the team of the Computing Centre SAS and the National Supercomputing Centre have also prepared a new comprehensive concept of services in the HPC field, which will be gradually implemented with an emphasis on their availability and quality. A new user and registration portal, a clear website with practical guides, and user support through a helpdesk based on the „next business day“ are already available.

A new user and registration portal, a clear website with practical guides, and user support through a helpdesk based on the “next business day” are already available.

Who and how can use Devana is defined by the access policy, which depends on the operator‘s capabilities and may change. Currently, all projects that fall strictly under open research and development can be implemented free of charge, whether the researcher is a university employee or a worker of a private company. Part of the computing capacity is also set aside for commercial use and also for clients of the two Slovak ECDIs (European Centers of Digital Innovations) – Hopero and the Center for Innovative Health. Among other things, ECDI provides companies with Devana’s computing capacity as part of the state aid scheme, i.e. also free of charge.

Above all, small and medium-sized enterprises, but also larger industrial companies, have the opportunity to use

the services of the National Competence Center for HPC and implement proof-of-concept and pilot projects with experts in the HPC field, who will help with the development and deployment of various solutions in the HPC environment free of charge with software parallelization, diagnostics and optimization

R&D STATE AID	ECDI	▪ assistance with digitalization ▪ test-before-invest ▪ education/training ▪ assistance with obtaining grants ▪ networking
R&D OPEN SCIENCE FREE OF CHARGE	NCC for HPC	▪ free cooperation on R&D projects ▪ education, courses, workshops ▪ international competence network in HPC+ ▪ assistance with access to EuroHPC systems
ACCESS TO COMPUTING RESOURCES	HPC CENTRE CC SAS NSCC	▪ access to the national HPC infrastructure ▪ professional user support ▪ technical assistance with HPC environment and applications

SCHEMA 1

Current overview of HPC services and providers in Slovakia.

02 How to start

As a first step, it is necessary to create a user account on the HPC portal register.nscc.sk. After submitting the registration form, it is necessary to wait for verification and notification, after which it is possible to submit a request for access to the computing infrastructure within various types of projects, manage projects and SSH keys, and monitor the utilization of allocated allocation in individual solver projects. There are also links to the Helpdesk and a „knowledge-base" page with instructions and guides. Computational time can be obtained for a specific project within one of the following options: projects for testing and benchmarking, standard projects, and commercial

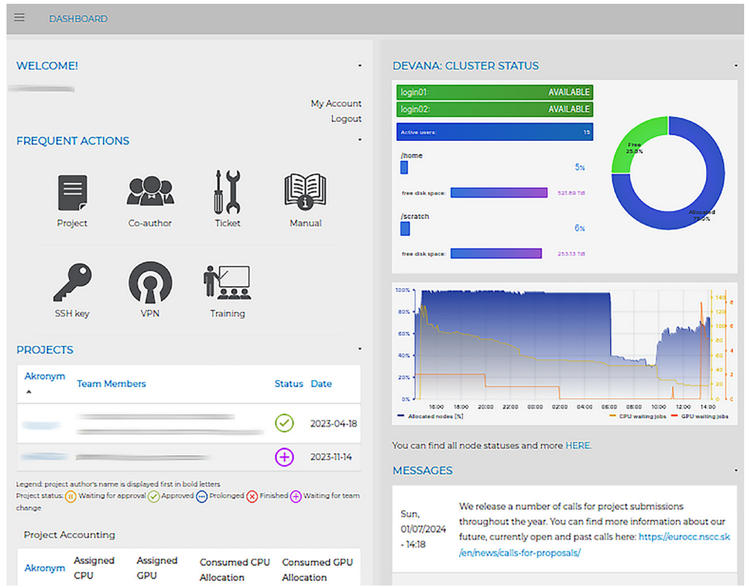


FIGURE 1

Sample dashboard in a user account on the HPC portal.

projects. Access for a test project can be requested at any time during the year, its aim should be benchmarking and obtaining data for subsequent submission of a standard project - this corresponds to a lower volume of computing capacity allocated for this type of project and a fast process of technical evaluation. The deadline for applications for standard projects, which are more complex and address specific scientific and research objectives, will be three times a year. These projects undergo technical evaluation as well as peer-review evaluation by independent experts in specific fields. The project duration is one year with the possibility of subsequent extension. To implement a commercial project, it is necessary to first conclude an individual contract with the operator according to general terms and conditions and the current price list.

More detailed practical information about access, HPC systems, software environment including manuals, documentation, and efficient usage of applications can be found by users and interested parties on the website userdocs.nscc.sk. Access to the HPC system itself is possible through the terminal with a command-line interface, but now also through a web interface. Access via a web browser is enabled by the new service Open OnDemand, which offers a user-friendly graphical environment for file management, running computational tasks, and interactive computing using applications such as Devana Desktop, Jupyter Notebook, or RStudio Server.

Devana is equipped with standard mathematical libraries and compilers including foss (free and open-source software) and Intel-based toolchains.

03 Application Software

Devana is equipped with standard mathematical libraries and compilers including foss (free and open-source software) and Intel-based toolchains. Users can also utilize containerization platform (operating system-level virtualization) Singularity for running computational tasks. Containerization allows the user to have control over the environment and „package" their software into a single file - a container, which is portable and reproducible across different platforms. Using Singularity, it is possible to create a container, for example, on your notebook and then run it on the HPC system.

Users can use their own software equipment if they have the appropriate license. Selected open-source applications, which have been among the most commonly used, are already available, and the personnel of the HPC center provide assistance with installing the necessary software based on users' requirements. As part of the National Competence Center for High-Performance Computing project (project number: 311071AKF2), popular software for multiphysics engineering simulations Ansys has also been procured and is now available to users. An overview of available applications and modules is continuously updated at userdocs.nscc.sk.

As part of the closed beta testing prior to the official launch of access, we approached several volunteers who helped us test the system's behavior under various types of loads and fine-tune the user environment:

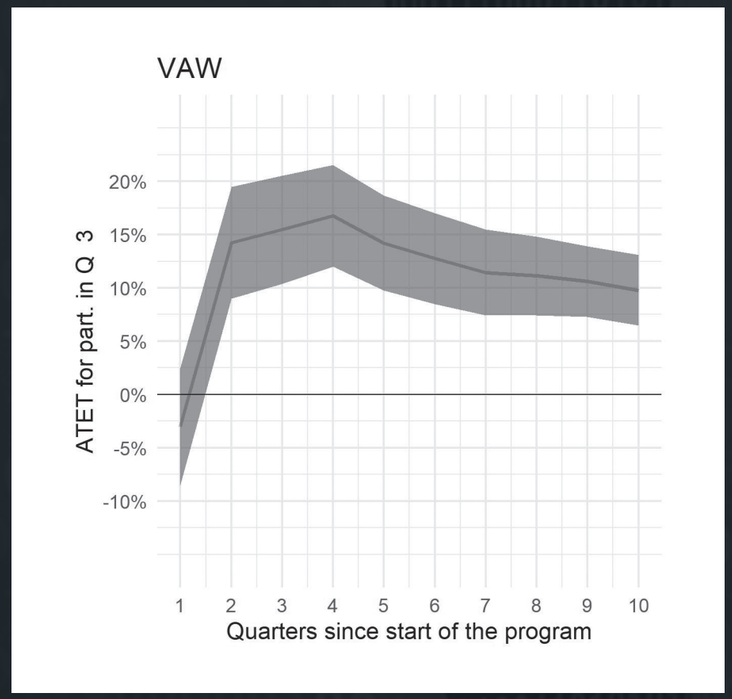


FIGURE 2

The Economic Institute of the Slovak Academy of Sciences (SAS) uses various statistical methods in its research, and the **DynamicALMP** project is now testing the use of HPC resources for processing extensive datasets. They are preparing an estimation model for the impact of participation in active labor market policy measures. The model is estimated based on administrative data on registered unemployed individuals. A dynamic estimator is used, taking into account the changing composition of participants with varying lengths of unemployment. Bootstrapping the errors of estimated coefficients with a sufficient number of repetitions is already pushing the limits of common PCs, hence the need to transfer the workload to an HPC environment.

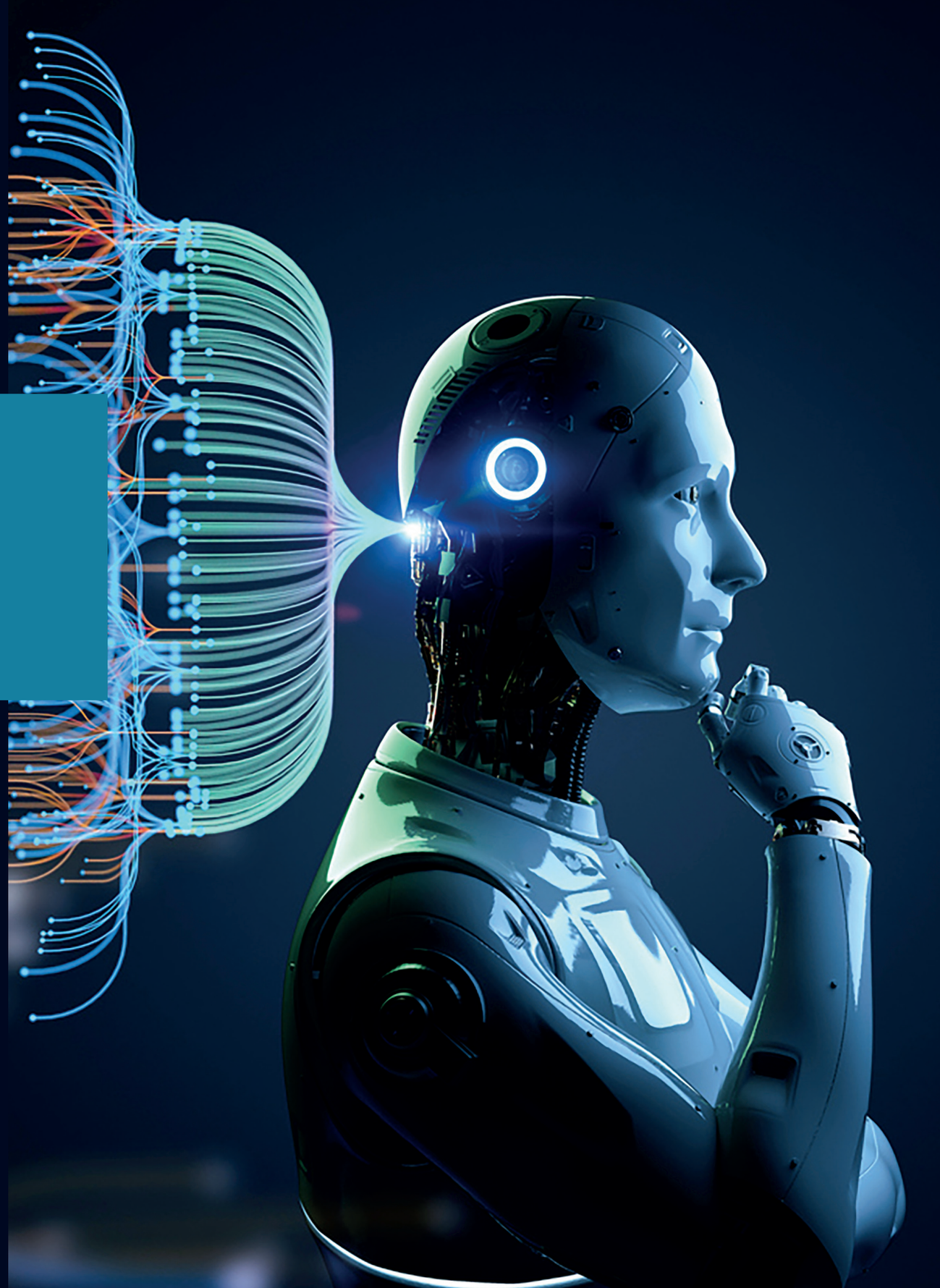
Testing different types of machine learning models in the HPC environment, with a focus on language and other computationally demanding models, is a project undertaken by a group of users from the **Kempelen Institute of Intelligent Technologies**. „In recent years, and especially in recent months, we have observed an accelerated pace of development of new, increasingly sophisticated and complex machine learning models. Dominant areas include natural language processing, where the language model ChatGPT (or models GPT-3.5 and GPT4) has achieved unprecedented success, and computer vision with generative models such as DALL-E, Stable Diffusion, and Midjourney. Within this testing project, we aim to enhance our own competencies in working with extremely complex and computationally demanding machine learning models, including currently the largest generative language

models. We then intend to incorporate these into our own research and existing as well as new research projects. We also aim to explore the characteristics of current complex models. For this purpose, it is necessary to train and subsequently perform inference with a large amount of data."

Devana opens up new perspectives for the Slovak scientific community and innovators.

Devana opens up new perspectives for the Slovak scientific community and innovators. We believe that it will contribute to the competitiveness of Slovak academic research and serve small and medium-sized enterprises and public administration in their efforts to develop and utilize more advanced digital solutions. Quality projects will certainly demonstrate the justification for further development of HPC in Slovakia and its application in various fields and industries.

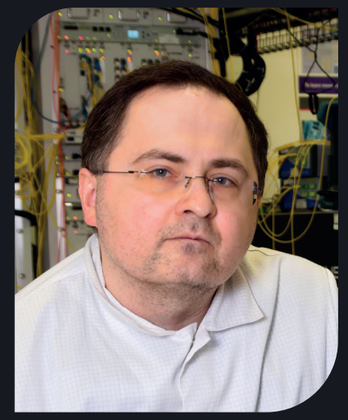
If you are a Slovak researcher or innovator, do not hesitate to seize the opportunity that Devana provides. Gain access, implement your projects, and become involved in the emerging HPC community in Slovakia.



QUANTUM TECHNOLOGIES BECOMING REALITY ACROSS EUROPE

In June 2023 the EuroHPC JU has signed hosting agreements with six sites across Europe to host & operate EuroHPC quantum computers. The six new EuroHPC quantum computers will be integrated into existing supercomputers in Czechia, France, Germany, Italy, Poland and Spain. The selection of these hosting entities, that will operate the systems on behalf of the EuroHPC JU, was made to ensure a diversity in quantum technologies and architectures, giving Europe an opportunity to be at the forefront of this still-novel field, and provide European users access to diverse and complementary quantum technologies. The quantum computers will be co-funded by the EuroHPC JU budget stemming from the Digital Europe Programme (DEP) and by contributions from the involved EuroHPC JU participating states. The JU will co-fund up to 50% of the total cost of the quantum computers with a planned total investment of more than EUR 100 million.

The EuroQCS-Poland consortium will host one of the quantum systems in the Poznan Supercomputing and Networking Center (PSNC), where it will be integrated into an existing HPC infrastructure. PSNC is leading the EuroQCS-Poland consortium consisting of two additional partners from Poland – Center for Theoretical Physics Polish Academy of Science and Creotech Instruments S.A, and one partner from Latvia – University of Latvia. The EuroQCS-Poland consortium aims to develop a quantum computer based on a trapped ions technology. We asked **Piotrovi Rydlichowskemu, Quantum technologies projects coordinator at PSNC**, several questions about the integration of the quantum computer in their datacenter and its potential for the scientific community.



Piotr Rydlichowski Ph.D. works as Quantum technologies projects coordinator at the Poznan Supercomputing and Networking Center.

There is a certain diversity in quantum computer technologies. What technology will your quantum computer be based on and why?

The EuroQCS-Poland consortium where PSNC is the leader and will host the machine aims to develop a quantum computer based on a trapped ions technology. This technology is already mature, and the ion traps themselves are well known and implemented in many other technologies such as atomic clocks. The very interesting property of this kind of quantum computing technology is the ability to implement flexible qubit connection topology. It is not predetermined, can be changed and adjusted according to specific use case and application scenario.

Will the quantum system be integrated with the existing (HPC) infrastructure? If so, is there a benefit to have them on the same premises?

EuroQCS-Poland infrastructure will be integrated with the existing PSNC HPC infrastructure in two ways. The first element is focused on the maintenance and access aspects to allow easy and efficient access to Quantum machine using existing EuroHPC tools and infrastructure. The second element will be focused on algorithms and this integration will allow to use hybrid classical-quantum algorithms and scenarios for research and integration. It is particularly interesting and fast-growing area. Quantum machines can potentially be efficient in solving specific problems and can complement existing HPC services and infrastructure. The benefits of this close HPC and QC integration at the same site – efficiency and ability for multiple layer and physical level access that cannot be achieved using distributed approach. It allows also for ef-

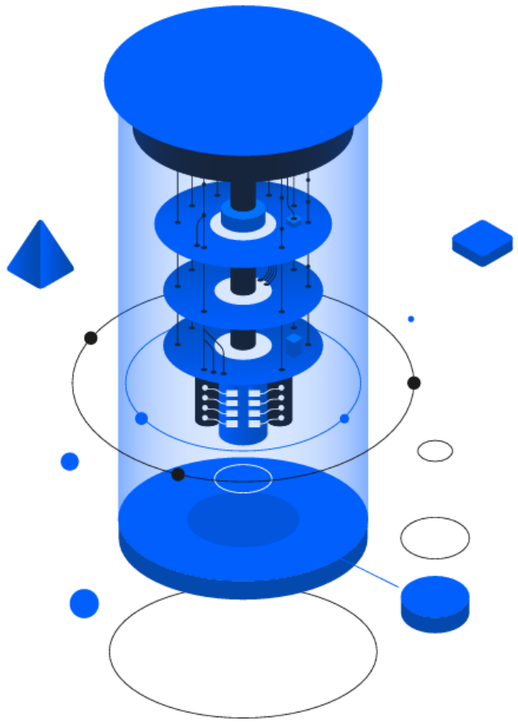


FIGURE 1

General layout of superconducting quantum computer.

ficient use and access to other resources required by QC and HPC infrastructure.

What are the biggest challenges of quantum computing nowadays that prohibit its application on real life problems?

Existing quantum computing technologies are developing in rapid way and also new ideas for physical qubit implementation emerge. We are still in the period of early development and exploring different technologies. It is time to test and verify multiple approaches and it is one of the reasons why EuroQCS will have in its infrastructure different quantum computing technologies – superconducting qubits, ion traps, optical etc. All these technologies have advantages and disadvantages that require further research. However, the biggest challenges that at the moment we can identify for quantum machines that limit its practical use are number of qubits, circuit depth and qubit readout errors. Number of qubits directly influence the size of problems that can be loaded and analyzed at given moment in the machine. Circuit depth and qubit readout errors are closely related. Quantum machines operate in terms of probability of a given result and qubit state readout process introduces errors that propagate along the circuit and limit the possible usable circuit depth and complexity. One of the key research areas now and in the future will be error correction and mitigation techniques that improve the quality of the circuits and its practical applications areas.

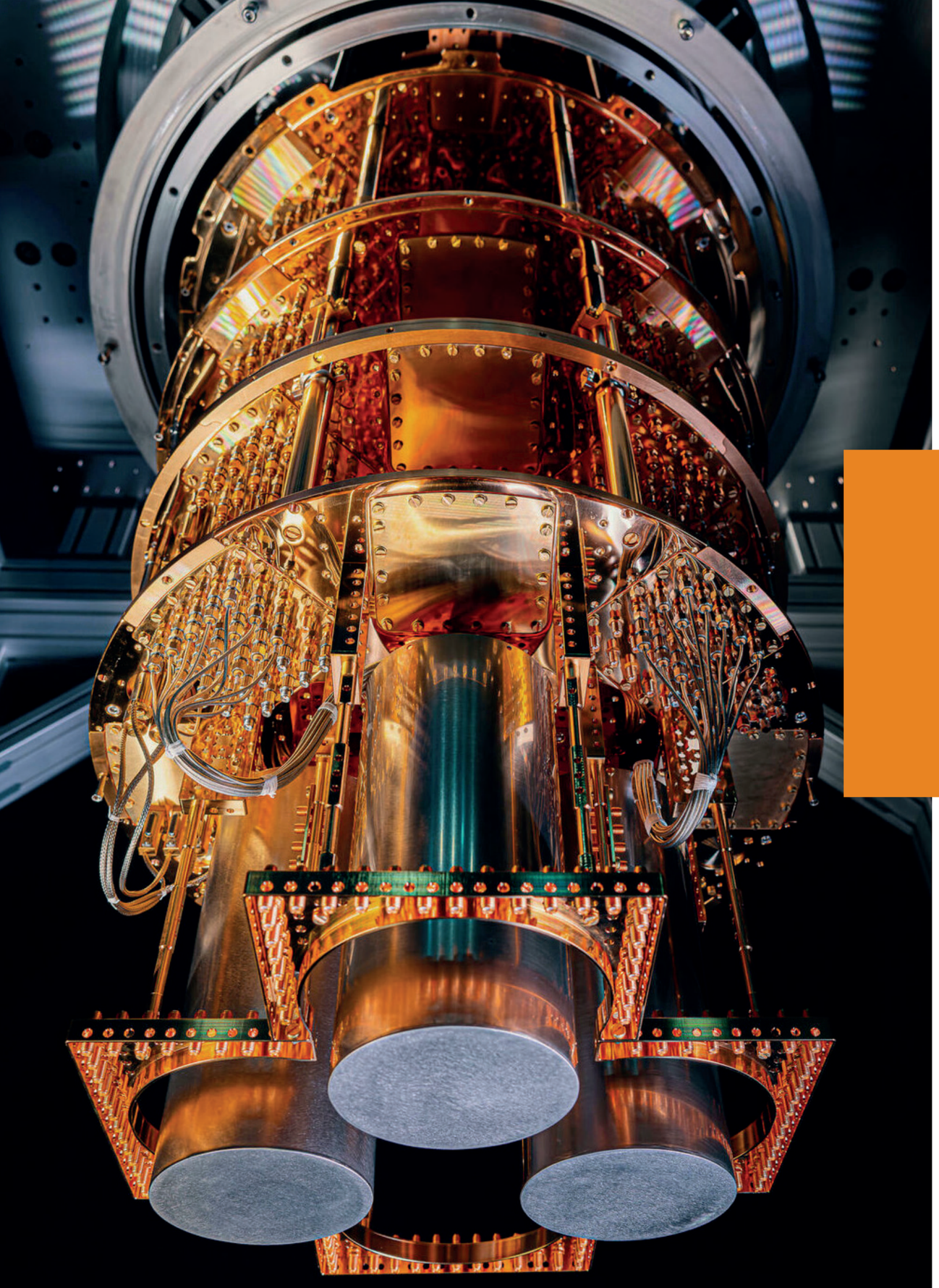
What kind of applications are you specifically planning to focus on, once the computer is operational?

PSNC is interested in number of QC applications that include pure quantum algorithms, hybrid QC-classical algorithms and quantum simulations. Due to PSNC close affiliation to Institute of Bioorganic Chemistry of the Polish Academy of Sciences quantum simulations in chemistry and medicine are of special interest. It is also worth to mention potential quantum machine learning applications in medical imaging analysis. Quantum simulations applications have the distinct characteristic that in many areas large number of qubits are not required as the main idea is to study quantum behavior of particular elements. There is large area of various quantum optimization use cases. There is also particular interest in integration with quantum communication scenarios.

Existing quantum computing technologies are developing in rapid way and also new ideas for physical qubit implementation emerge.

Are the necessary quantum algorithms for solving these problems available already, or will it require additional research?

Algorithms for the above-mentioned simulations, applications and scenarios are partly existing as the area is being extensively researched at the moment. Such examples are VQE (Variational Quantum Eigensolver) or QAOA (Quantum Approximate Optimization Algorithm). These already have been used in many examples and research scenarios. However, in some areas, such as quantum simulations, further research is required. Existing algorithms need to be constantly adapted to new hardware implementations, as well. There is also open field for potential new algorithms along with new use cases and applications.



Are/will quantum computers be suitable also for processing large amounts of data?

Quantum computing infrastructure can be also efficiently used for large problems with large data analysis, but it requires efficient classical-quantum interface and algorithms. Currently the limitation is the size of problems determined by number of qubits available in the system.

Many universities and schools introduce quantum mechanics courses that help better understand the effects and principles and quantum computing and communications.

What are the prerequisites for a scientist to be able to get involved in quantum computing?

Currently the input threshold to be involved in quantum computing or quantum communication is quite high as the use cases and implementations require good knowledge of quantum computing infrastructure also on physical level. It is like programming in assembler. Knowledge and understanding of quantum mechanics in many areas is also required. This is one of the reasons why many universities and schools introduce quantum mechanics courses that help better understand the effects and principles and quantum computing and communications. Once a higher-level programming interfaces are developed it may be easier to work with quantum computing infrastructure.

02

HPC Applications

hpc focus

Anomaly detection in time series data:

Gambling prevention using Deep Learning

BIBIÁNA LAJČINOVÁ
MARIÁN GALL
MICHAL PITOŇÁK

Gambling prevention of online casino players is a challenging ambition with positive impacts both on player's wellbeing, and for casino providers aiming for responsible gambling. To facilitate this, we propose an unsupervised deep learning method with an objective to identify players showing signs of problem gambling based on available data in a form of time series. We compare the transformer-based autoencoder architecture for anomaly detection proposed by us with recurrent neural network and convolutional neural network autoencoder architectures and highlight its advantages. Due to the fact that the players' clinical diagnosis was not part of the data at hand, we evaluated the outcome of our study by analyzing correlation of anomaly scores obtained from the autoencoder and several proxy indicators associated with the problem gambling reported in the literature.

Gambling prevention of players with problem or pathological gambling, currently conceptualized as a behavioral pattern where individuals stake an object of value (typically money) on the uncertain prospect of a larger reward [1], [2], is of high societal importance. Research over the past decade has revealed multiple similarities between pathological gambling and the substance use disorders [3]. With the high accessibility of the Internet, the incidence of pathological gambling has increased. This disorder can result in significant negative consequences for the affected individual and his/her family too. Therefore detecting early warning signs of problem gambling is crucial for maintaining player's wellbeing. This work is a joint effort of Slovak National Competence Center for High-performance Computing, DOXXbet, ltd. – sports betting and online casino, and Codium, ltd. – software developer of the DOXXbet sports betting and iGaming platform, with the goal to enhance customer service and players' engagement via identification and prevention of gambling behavior. This proof of concept is a foundation for future tools, which will help casino mitigate negative consequences for players, even for a price of less provision for the provider, as in line with European trends in risk management related to problem gambling.

We propose a completely unsupervised deep learning approach using transformer-based AE architecture to detect anomalies in the dataset - players with anomalous behavior. The dataset at hand does not comprehend the clinical diagnosis, and amongst other proxy indicators mentioned before only few are available - requests to increase spending limits, chasing losses by gambling more (referred to as chasing episodes later in this article), usage of multiple payment methods, frequent withdrawals of small amount of money and other mentioned later in the text. Clearly, not all the anomalous users must necessarily have problem gambling, hence the proxy indicators are used in combination with AE results, namely the anomaly score. The foundation of our approach rests on the idea that a compulsive gambler is an anomaly within the active casino players, with the literature mention-

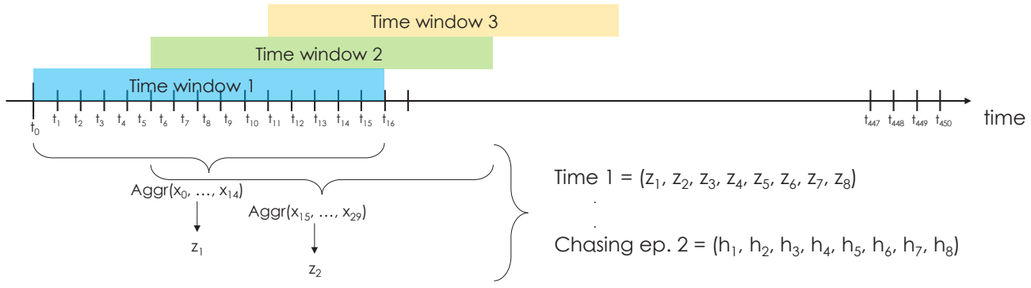
Gambling prevention of players with problem or pathological gambling, currently conceptualized as a behavioral pattern where individuals stake an object of value (typically money) on the uncertain prospect of a larger reward.

ing their fraction amongst all players being between 0.5% to 5% for chance-based games.

The data acquired for this research consist of sequences of data points collected over time, tracking multiple aspects of player's behavior such as frequency and timing of their gaming activities, frequency and amount of cash deposits, payment methods used when depositing cash, information about the bets, wins, losses, withdrawals and requests for change of deposit limit. Feature engineering resulted in 19 features in a form of time series (TS), so that each feature consists of multiple time stamps. These features can be classified into three categories - "time", "money" and "despair", as inspired by Seth et al. [4]. Each feature is a sequence of N values, where each value stands for one out of N consecutive time windows. This value was produced by aggregating daily data in the respective time window, with the time window lengths varying from 15 days to 3 months, either sliding or not. Hence, for each sample we needed a history of N time windows, where the value of N was set to 8 after several experiments. Feature engineering procedure is displayed in Figure 1 and the final data shape is depicted in Figure 2.

FIGURE 1

Visualization of the data aggregation from daily basis into time windows, and eventually to TS features.
$t_1, ..., t_{450}$ represent time stamps for daily data $x_1, ..., x_{450}$. Daily data points from a time window are aggregated into a single value z_i for all $i \in (1, \ldots, 8)$.



Model architecture

Autoencoder is an unsupervised Deep learning technique suitable for anomaly detection for TS data. The idea behind using this type of neural network to detect anomalies in data is based on reconstruction ability of the model. The AE model learns to reconstruct the data in the training set and since the training set is expected to contain mostly non anomalous

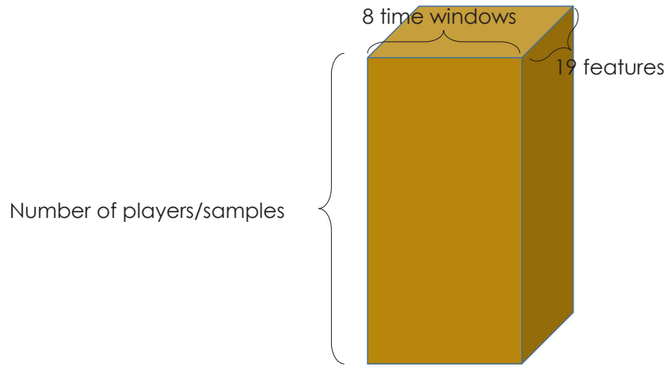


FIGURE 2

Final data shape obtained after feature engineering. Each sample is represented by 19 features consisting of 8 time windows.

data points the model learns to reconstruct well only samples that are normal. Therefore, when the input data sample is anomalous, trained AE model cannot reconstruct this data point well enough which results in high reconstruction error. Reconstruction error can be used to as anomaly score of a data sample, where a higher score indicates a greater likelihood of the sample being an anomaly.

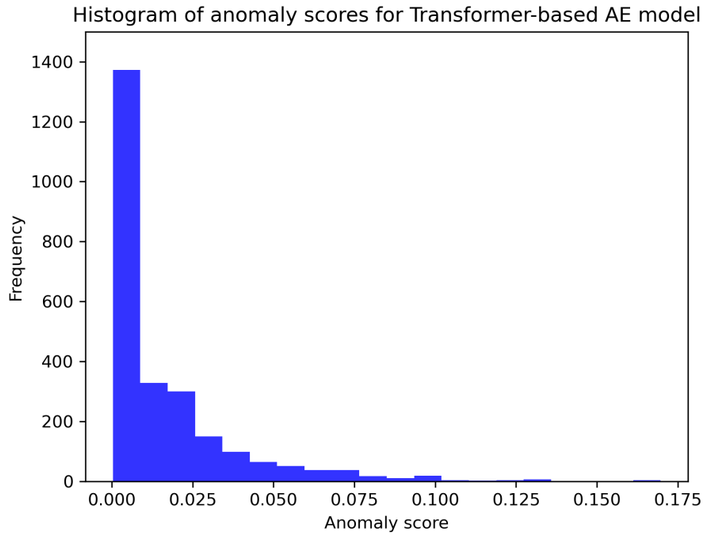
We trained AE with architecture based on transformers, where both encoder and decoder contain Multi Head Attention layer with four heads and 32 dimensional keys and values vectors. This self-attention layer is followed by Feed Forward Network with dropouts and residual connections between them, with overall about 100k parameters.

Reconstruction loss and the prediction ability

We performed a 3-fold cross-validation by splitting the data into training, validation, and test sets, and trained the model for each split to assess their stability. Generally, the test loss is observed to be always higher than train and validation losses. The reason for this is that those 211 data points that were removed from the training set in the data cleaning process, were moved to the test set. Without moving these samples the test loss for transformer-based model would be as low as 0.012. Model's performance is displayed on the Figure 3 as a histogram of loss values of the test set. Histogram has a heavy right tail, which is expected for datasets containing anomalies.

FIGURE 3

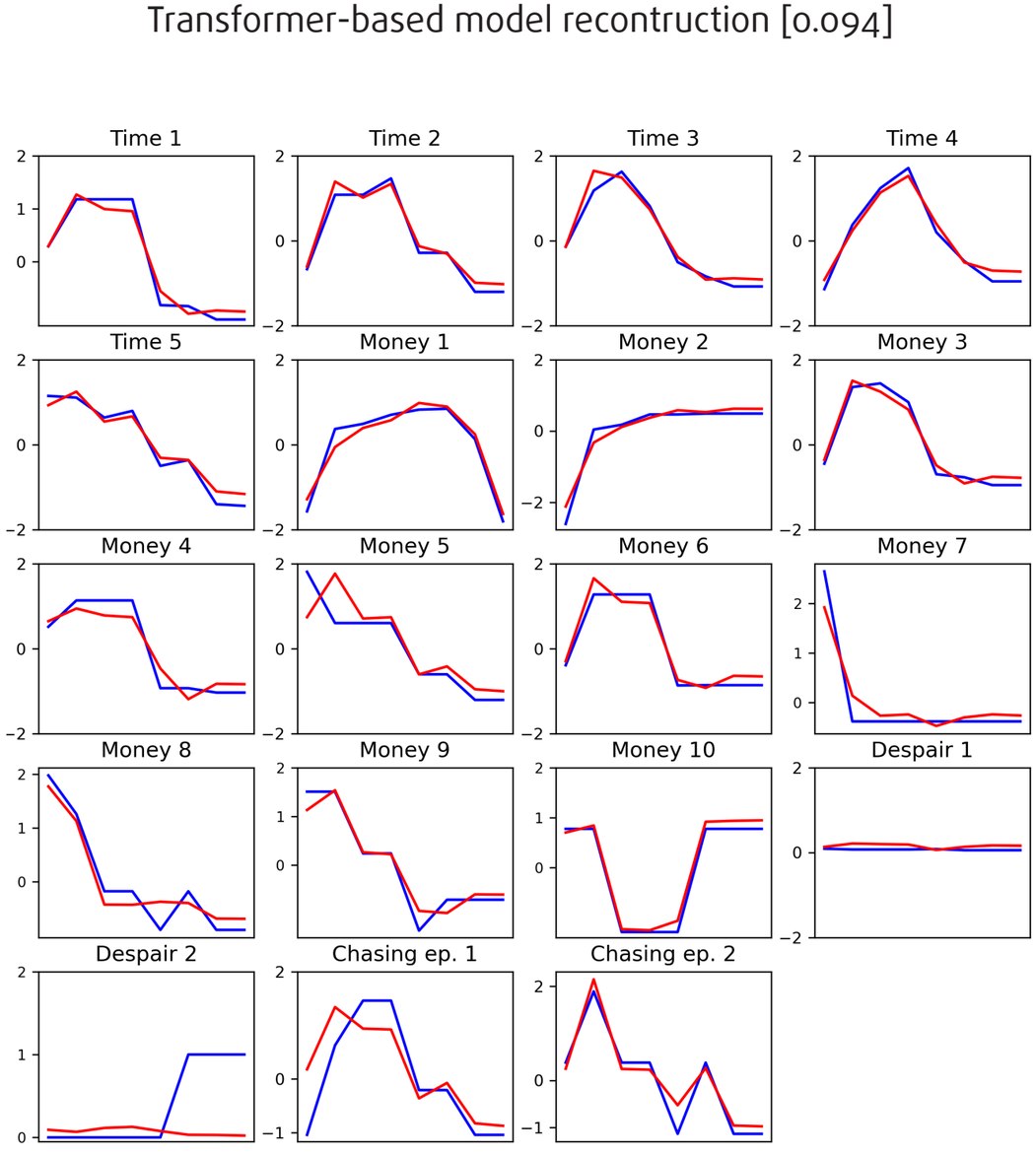
Histogram of the transformer reconstruction errors for the test set. The x-axis displays the value of anomaly score and the y-axis shows the frequency of the respective value.



To get a visual impression of how well is a TS reconstructed by all four evaluated models, Figure 4 depicts original (blue line) and predicted data (red line) for a randomly selected anomalous sample. Value of the anomaly score measured by MSE for each method is specified in the figure's heading.

FIGURE 4

Prediction performance, blue lines represent the original data and red lines represent the reconstruction by the Transformer AE model. The number in the heading is the anomaly score of the sample.



This proof of concept is a foundation for future tools, which will help casino mitigate negative consequences for players.

Results

Due to the lack of clinical diagnosis in our dataset, we can only rely on proxy indicators when identifying players with potential problem gambling. Our approach is to detect the anomalies in the dataset, but we are aware that not all the anomalies must indicate the gambling disorder. Therefore, we correlated the results of the AE model with these proxy indicators:

- Mean number of logins in a time window.
- Mean number of withdrawals in a time window.
- Mean number of small and frequent withdrawals in a time window.
- Mean number of requests for the change of the deposit limit in a time window.
- Sum of the chasing episodes in the time slot of N time window

Figure 5 depicts the correlation of the anomaly score with the proxy indicators. Each subplot contains 10 bars, each bar representing one decile of the data samples (i.e. each bar represents 10% of data samples sorted by anomaly score). The bar colors represent the category value of the respective proxy indicator.

A distinctive pattern in players' behavior can be observed, where players with larger anomaly scores tend to exhibit high values for all the indicators evaluated. Higher frequency of logins is proportionate to higher anomaly score with more than half of the players in the last decile of reconstruction error having a mean number of logins in a time window greater than 50. The same applies for mean number of cash withdrawals in a time window. Players with low anomaly score have almost none or very few withdrawals, whilst more than one fourth of players in the last anomaly score decile have two or more withdrawals in a time window on average. Another secondary indicator we utilize is the number of small and frequent withdrawals. Most of the players with at least one of these events is in 10% of players with the highest MSE. When analyzing another indicator, namely

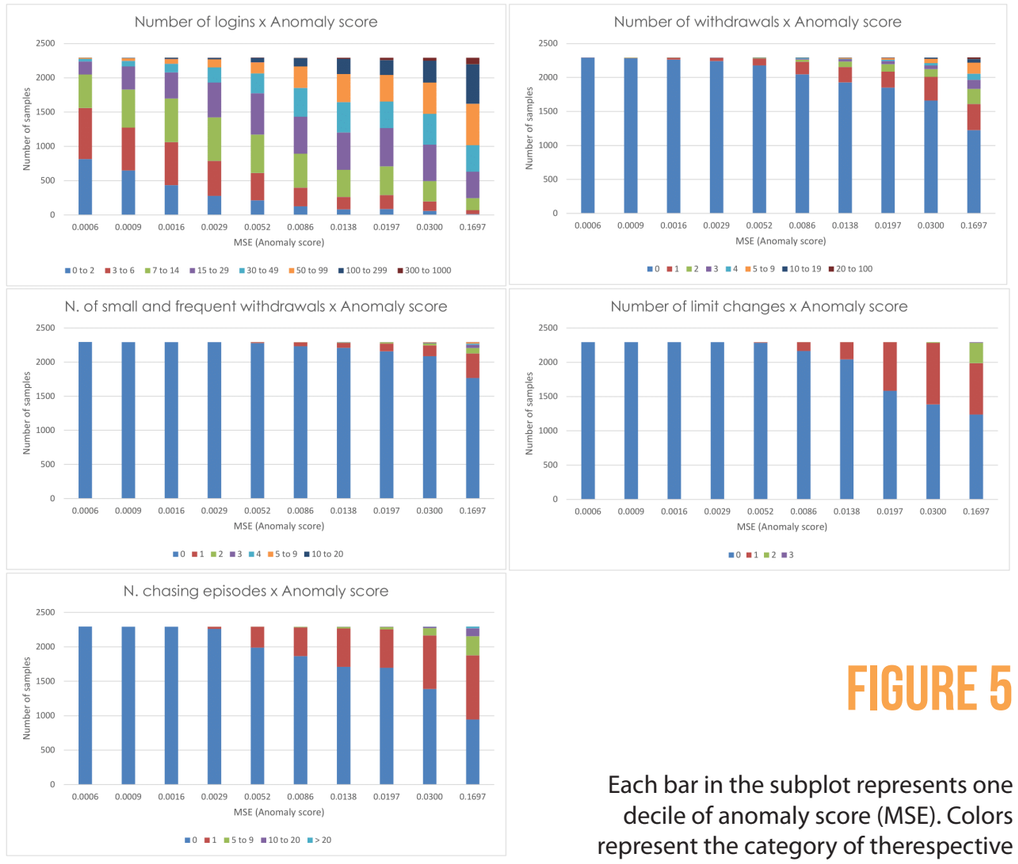


FIGURE 5

Each bar in the subplot represents one decile of anomaly score (MSE). Colors represent the category of therespective proxy indicator being analyzed with category values specified in the legend.

the number of requests for a deposit limit change, we observe a more subtle pattern. It is evident that players in the first five deciles generally have no requests for a limit change (with very few exceptions), while as the anomaly score increases, the frequency of limit change requests also tends to rise. The last proxy indicator depicted is the number of chasing episodes. A rising frequency of these events proportionate to their anomaly score can be observed. More than half of the players in the last decile have at least one chasing episode in the time window. If these plots are overlapped in order to identify the portion of players fulfilling multiple proxy indicators, following observations result: in the last five percentiles of the anomaly scores 98.6% of players satisfy at least one proxy indicator, and 77.3% satisfy at least three indicators. As for the last two percentiles, so 2% of players with the highest reconstruction error, almost 90% of them satisfy at least three indicators. The thresholds used to calculate these proportion are >= 1 chasing episode, >= 1 limit

Early detection of warning signs of problem gambling is key to maintaining the player's well-being, but also the financial situation of the affected persons.

change, >= 1 small and frequent withdrawal, >= 31 logins and >= 1.25 withdrawal on average per time window..

In this work, we successfully applied a transformer-based autoencoder (AE) to detect anomalies in the dataset of online casino players. The aim was to detect problem gamblers in dataset at hand in an unsupervised manner. 19 features were derived from the raw time series (TS) data reflecting players' behavior in the context of time, money and despair. We compared the performance of this architecture with three other AE architectures based on LSTM (Long Short-Term Memory) and convolutional layers and found that the transformer-based AE achieved the best results amongst the four models in terms of reconstruction capability. This model also showcases high correlation with proxy indicators such as the number of logins, number of player's with drawals, number of chasing episodes and other, that are commonly mentioned in literature in relation to the gambling disorder. This alignment of AE's anomaly score with proxy indicators enables us to gain insights into rediction's effectiveness in identifying players with potential problem gambling.

Even though these proxy indicators were also used as predictors, we suggest to use them as a secondary check when detecting players with potential problem gambling in order to avoid false positives, as not all anomalies must be linked to the condition of gambling disorder.

Our findings demonstrate the potential of transformer-based AEs for unsupervised anomaly detection tasks in TS data, particularly in the context of online casino player behavior.

The full version of the paper is available on the website of the National Competence Center.

REFERENCES

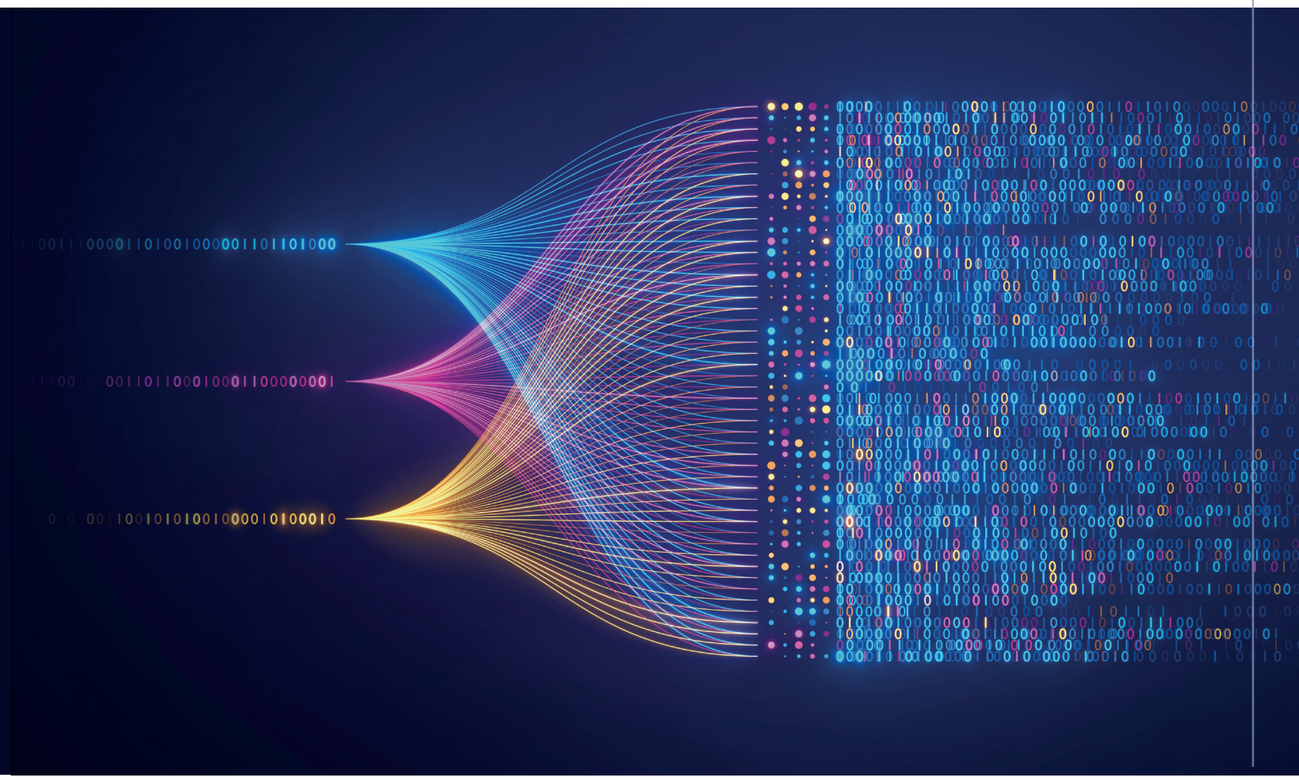
[1] Alex Blaszczynski and Lia Nower. *A Pathways Model of Problem and Pathological Gambling*. In: Addiction (Abingdon, England) 97 (June 2002), pp. 487–99. doi: 10.1046/j.1360-0443.2002.00015.x.

[2] National Research Council. *Pathological Gambling: A Critical Review*. Washington, DC: The National Academies Press, 1999. isbn: 978-0-309-06571-9. doi: 10.17226/6329. url: https://nap.nationalacademies.org/catalog/6329/pathological-gambling-a-critical-review.

[3] Luke Clark et al. *Pathological Choice: The Neuroscience of Gambling and Gambling Addiction*. In: Journal of Neuroscience 33.45 (2013), pp. 17617–17623. issn: 0270-6474. doi: 0.1523/JNEUROSCI.3231-13.2013. url: https://www.jneurosci.org/content/33/45/17617.

[4] Deepanshi Seth et al. *A Deep Learning Framework for Ensuring Responsible Play in Skill-based Cash Gaming*. In: 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA) (2020), pp. 454–459.

Measurement of microcapsule structural parameters using **artificial intelligence (AI)** & **machine learning (ML)**



The main aim of collaboration between the National Competence Centre for HPC (NCC HPC) and the Institute of Polymers of SAS (IP SAS) was design and implementation of a pilot software solution for automatic processing of polymer microcapsules images using artificial intelligence (AI) and machine learning (ML) approach. The microcapsules consist of semi-permeable polymeric membrane which was developed at the IP SAV.

Automatic image processing has several benefits for IP SAV. It will save time since manual measurement of microcapsule structural parameters is time-consuming due to a huge number of images produced during the process. In addition, the automatic image processing will minimize the errors which are inevitably connected with manual measurements. The images from optical microscope obtained with 4.0 zoom usually contain one or more microcapsules, and they represent an input for AI/ML process. On the other hand, the images from optical microscope obtained with 2.5 zoom usually contain (three to seven) microcapsules. Herein, a detection of the particular microcapsule is essential.

The images from optical microscope are processed in two steps. The first one is a localization and detection of the microcapsule, the second one consists of a series of operations leading to obtaining structural parameters of the microcapsules.

Microcapsule detection

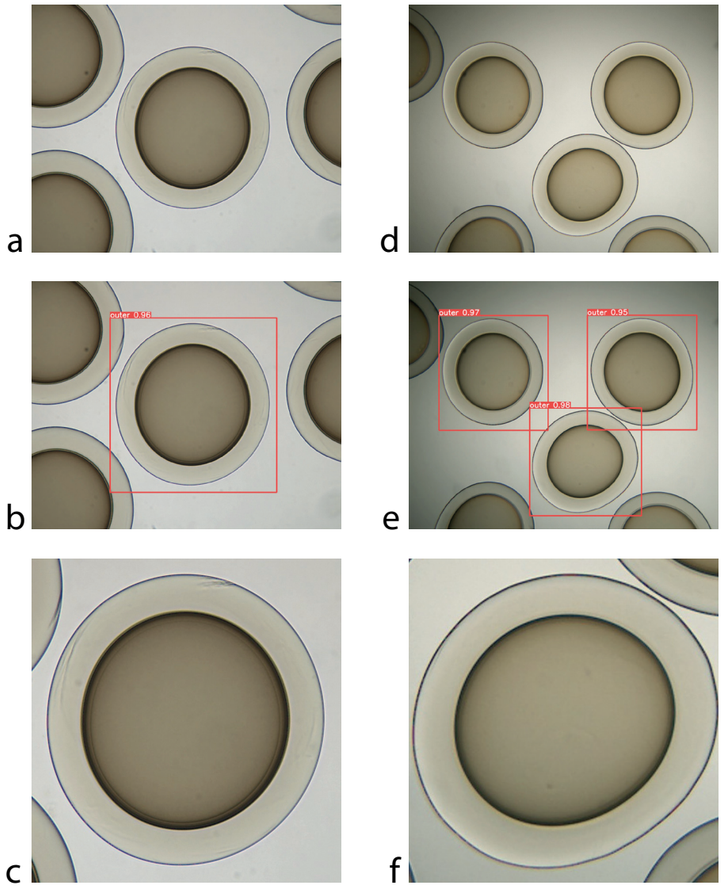
YOLOv5 model [1] with pre-trained weights from COCO128 dataset [2] was employed for microcapsule detection. Training set consisted of 96 images, which were manually annotated using graphical image annotation tool LabelImg [3]. Training unit consisted of 300 epochs, images were subdivided into 6 batches per 16 images and the image size was set to 640 pixels. Computational time of one training unit on the NVIDIA GeForce GTX 1650 GPU was approximately 3.5 hours.

The detection using the trained YOLOv5 model is presented in Figure 1. The reliability of the trained model, verified on 12 images, was 96%, with the throughput on the same graphics card being approximately 40 frames per second.

Microcapsules serve as packaging for pancreatic islets forming a promising drug for diabetes mellitus.

FIGURE 1

(a) microcapsule image from optical microscope (b) detected microcapsule (c) cropped detected microcapsule for 4.0 zoom, (d) microcapsule image from optical microscope (e) detected microcapsule (f) cropped detected microcapsule for 2.5 zoom.



Measurement of microcapsule structural parameters using AI/ML

The binary masks of inner and outer membrane of the microcapsules are created individually, as an output from the deep-learning neural network of the U-Net architecture [4]. This neural network was developed for image processing in biomedicine applications. The first training set for the U-Net neural network consisted of 140 images obtained from 4.0 zoom with the corresponding masks and the second set consisted of 140 images obtained from 2.5 zoom with the corresponding masks. The training unit consisted of 200 epochs, images were subdivided into 7 batches per 20 images and the image size was set to 1280 pixels (4.0 zoom) or 640 pixels (2.5 zoom). The 10% of the images were used for validation. Reliability of the trained model, verified on 20 images, exceeded 96%. Training process lasted less than 2 hours on the HPC system with IBM Power 7 type nodes, and it had to be repeated several times. Obtained binary masks were subsequently post-processed using fill-holes [5] and watershed [6], operations, to get rid of the unwanted residues. Subsequently, the binary masks were fitted with an ellipse using scikit-image measure library [7]. First and second principal axis of the fitted ellipse are used for the calculation of the microcapsule structural parameters. An example of inner and outer binary masks, and the fitted ellipses is shown in Figure 2.

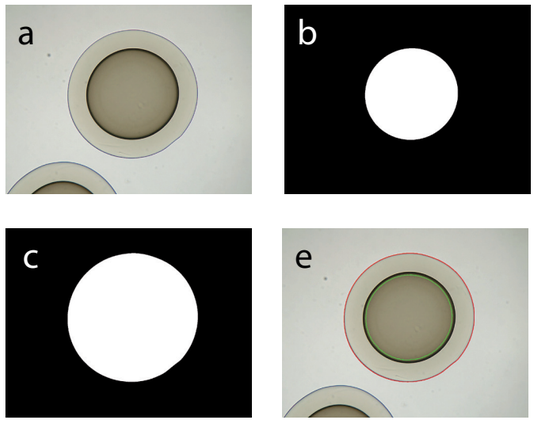
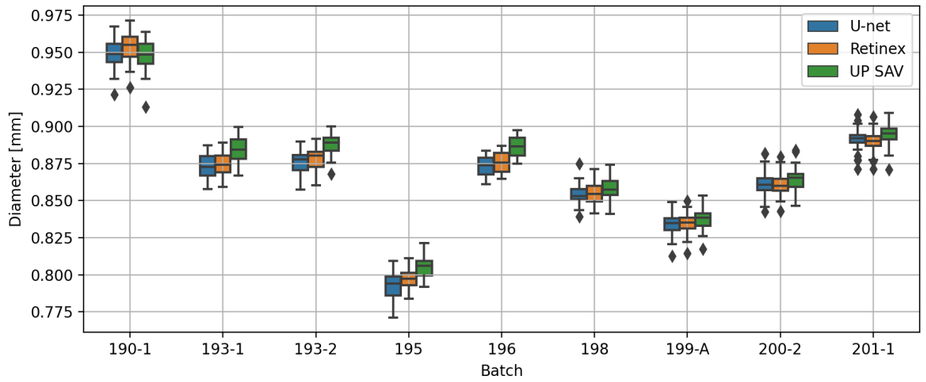


FIGURE 2

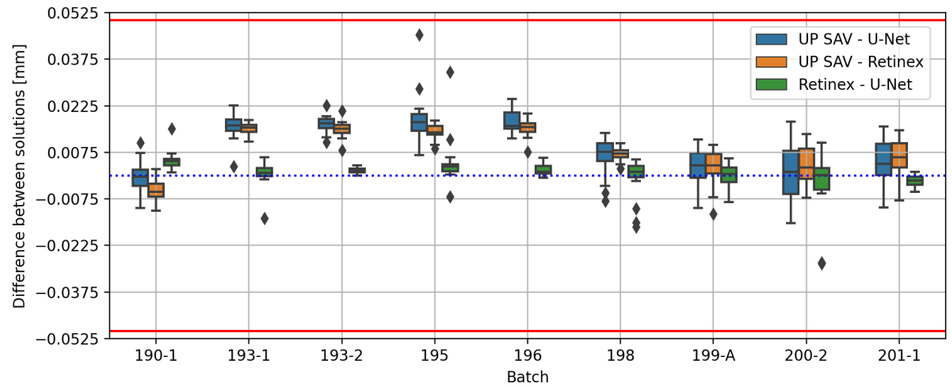
(a) input image from optical microscope
(b) inner binary mask
(c) outer binary mask
(d) output image with fitted ellipses.

Structural parameters obtained by our AI/ML approach (denoted as “U-Net”) were compared to the ones obtained by manual measurements performed at the IP SAV. A different model (denoted as “Retinex”) was used as another independent source of reference data. The Retinex approach was implemented by **RNDr. Andrej Lúčny, PhD.** from the Department of Applied Informatics of the Faculty of Mathematics, Physics and Informatics in Bratislava. This approach is not based on the AI/ML, the ellipse fitting is performed by the aggregation of line elements with low curvature using so-called retinex filler [8]. The Retinex approach is a good reference due to its relatively high precision, but it is not fully automatic, especially for the inner membrane of the microcapsule.

a



b



C

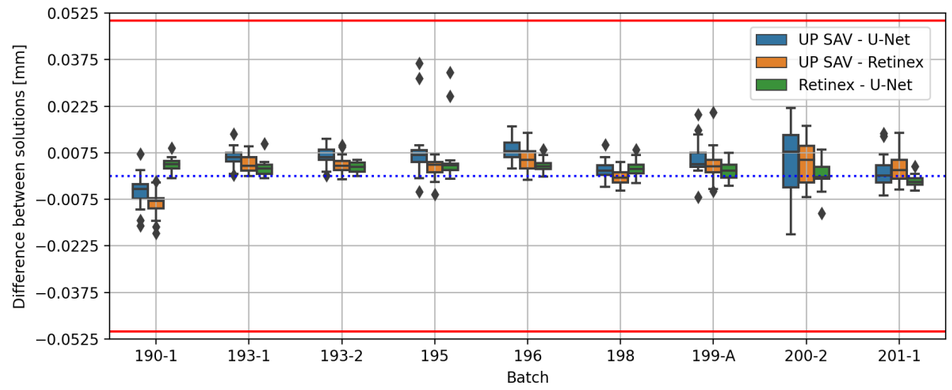


FIGURE 3

(a) microcapsule diameter for different batches
(b) difference between the diameters of the fitted ellipse (first principal axis) and microcapsule
(c) difference between the diameters of the fitted ellipse (second principal axis) and microcapsule.
Red lines in (b) and (c) represents the threshold given by IP SAV.
The images were obtained using 4.0 zoom.

Figure 3 summarizes a comparison between the three approaches (U-Net, Retinex, UP SAV) to obtain the 4.0 zoom microcapsule structural parameters.

All obtained results, except 4 images of batch 194 (ca 1.5%), are within the threshold defined by the IP SAV. As can be seen from Figure 3(a), the microcapsule diameters calculated using U-net and Retinex are in a good agreement to each other. The U-Net model performance can be significantly improved in future, either by the training set expansion or by additional post-processing. The agreement between the manual measurement and the U-Net/Retinex may be further improved by unifying the method of obtaining microcapsule structural parameters from binary masks.

The AI/ML model will be available as a cloud solution on the HPC systems of CoO SAS. Additional investment into the HPC infrastructure of IP SAS will not be necessary. Production phase, which goes beyond the scope of the pilot solution, accounts for an integration of this approach into the desktop application.

REFERENCES

[1] https://github.com/ultralytics/yolov5

[2] https://www.kaggle.com/ultralytics/coco128

[3] https://github.com/heartexlabs/labelImg

[4] https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/

[5] https://docs.scipy.org/doc/scipy/reference/generated/scipy.ndimage.binary_fill_holes.html

[6] https://scikit-image.org/docs/stable/auto_examples/segmentation/plot_watershed.html

[7] https://scikit-image.org/docs/stable/api/skimage.measure.html

[8] D.J. Jobson, Z. Rahman, G.A. Woodell, IEEE Transactions on Image Processing 6 (7) 965-976, 1997.

COMPUTATIONAL MATERIALS DESIGN RESEARCH GROUP IN COLLABORATION WITH THE DEPARTMENT OF ION BEAM TECHNOLOGIES at ATRI MTF STU IN TRNAVA

NOVEL OXIDES OF NICKEL, COPPER & PALLADIUM

Following our previous article (HPC FOCUS 2022), which was dedicated to computer modelling of silver halides, in this contribution, we will focus on introducing our research of new oxides of nickel, copper, and palladium.

Nickel, copper, and palladium belong to the main transition elements in the Periodic Table of Elements, also known as the d-block elements, because their valence (or chemically active) electrons are located in d orbitals. This group of elements is characteristic of a diversity of physical and chemical properties, resulting in their significant importance in both fundamental and applied research as well as in industry. They often appear in the form of compounds with other elements, and it is the interaction with other elements that contributes the most to the diversity of their properties. Therefore, the study of new compounds of transition metals is among the primary tasks of basic research.

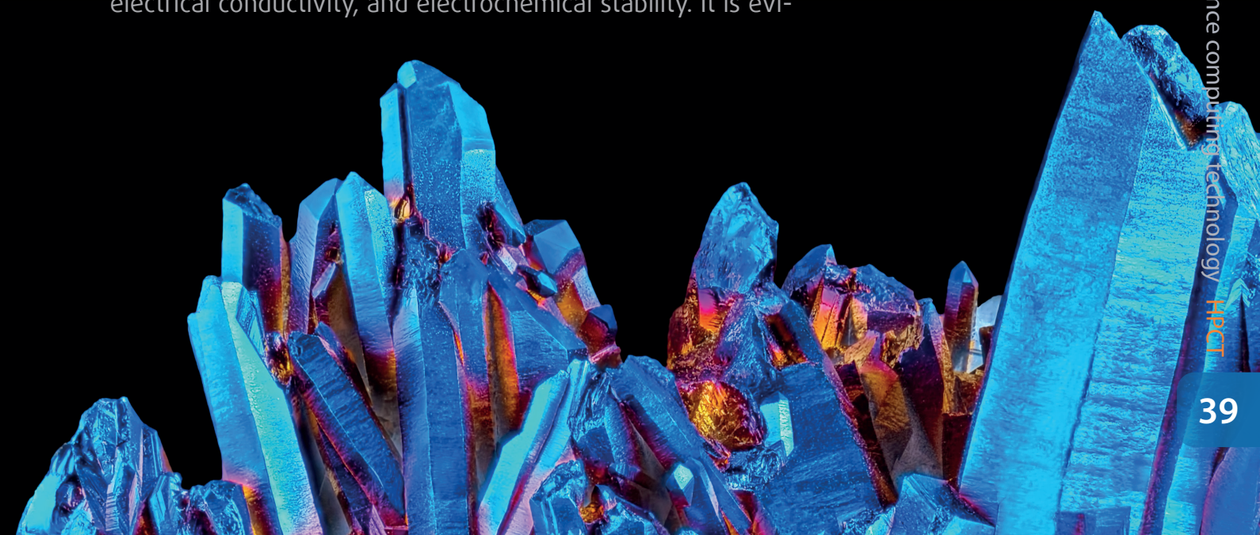
The most significant and extensively studied compounds of transition metals are their oxides. This is because oxygen is one of the most common chemical elements on Earth, and transition metals readily react with it under natural conditions, whether in soil, aqueous environments, or in the air. It's a common natural phenomenon well known, for example, as the chemical corrosion of metals, where the contact of a metal with oxygen results in the formation of a more stable oxide. Although corrosion is often associated with material degradation or deterioration, in general, the formation of transition metal oxides significantly contributes to the diversification of their properties and, consequently, their functionality.

Transition metal oxides are particularly important for modern technological applications because they exhibit novel phenomena such as high-temperature superconductivity, giant magnetoresistance, and various new magnetic and electrical properties that can be intricately coupled. This makes oxides candidates for multifunctional materials capable of performing multiple functions simultaneously with minimal intervention.

The formation of oxides can also lead to the improvement of already existing attractive properties of transition metals, such as their catalytic abilities and redox properties, which are significant in the development of new energy materials.

From the perspective of the significant potential of transition metal oxides, it is especially noteworthy that each transition element can form many compounds with oxygen, varying in the ratio of atoms of the transition metal and oxygen. Additionally, each of these compounds can exist in various crystal structure types, characterized by different arrangements of atoms, and thus, a different set of properties or modifications. An excellent example of this is titanium and its oxides. Titanium is an exceptionally strong and lightweight metal with high corrosion resistance in aqueous environments, making it widely used in various applications, ranging from kitchenware, sports equipment, structural components to coatings, from medicine to the navy. In combination with oxygen, it forms over 20 different compounds or phases Ti_2O, TiO, TiO_2, Ti_2O_3, etc.), and each of them can adopt different crystal structures (for TiO_2 alone, there are over 10 crystal structures known). This diversity gives rise to countless additional properties and applications. For example, the most well-known form of TiO2, known as rutile, has excellent properties as a pigment blocking UV radiation, making it widely used as white pigment and UV filter in cosmetics. Several forms of TiO2 are promising candidates for photovoltaic applications. Another group of titanium oxides with a general formula Ti_nO_{2n-1} (n=2,3, ...) known as the Magnéli phases, named after their discoverer Arne Magnéli, holds great technological potential due to their high resistance to corrosion in acidic and alkaline solutions, electrical conductivity, and electrochemical stability. It is evi-

Corrosion is often associated with material degradation or deterioration, in general, the formation of transition metal oxides significantly contributes to the diversification of their properties and, consequently, their functionality.



The Computational Materials Design Research Group at MFT STU in Trnava was founded by Dr. hab. Mariana Derzsi (left). Together with Dr. Kamil Tokár (right) they lead a young team consisting of undergraduate, graduate, PhD. students and postdocs.

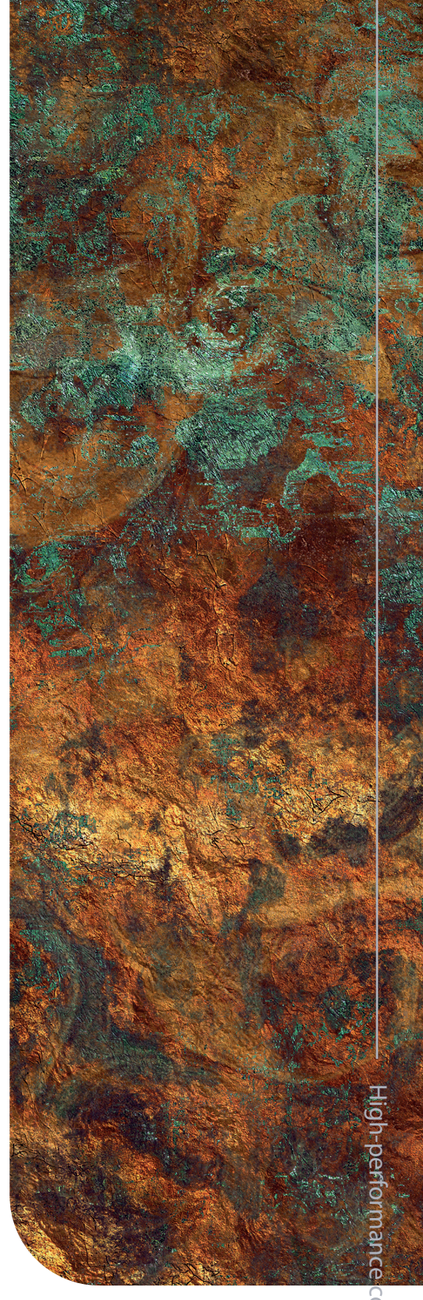
dent that such a vast diversity of Ti-O phases provides ample room for harnessing their technological potential.

Our goal is to achieve a similar structural diversity in the oxides of nickel, copper, and palladium, which, unlike titanium, are far less explored. All three metals have significant industrial applications, especially in construction and the automotive industry. Copper is known for its exceptional electrical conductivity, often used in electrical cables. Nickel, as a key component of stainless steel, enhances its ductility, weldability, tensile strength, and high-temperature corrosion resistance. Palladium plays a crucial role in catalytic converters for automobiles, converting toxic emissions from vehicles into less harmful substances. Although the oxides of these metals are known, their phase diversity lags significantly behind the one that we know, for example, for titanium oxides. Current research strongly suggests the exceptional nature of these oxides, particularly in electronic applications and energy-related fields.

In our research, we primarily focus on studying the phase and structural diversity of nickel, copper, and palladium oxides. Through computer simulations based on quantum-mechanical calculations and evolutionary algorithms, we investigate the possibility of forming a wide range of phases, propose their crystal structures, and predict the properties that characterize them.

The dominant tool for discovering new materials today is computer modelling based on quantum-mechanical calculations. A critical step preceding quantum-mechanical calculations is the design of the molecular structure model of the desired compound. Our initial source of input models comes from databases that store information about all known types of molecular and crystal structures. The prediction of the structure of a new compound begins with selecting known structure types for the given stoichiometry. For example, if our proposed compound is Cu_2O_3, which contains two chemical elements, Cu and O, in a ratio of 2:3, we will select structures from the database for all known compounds containing only two chemical elements, let's call them A and B, in the same 2:3 ratio. In the selected A_2B_3 structures, we then replace element A with copper and element B with oxygen. These input models are subjected to a „correctness test" in the form of quantum-mechanical calculations. Quantum mechanics is necessary for the accurate description of interatomic interactions that dictate the arrangement of atoms in the structure. The optimal arrangement is one where all forces between atoms are balanced, resulting in the structure having minimum energy. In predicting new compounds, quantum-mechanical calculations are therefore used to optimize the proposed models. During the optimization process, all atoms are moved until their optimal arrangement is achieved. Subsequent calculations assess the stability of the optimized models under various conditions, including the possible synthesis routes, and calculate their properties. The result of all these calculations is the prediction of the existence of a new compound, which includes a description of the expected structure, key properties, and stability conditions.

The procedure described above for selecting input models does not always lead to the prediction of the most optimal structure. Therefore, more sophisticated methods, such as evolutionary algorithms, are increasingly used for this purpose. These algorithms do not rely on structural databases but generate random atomic arrangements while adhering to a minimal set of rules, such as chemically sensible atomic spacings. Models generated in this way also undergo quantum-mechanical optimization. The difference is that the optimization process does not end but begins here. In the next step, models with the lowest energies are selected, and they are used to create a new generation of models, which are again subject to optimization. This entire process is repeated several times, generating dozens of new generations, dramatically increasing the chances of finding the optimal structure. The way new



An extensive study of the Cu-O system was conducted by Dr. Sankari Sampath, who was part of our research group from 2019 to 2021.

generations of models are created is inspired by processes in biological evolution, such as reproduction (creating a n offspring model from two parent models) or mutation (introducing various changes into selected models).

Below, we present some of our first interesting results obtained in the study of new nickel, copper, and palladium oxides using the theoretical methods introduced earlier.

Copper
Cu
29

Copper

Binary copper oxides [Cu-O] are receiving significant attention from the broad scientific community due to their electronic and optical properties, making them promising candidates for photovoltaics and all-oxide electronics. They are excellent light absorbers (electromagnetic radiation across the entire visible spectrum), exhibit outstanding stability, are non-toxic, and are easily manufacturable. Furthermore, they form an extensive family of semiconductors and superconductors, ranking them among the most studied transition metal oxides. To date, only three binary phases are known: Cu_2O, $Cu^{II}O$ a $Cu^{I}_2Cu^{II}_2O_3$, even though there are numerous hints in the scientific literature of possible formation of various other stoichiometries.

One of the interesting results we obtained while studying the Cu-O system pertains to the group of binary oxides with the general chemical formula M_2O_3, known as sesquioxides. Sesquioxides can be formed by both metallic and non-metallic elements, with the largest group consisting of sesquioxides of transition metals. Among the sesquioxides of transition metals from the 4th period, only Cu_2O_3 remains unknown to this day. Our results obtained through evolutionary algorithms helped us unveil its chemical and structural nature (Figure 1). We found that Cu_2O_3 contains two types of oxygen anions, namely the common O^{2-} and the superoxide anion O^{-}_2. The presence of these two different anions makes Cu_2O_3 exceptional among all known sesquioxides. Most sesquioxides contain the common O^{2-} anion. An exception is found in the sesquioxides of alkali metals, which contain the peroxide anion O^{2-}_2 and the superoxide anion O^{-}_2. The presence of both peroxide and superoxide anions in alkali metal sesquioxides can be easily understood. The presence of the common oxygen anion O^{2-} requires the presence of a cation with an oxidation state of 3+ to ensure the compound's neutrality ($M^{3+}_2O^{2-}_3$). Since alkali metals can only have an oxidation state of 1+ in compounds, the only way they can form sesquioxides with oxygen is if the oxygen atoms within them create superoxide and peroxide anions (e.g., $Rb_2O_3 = Rb^{+}_4(O^{-}_2)_2O^{2-}_2$). Returning to the unknown Cu_2O_3, the presence of the superoxide anion in its structure suggests the impossibility of stabilizing Cu^{3+} in copper sesquioxide. Copper stands out with this property among other transition metals, all of which exhibit an oxidation state of 3+ in sesquioxides.

A comprehensive study of the Cu-O system was undertaken by Dr. Sankari Sampath, who was a part of our research group from 2019 to 2021. Dr. Sampath earned her Ph.D. in Materials Science from Ruhr-Universität Bochum in Germany and focused on DFT modelling of the Cu-O system, thanks to the Homecoming grant of the Ministry of Education, Science, Research, and Sport of the Slovak Republic and a postdoctoral grant from the Slovak University of Technology in Bratislava.

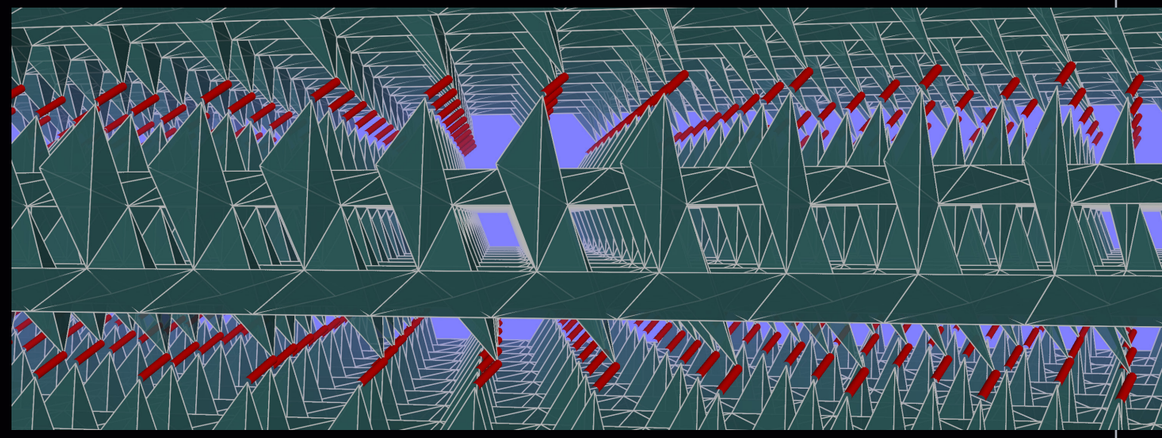


FIGURE 1

New Cu_2O_3 structure predicted using evolutionary algorithms. The illustration highlights layers of the [CuO_4] units (dark green polyhedrons) linked via oxygen O-O bridges (red linkers).

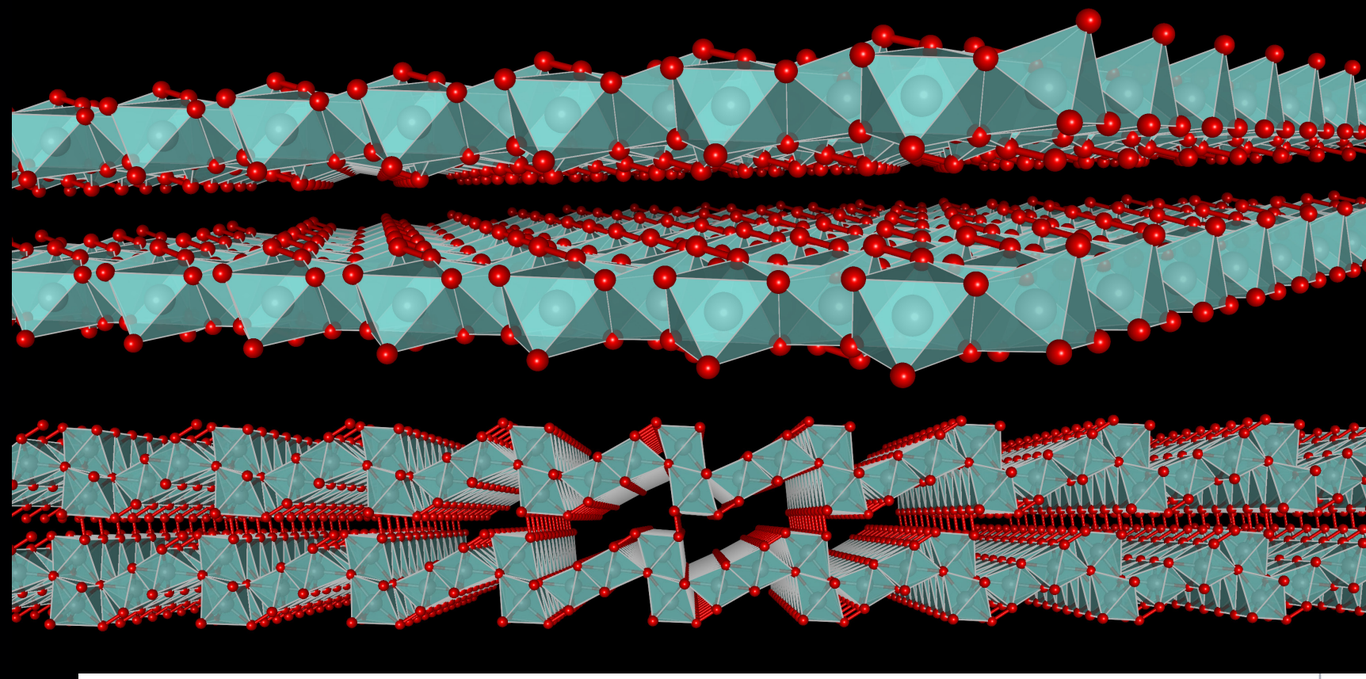
Ing. Radovan Bujdák is a PhD student at MTF STU. His PhD project is dedicated to predictive DFT+EA modelling of new binary nickel oxide phases.

Nickel
Ni
28

Nickel

The binary nickel-oxygen system Ni-O contains only one well-characterized phase, nickel oxide (NiO). Often considered as a known phase is also nickel sesquioxide, Ni_2O_3, which is also known as black nickel. Despite its familiar name, it is striking that Ni_2O_3 is poorly characterized, and its crystal structure remains unknown. Given that black nickel is increasingly being used as a component in many functional materials due to its declared photocatalytic activity, Ni_2O_3 was the first phase we began studying in the Ni-O system.

Like the previously mentioned copper sesquioxide Cu2O3, Ni2O3 also exhibits unusual behavior among known sesquioxides. Our quantum-mechanical calculations suggest the presence of two different nickel cations. Evidence for their presence can be seen in the different lengths of chemical bonds formed with oxygen anions, as well as different magnetic moment values. This is unusual because the vast majority of sesquioxides contain only one type of cation. Ni^2O_3, therefore, mimics the behavior of ternary oxides that contain two different metallic elements (e.g., Fe and Ti) and are known as ilmenites ($FeTi_2O_3$). This finding could have significant implications for understanding the declared photocatalytic activity of Ni_2O_3, as the presence of two different Ni cations may be a key factor.



As the second phase of interest, we were intrigued by nickel pentoxide, Ni_2O_5. Nickel pentoxide represents a completely unknown and unexplored phase. It caught our attention because pentoxides of transition metals are known and exhibit rich polymorphism with varying levels of structural complexity. The oxidation state of 5+, which is characteristic of pentoxides, is exotic in the case of nickel. On the other hand, the d5 electron configuration, which corresponds to the Ni^{5+} cation, with each d orbital occupied by one electron, is one of the most stable ones. Our initial results suggest that under normal conditions, Ni^{5+} tends to reduce in Ni_2O_5, leading to the formation of oxygen bridges. This means that we can expect the formation of peroxide or superoxide in the form of attractive layered structures (Figure 2).

The study of Ni_2O_3 was undertaken by Ing. Michaela Gašpárková as part of her master's thesis and subsequent scientific research activities from 2019 to 2021. Following her work on Ni_2O_3, Ing. Radovan Bujdák continued the research. Starting with Ni_2O_5, he has been focusing on a broader spectrum of previously unknown nickel oxides as part of his doctoral studies.

FIGURE 2

New layered Ni_2O_5 structures predicted by evolutionary algorithms. Top figure - structure with straight layers. Bottom figure - structure with folded layers linked by O-O bridges. Small red balls represent oxygen atoms, big balls represent nickel atoms.

Ing. Diana Fabušová focused on DFT+EA calculations of PdO2 as a bachelor and master student at MTF STU within her BSc and master thesis and extracurricular activities. Currently she continues her research of new binary palladium oxides as a PhD student.

Palladium

Pd

46

Palladium

The situation in the binary palladium-oxygen system resembles that in the Ni-O system. Just like nickel, palladium has one well-characterized phase, palladium oxide (PdO), and another poorly characterized phase, palladium dioxide (PdO_2). Palladium dioxide was first obtained in high-pressure and high-temperature synthesis in 1978, and its crystal structure was proposed based on X-ray diffraction data. However, there has been no subsequent study to confirm or refute the proposed PdO_2 structure. Palladium and its oxides are crucial catalysts in many industries, with significant importance in the automotive industry. Therefore, deepening our understanding of palladium oxides and predicting new phases is highly desirable.

Ing. Diana Fabušová focused on quantum-mechanical modeling of palladium dioxide as part of her bachelor's and master's theses. In her bachelor's thesis, she studied the polymorphism of PdO_2 under normal conditions, while her master's thesis explored structural changes and phase transitions due to increased temperature and pressure. Thanks to her work, we obtained detailed information about the crystal structure proposed in the original experiment, and we refined the crystallographic system in which PdO2 crystallizes.

We discovered that, like its better-studied heavier counterpart PtO_2, PdO_2 can exist in various polymorphs, which was an expected result since palladium and platinum belong to the same Group 10 in the periodic table of elements. What came as a surprise was the prediction of a new porous structure (Figure 3), which is entirely unknown in the case of PtO_2. Our calculations suggest that this new porous PdO_2 structure is highly stable. This is a significant finding because porous structures represent attractive materials with a wide range of applications, including gas storage and separation, catalysis, energy storage and conversion, ion exchange, various types of sensors, including biosensors, and many other applications.

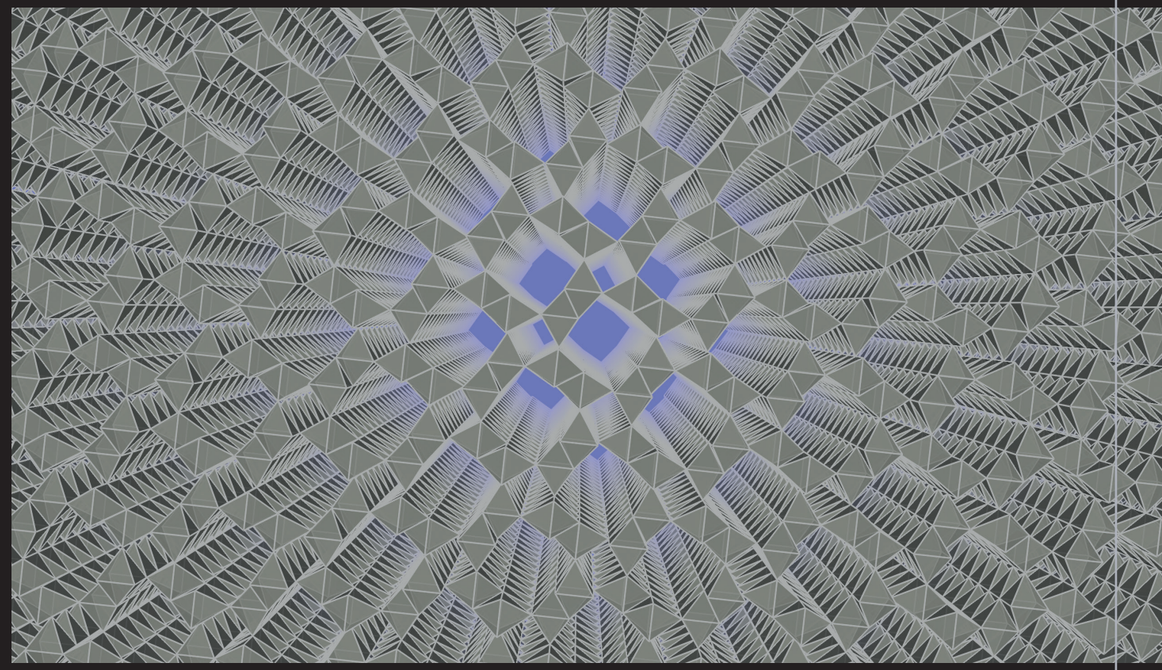


FIGURE 3

New porous PdO_2 structure predicted using Density Functional Theory calculations.

MODELLING & EXPERIMENTATION

Together with Dr. Pavol Noga (left) and Prof. Martin Kusý (right), we are training a new generation of young materials researchers in a joint approach in the search for new functional materials, which combines computer modelling with experimental techniques.

COMPUTATIONAL MODELLING & EXPERIMENT

Computational modelling at the atomic scale represents the most advanced approach to discovery of new materials. The mission of our research team and the Computational Materials Design laboratory is not only to conduct research but also to make computational modelling an integral part of the development of new materials at the Advanced Technologies Research Institute (ATRI), which is part of Faculty of Materials Science and Technology in Trnava of Slovak University of Technology in Bratislava. For this purpose, we closely collaborate with the ATRI's Department of Ion-beam technologies and the Institute of Materials MTF STU, where our colleagues work diligently to realize the materials we predict.

Under the guidance of Ing. Pavol Noga, PhD., the compounds we predict are being realized using physical and physico-chemical methods such as ion-beam synthesis and reactive magnetron sputtering, often in combination with plasma-assisted ion implantation. Characterization of the obtained materials is conducted in collaboration with the Institute of Materials. The primary method for detecting crystal structure is X-ray diffraction, overseen by Professor Martin Kusý. Together with Professor Kusý and Dr. Noga, we are training a new generation of young materials researchers in a combined approach to discovery of new functional materials that combine computational modelling with experimental techniques.

COMPUTING RESOURCES, COLLABORATIONS & OUTPUTS

Every result we have obtained and described above represents days, weeks, and even months of pure computational time. Our research would not have been possible without a well-functioning and powerful high-performance computing (HPC) infrastructure, as most of our simulation tasks would not have been feasible. We used the SIVVP infrastructure for parallelized computations, primarily utilizing the Aurel supercomputer at the Slovak Academy of Sciences in Bratislava and a cluster located in Košice. Additionally, we used the HPC infrastructure PRACE FENIX (project fnxpo70004 at TGCC) in France.

The results obtained through the Slovak supercomputing infrastructure have been presented at several international conferences [1-14]. They have also been the subject of one bachelor's thesis [15] and two master's theses [16,17]. Currently, they are the focus of two doctoral projects [18,19].

Our research would not be possible without a well-functioning and powerful supercomputing HPC infrastructure, as most of our simulation tasks would not be feasible.

REFERENCES

[1] D. Fabušová (študentka MTF STU), M. Derzsi, K. Tokár: *Design of atomistic models of little-known palladium oxide PdO_2*, The 2nd International Online Conference on Crystals MDPI, 2020.

[2] S. Sampath, K. Tokar, M. Derzsi: *Mapping of novel binary copper oxides with density functional theory modelling*, The 2nd International Online Conference on Crystals MDPI, 2020.

[3] S. Sampath, K. Tokár, M. Derzsi: *New binary copper oxide phases from ab initio*, YOUNG MULTIS 2021.

[4] S. Sampath, K. Tokár, M. Derzsi: *Exploring novel binary copper oxides using density functional theory and evolutionary algorithms,* ASC Spring 2021 Virtual.

[5] 5. M. Gašpárková, K. Tokár, M. Derzsi: *Crystal and electronic structure and thermodynamic stability of nickel sesquioxide Ni_2O_3 from ab initio*, EUROMAT 2021 ONLINE.

[6] D. Fabušová, M. Derzsi, K. Tokár: *Ab Initio Study of Impact of Pressure and Temperature on Stability and Polymorphism of PdO_2*, ES22 Online, Columbia University in the City of New York, 2022.

[7] R. Bujdák, M. Gašpárková, M. Derzsi, K. Tokár: *Exploring new phases in Ni-O binary system from ab initio*, ES22 Online, Columbia University in the City of New York, 2022.

[8] F. Ferenčík, J. Halanda, M. Kusý, T. Shen, D. Vaňa, M. Beňo, M. Derzsi, P. Noga: *Ion beam synthesis of high oxidation state palladium oxide nanoparticles*. ANPC 2021, Praha.

[9] F. Ferenčík, J. Halanda, D. Vaňa, M. Derzsi, P. Noga: *Ion beam synthesis of high oxidation state palladium and copper oxides*, IBMM 2022, Lisbon.

[10] D. Fabušová, K. Tokár, M. Derzsi, P. Piekarz, and P. T. Jochym: *Phase stability of PdO_2: The role of temperature and electron correlations*, Recent progress in ab initio phonon calculations, Cracow, 2023.

[11] K. Tokár, M. Derzsi: *Dynamical properties of isolated silver difluoride 1D structure under strain*, Recent progress in ab initio phonon calculations, Cracow, 2023.

[12] D. Fabušová, K. Tokár, M. Derzsi: *Study of PdO_2 polymorphism using Density Functional Theory (DFT) modelling*, 2nd Super school on physics of condensed matter (Original SK name: 2. letná škola fyziky kondenzovaných látok), Liptovský Ján, 2023.

[13] R. Bujdák, D. Fabušová, K. Tokár and M. Derzsi: *Ab initio study of novel Ni-O phase Ni_2O_5*, 2nd Super school on physics of condensed matter (Original SK name: 2. letná škola fyziky kondenzovaných látok), Liptovský Ján, 2023.

[14] R. Bujdák, D. Fabušová, K. Tokár and M. Derzsi: *First principles study of new nickel pentaoxide phase*, Open Science with Atomic Simulation Environment (workshop), Daresbury Laboratory, 2023.

[15] D. Fabušová, Bachelor's work (2020): *Design of atomistic models of little-known palladium oxide PdO_2* (Original SK title: *Návrh atomistických modelov málo známeho oxidu paládia PdO_2),* available online on https://opac.crzp.sk/.

[16] M. Gašpárková, Diploma thesis (2020): *Modelovanie kryštálovej štruktúry a termodynamickej stability nového oxidu niklu Ni_2O_3 (*Original SK title: *Modelovanie kryštálovej štruktúry a termodynamickej stability nového oxidu niklu Ni_2O_3),* available online on https://opac.crzp.sk/.

[17] D. Fabušová, Diploma thesis (2022): *Computational modelling of pressure induced phase transitions in PdO_2 (*Original SK title: *Modelovanie kryštálovej štruktúry a termodynamickej stability nového oxidu niklu Ni_2O_3),* available online on https://opac.crzp.sk/

[18] R. Bujdák: *Structural characterization of novel oxides of nickel and copper*, PhD project 2022-2025.

[19] D. Fabušová: *: New compounds with palladium for technological applications (*Original SK title: *Nové zlúčeniny s paládiom pre technologické aplikácie),* PhD project 2023-2026.

ACKNOWLEDGMENTS

The Slovak Research and Development Agency project
No. APVV-18-0168

"Quest for novel inorganic compounds with nickel, palladium,copper and silver by DFT modelling and ion beam synthesis"

Scientific Grant Agency of the Slovak Republic project VEGA
No. VG 1/0223/19

"Computational design of novel functional materials"

Operational Programme Research and Innovation for the project:
TMS2014+: 313011W085 co-financed by the European Regional Development Fund (ERDF)

"Scientific and Research Centre of Excellence SlovakION for Material and Interdisciplinary Research"

Computer modelling on the atomics cale represents the state-of-the-art approach to the discovery of new materials.

Use of HPC to understand changes in the diversity of fungal communities and monitor biological invasions

MIROSLAV CABOŇ
SLAVOMÍR ADAMČÍK

Dr. Miroslav Caboň and Dr. Slavomír Adamčík work as independent researchers at the Department biodiversity and ecology of the PSBC SAS Botanical Institute.

Fungal communities represent a dynamic environment whose enzymatic activities and interactions with plants significantly impact soil quality and the makeup of terrestrial plant communities. Estimating local fungal diversity is challenging due to the varied climatic needs of individual fungi for fruiting body production. Most fungi within these communities exist as underground hyphae.

In recent years, advanced molecular methods, notably second-generation sequencing, have enabled more precise estimations of fungal diversity than traditional methods involving prolonged monitoring and repeated site visits. This process, involving the isolation of fungal DNA directly from soil followed by sequencing, yields millions of short DNA segments. These segments can be specifically linked to various types of soil organisms. Through a technique known as metabarcoding, we gain a comprehensive view of soil community diversity, allowing us to detect shifts in diversity composition due to various external factors. However, processing such an extensive array of sequences swiftly necessitates substantial computing power, making access to a robust HPC (High-Performance Computing) infrastructure essential. Researchers from the Laboratory of Molecular Ecology and Mycology (MEM) are leading a national APVV project (**APVV 20-0257** *Tree and landscape – the influence of woody plants on the diversity of soil microorganisms in agricultural land*). This project uses metabarcoding to track changes in fungal communities influenced by biotic or abiotic factors surrounding solitary trees in landscapes. Our goal is to understand how solitary trees and their effects, such as rainfall, wind, and sun shade, influence the surrounding soil quality and local fungal communities. Given the need to process hundreds of soil samples, we anticipate continuing to utilize the HPC infrastructure provided by the SAS Computing Center.

In our Laboratory, the second research focus is on understanding the impact of plant invasions on native habitats and their direct effects on soil diversity. Biological invasions rank among the leading causes of habitat degradation and biodiversity loss. We define invasive organisms as artificially introduced species that adapt aggressively in new environments, expanding their territories and often outcompeting native species, thereby disrupting ecological balances. It's estimated that fewer than 10% of non-native species become invasive, yet invasive organisms span all kingdoms, from plants and fungi to animals. The restoration of habitats invaded by these organisms is typically challenging and costly. Consequently, biological invasions receive considerable attention, and various national and international lists of invasive and non-native organisms are

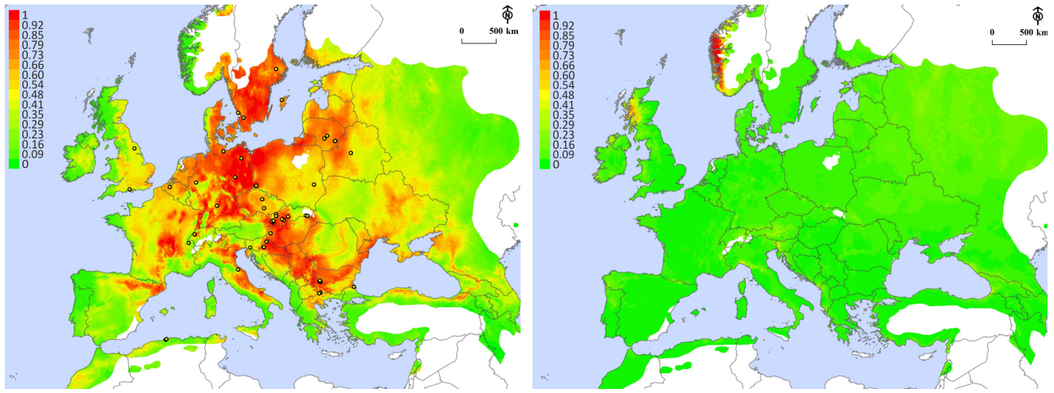
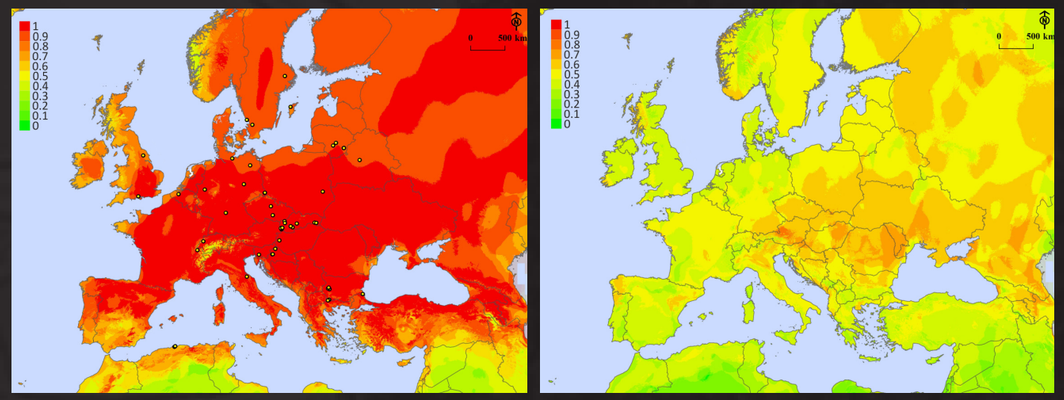


FIGURE 1

Models of the potential distribution of the European species *Phyllactinia fraxini* (left) and the Asian *P. fraxinicola* (right) in Europe in 2050, modeled under current climate scenarios. Published in Pastirčáková et al. 2021, (freely redistributable, CC BY 4.0 license)
Author of calculations: RNDr. Dušan Senko, PhD. (Department of Evolution and Systematics of the PSBC SAS Botanical Institute).

regularly updated. This topic ties directly into our ongoing project, **APVV 19-0510** *„Invaders Among Us: Time-space Dynamics of Plant Invasions and Their Adverse Impact on Ecosystems,"* led by R. Hrivnák. Our goal is to identify the fungal component in the soil and understand their responses to environmental changes.

Another area where we actively utilize HPC infrastructure is in modeling the climatic demands of species with invasive potential and identifying changes in diversity in invaded areas. This modeling work relates to the completed project **APVV 15-0210** *„Distribution Potential of Different Trophic Groups of Fungi in Europe,"* which focused on phytopathogenic fungi known as powdery mildews. These fungi are biologically specialized to inhabit only certain families, genera, or species of host plants. While some mildew types proliferate across their host plant's range, others are restricted to specific climatic regions. Most species are geographically limited to individual continents or large areas, with only a few being globally distributed. The highest diversity of mildews is found in temperate and humid climates, with their occurrence and species spectrum decreasing significantly towards the equator and the poles. Although knowledge about some economically significant mildew species is rapidly expanding, little is known about their species diversity over large geographical areas. Our project focused on mealybugs of the genus *Phyllactinia* colonizing ash trees, found in both Europe and Asia. Literature acknowledges two species: the European *Phyllactinia fraxini*, present on all three European ash species, and the Asian *Phyllactinia fraxinicola*, attacking a wide range of ash trees in



Southeast Asia. Our research aimed to determine whether these species differ in their climatic requirements and whether the introduction of non-native ash species poses a risk for their spread into new areas.

In Europe, our research confirmed the presence of *Phyllactinia fraxini* on all three native European ash species as well as on two Asian species planted in the arboretum (*Fraxinus. chinensis* ssp. *Rhynchophylla* a *Fraxinus mandshurica*). Our modeling of the potential distribution for both *Phyllactinia* species indicates that their habitats do not overlap in Europe. Additionally, our data suggest no preference of either species for urban or native habitats, nor for a specific ash species. The Asian powdery mildew, *P. fraxinicola*, appears unsuited to the conditions in most parts of Europe and currently does not pose a significant threat. However, some areas in the northwest of the continent might be vulnerable to invasion. Therefore, non-native ash trees in these regions should be more rigorously monitored for signs of powdery mildew to prevent this species from gaining a foothold in Europe.

Climatic projections up to 2050 indicate an expansion in areas suitable for the distribution of both species, with the distribution center of the Asian species shifting towards Central and Southeastern Europe. For the European species, suitable climatic areas are expected to grow substantially, covering most of Europe. As a result, it is prudent to pay closer attention to this species in the future and consider its potential impact in planting strategies.

FIGURE 2

Prediction of the potential distribution of the European species *Phyllactinia fraxini* (left) and the Asian *P. fraxinicola* (right)) in Europe. Published in Pastirčáková et al. 2021, (freely redistributable, CC BY 4.0 license).
Author of calculations: RNDr. Dušan Senko, PhD. (Department of Evolution and Systematics of the PSBC SAS Botanical Institute).

Transfer and optimization of CFD calculations workflow in HPC environment

Ján Škoviera[1]
Sylvain Suzan[2]

[1]National Competence Centre for HPC, Slovak National Supercomputing Centre
[2]Shark Aero

Shark Aero specializes in designing and manufacturing ultralight sport aircraft, featuring two-seater tandem cockpits. In their design process, they utilize the popular open-source software package, OpenFOAM [1], with a focus on Computational Fluid Dynamics (CFD) simulations employing the Finite Element Method (FEM). The initial step involves creating a model using Computer-Aided Design (CAD) software, which is then segmented into discrete cells, forming what is known as a mesh. The precision of these simulations is closely tied to the mesh density, with both computational and memory requirements increasing exponentially, specifically to the third power of the number of vertices. In certain cases, the computational demands of these simulations can become a significant bottleneck, particularly when using standard computing technology. To address this challenge, we have explored transferring the simulation workflow to a High-Performance Computing (HPC) environment. Our primary focus has been on examining the effectiveness of parallelizing computational tasks for specific model types in this advanced computational setting.

In certain cases, the computational demands of these simulations can become a significant bottleneck, particularly when using standard computing technology.

Compute nodes with 2x6 cores Intel Xeon L5640 @ 2,27GHz, 48 GB RAM and 2x500 GB were used for this project. All calculations were done in a standard HPC environment using Slurm job scheduling system. This is an acceptable solution for this type of workloads where no real-time response, nor

immediate data processing is required. For the CFD simulations we continued to use OpenFOAM & ParaView version 9 software packages. Singularity container was used for calculation deployment, having in mind potential transfer of the workload to another HPC system. The speed-up gained from just

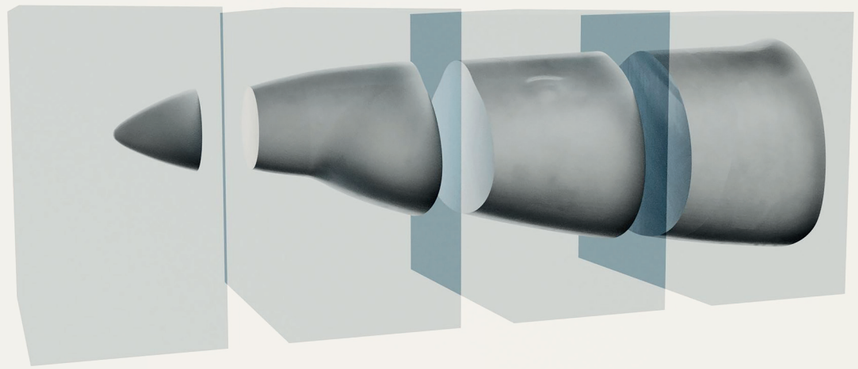


FIGURE 1

Illustration of mesh segmentation. The encoling mesh is represented by the transparent boxes.

straight away transfer to HPC system was approximately 1.5x compared to a standard laptop.

Parallelized task execution can increase the speed of the overall calculation by utilizing more computing units concurrently. In order to parallelize the task one needs to divide the original mesh into domains - parts that will be processed concurrently. The domains, however, need to communicate through the processor boundaries i.e. domain sides where the original enclosing mesh was divided. The larger the processor boundary surface is, the more I/O is required in order to resolve the boundary conditions. Processor boundary communication is facilitated by the distributed memory Message Passing Interface (MPI) protocol, and the distinction of difference between CPU cores and different compute nodes is abstracted from user. This leads to certain limitations on efficient usage of many parallel processes, since overly parallelized job executions can be actually slower due to communication and I/O bottlenecks. Therefore, the domains should be created in a way that minimizes the processor boundaries. One possible strategy is to divide the original mesh only in co-planar direction with the smallest side of the original enclosing mesh. By careless division into domains the amount of data to be transferred increases beyond reasonable measure. If one chooses to use mesh division in multiple axes, one also creates more processor boundaries.

Parallelized task execution can increase the speed of the overall calculation by utilizing more computing units concurrently.

The calculations were done in four steps: enclosing mesh creation, mesh segmentation, model inclusion and CFD simulation. The enclosing mesh creation was done using the blockMesh utility, the mesh segmentation step was done using the decomposePar utility, the model inclusion was done using the snappyHexMesh program, and the CFD simulation itself was done using SimpleFoam. The most computationally demanding step is snappyHexMesh. This is understandable from the fact that while in CFD simulation the calculation needs to be done several times for every edge of the mesh and every iteration, in the case of model inclusion one creates new vertices and deletes old ones based on the position of vertices in the model mesh. This requires creation of an “octree” (partitioning of three-dimensional space by recursively subdividing it into eight octants), repeated inverse search, and octree re-balancing. Each of these processes is N*log(N) in the best case scenario, and N2 in the worst case, N being the number of vertices. The CFD itself scales linearly with number of edges, i.e. “close to” linearly with N (only spatially proximate nodes are interconnected).2 We developed a workflow that creates a number of domains that can be directly parallelized with the yz plane (x be-

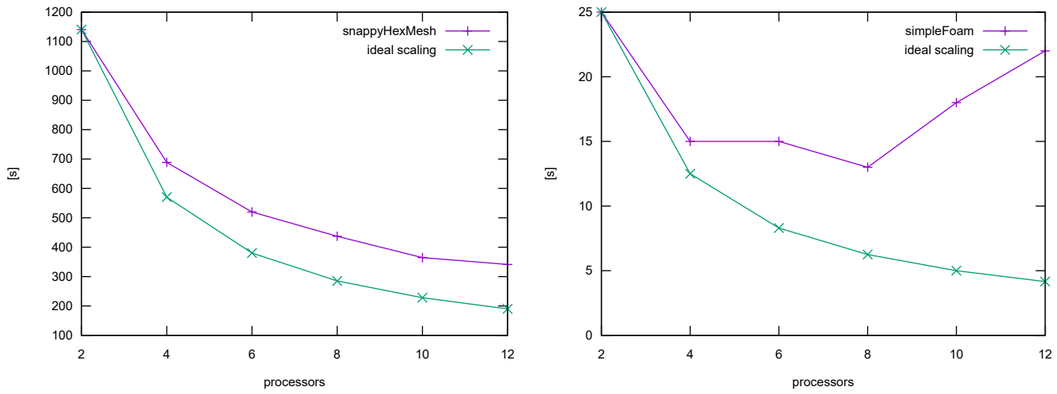


FIGURE 2

Dependence of real elapsed time on the number of processes for snappyHexMesh and simpleFoam. In the case of simpleFoam the time starts to diverge for more than 8 processes, since the data trafic overcomes the paralellization advantage. Ideal scaling shows the theoretical time needed to finish the calculation, if the data trafic and processor boundary condition resolution was not involved.

ing the axis of the aircraft nose), which simplifies the decision making. After inclusion of a new model, one can simply specify the number of domains and run the calculation minimizing the human intervention needed to parallelize the calculation.

The relative speedup of the processes calculation is mainly determined by limited I/O. If the computational tasks are well below I/O bounding, the speed is inversely proportional to the number of domains. In less demanding calculations, i.e. for small models, the processes can be easily over-parallelized. Once the mesh density is high enough, the time to calculate the CFD step is also inversely proportional to the number of parallel processes. As shown in the second pair of figures with twofold increase in mesh density, the calculations are below I/O bounding even in the CFD step. Even though the CFD step is in this case comparatively fast to the meshing process, the calculation of long time intervals could make it the most time consuming step.

The aircraft parts design requires simulations of a relatively small models multiple times under altering conditions. The mesh density needed for these simulations falls into medium category. When transferring the calculations to the HPC environment, we had to take into account the real needs of the end user in terms of model size, mesh density and result precision required. There are several advantages of using HPC:

- The end user is relieved of the need to maintain his own computational capacities.
- Even when restricted to single thread jobs the simulations can be offloaded to HPC with high speed up, making even very demanding and precise calculations feasible.
- For even more effective calculations a simple way of utilizing parallelization was determined, for this particular workload. Limitations of parallel runs for the given use case and conditions were identified. The total increase in speed that was reached in practical conditions is 7.3 times. The speed-up generally grows with the calculation complexity and the mesh precision.

03

HPC Popularization

hpc focus

Short news

EuroCC 2

In January of this year, the EuroCC 2 project started, as part of which we implement the activities of the National Competence Centre for HPC. The host institution is the National Supercomputer Centre. The project is mainly focused on the development of competences and services in high performance reading and supporting small and medium enterprises, academic workplaces and public reports in the form of cooperation in the development of solutions based on HPC technologies.

National Competence Centre for HPC on Europe Day

On May 9, the „birthday celebration" of the European Union took place in Bratislava's Stara tržnica. The Europe Day event was organized by the Office of the European Parliament and the Representation of the European Commission. The team of the National Competence Centre for High-Performance Computing also presented its activities on Europe Day. On Europe Day, we commemorate the Schuman Declaration which started modern European integration and led to the creation of the current European Union as a unique peaceful and democratic project.

Information stands of EU member states, the European Parliament and the European Commission were available to the general public. The event also featured the stand of the National Competence Centre for High-Performance Computing (NCC for HPC). The NCC team presented the activities and services of the National Competence Centre for HPC. Through an interactive demonstration of simulations of potential drugs against the Covid-19 disease, we brought the participants closer to what calculations supercomputers can be used for. During the event, Lucia Demovičová spoke on stage and presented the National Competence Centre for HPC to the public. She explained how the competence centre can help scientists, researchers, and also small and medium-sized entrepreneurs. She also presented successful examples of cooperation with the private and public sector.

eurocc.nscc.sk

TREX CoE & NCCs collab
Code tuning for the exascale

TREX centre of excellence together with 3 national competence centres delivered a three-day workshop „Code tuning for the exascale". NCC Slovakia, Austria and Czech Republic teamed up to bring an interesting programme including advanced parallel programming, energy efficiency analysis and HPC application optimization. The workshop took place in Bratislava, Slovakia during June 5 – 7, 2023 and was organized in a face-to-face format.

This was a nice example of collaboration and support between the NCCs and a CoE. Three complementary topics were covered by the agenda. During the first day the team from Austrian NCC presented advanced parallel programming techniques (MPI, OpenMP), followed by the presentation of a trending topic of energy efficiency in HPC by the Czech NCC on the second day. The third day belonged to the TREX CoE, which presented a tool for HPC applications performance analysis and optimization called MAQAO. Organization, access to HPC and technical HPC support was given by the Slovak NCC.

TREX
Targeting Real chemical accuracy at the EXascale

Central European NCC working group meeting

On June 12, representatives from national competence centres for HPC situated in the central Europe region met during the first Central European NCC working group meeting. The event was organized by NCC Slovenia and NCC Austria in Maribor, Slovenia and online.

Three sessions were on the agenda, covering the topics of interaction with industry, training and communication. Participants worked in small groups, discussing specific points and sharing best practices within the topics, such as SME approach strategies, service portfolio and how to deal with the state aid issues, rules of engagement with a private company.

Training event formats were discussed, considering post-pandemic trends; experiences with NCC-CoE collaborations were highlighted and different strategies of attracting SME participants to training courses were shared.

The communication session focused on promotion and communication channels, build-up of an audience for different topics and organization of theme-specific webinar series. In each section, opportunities and invitations for collaboration were discussed, as well. The event was a great opportunity for getting to know colleagues from neighboring NCCs, informal networking, fostering collaboration and finding new inspiration.

hpc focus

2023

DIRECTOR OF CC SAS
Michal Kadúc

RESPONSIBLE REDACTOR
Lucia Demovičová

GRAPHIC DESIGN & DTP
Gabriela Obadalová

PHOTOS
Pavol Novák
Shutterstock

PERIODICITY
Raz ročne

PUBLISHER
Centre of Operations of SAS
Computing Centre of SAS

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

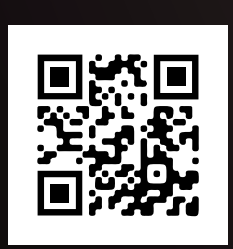
www.vs.sav.sk

ISBN 978-80-89871-20-9
ISSN 2729-9090

Centrum spoločných činností SAV, v. v. i.
Výpočtové stredisko SAV

Dúbravská cesta 9
845 35 Bratislava
Slovenská republika

www.vs.sav.sk



ISBN 978-80-89871-20-9
ISSN 2729-9090