HPC ECHO

From science to innovation



ISBN: 978-6-15019-180-5

2024

Content

A new dimension in the clinical significance of mutations: Alphafold based genetic variant annotation	3
Axle load estimation based on data provided by strain gauges installed on a bridgetext	6
The Super Team: Supernovae Near Supermassive Black Holes	8
Synthetic language training in an HPC environment	10
Parallel multi-GPU algorithm research for brain EEG data processing and analysis	12
How can weather forecasts be more accurate?	14
Simulation of urban air pollution, development of a general-purpose HPC portal and fluid dynamics solvers	16
Advancing Microbiome Engineering with High-Performance Computing	18
The Magnetic Exchange — MagEx project	20

Foreword

Krisztián Kongó

Vice-President



For the fourth time, we are launching the HPC ECHO "From Science to Innovation", where publication with the support of the HPC Competence Centre, which will inform you how the recently delivered Komondor supercomputer (HPC) is serving research in almost all scientific disciplines. Already during its trial phase, Komondor contributed to numerous scientific works and developments, and its continuous operation now offers significant research and innovation opportunities, even on a global scale, for Hungarian higher education, large industry and the growing number of small and medium-sized enterprises. The current edition covers a wide range of topics, from supernova observation to microbiome design and supercomputer-based medical research. I hope that many of you will decide to join the HPC user community and take advantage of one of the most powerful supercomputers in Central Europe, which is ready to serve professionals with its 6-petaflop performance, comprehensive software support, and full user assistance.

HPC Competence Centre

HPC @hu
Competence Centre

A new dimension in the clinical significance of mutations: Alphafold based genetic variant annotation

Genetic variants are alterations in the DNA sequence that can impact protein function through various mechanisms. Some of these variants are benign and do not affect significant biological processes, while others can cause changes in the protein structure, splicing, or expression and can lead to cancer development. Over time, these mutations accumulate, contributing to the complexity and uncontrolled growth of cancer cells. Understanding the impact of these variants on gene function and their phenotypic consequences is crucial for comprehending how they influence cancer development.[1]

The development of next-generation sequencing techniques has greatly facilitated genetic screening for both cancer patients and healthy individuals.[2] These techniques have rapidly become an integral part of clinical genetics, enabling the identification of hundreds and thousands of missense variants. However, this extensive screening has also led to the emergence of a significant issue: for a considerable number of these genetic variants, we lack information on how they affect protein function and, ultimately, patients' risk of developing cancer.[3] These variants, classified as variants of unknown significance (VUS), pose a major challenge for physicians when identified, causing significant stress, particularly for the patient.

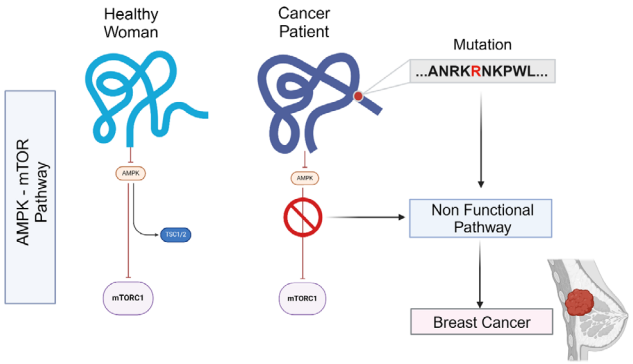


Figure 1: *illustrates the impact of a pathogenic mutation on the normal functioning of a hypothetical biological pathway. The diagram compares the pathway's operation under normal conditions with its altered state due to the mutation, highlighting key points where the mutation disrupts normal processes and leads to functional impairment.*



HUN-REN Biological Research Centre, The Mutagenesis and Carcinogenesis Research Group
Delta Bio 2000 Kft
European Life Technologies Zrt.

The Mutagenesis and Carcinogenesis Research Group at the Biological Research Centre in Szeged, led by Lajos Haracska, has been studying the mutational events leading to cancer development for over two decades. The research group employs the most advanced biochemical methods, complemented by genomic and transcriptomic studies. Their models are validated using human tissue cultures and microscopic techniques, ensuring the reliability and accuracy of their research results.

Delta Bio 2000 Ltd. specializes in developing methods and services necessary for the molecular genetic analysis of tumors. The company faces the challenges posed by somatic variants with unknown outcomes on a daily basis and continuously works on solving them to facilitate accurate diagnosis and effective treatment.

European Life Technologies Plc. focuses primarily on molecular biological research and professional IT solutions related to the monitoring and screening of cancer diseases through liquid biopsy. They also play a significant role in storing the large amounts of data generated during the project and making it accessible to the community, ensuring the wide accessibility and usability of the research results.

To aid in the classification of these VUS, machine learning algorithms have been developed and utilized over the past decade.[4] These methods have demonstrated sufficient effectiveness to be incorporated into the ACMG/AMP guidelines for variant interpretation (Richards et al., 2015); nevertheless, they possess inherent limitations that affect their utility in clinical settings (Sun and Yu, 2019; Liu et al., 2022), and many variants remain inconclusive. This highlights the need for continued efforts in developing more robust algorithms or enhancing their predictive capabilities. Protein functions are largely dependent on the tertiary structure, which is defined by the amino acid sequence; therefore, if we can predict how a change in the amino acid sequence affects the tertiary structure, we may have a chance to predict how it affects the function. However, these amino acid changes could affect protein structure in a plethora of ways; thus, without a computational model

it is impossible to give a prediction that is accurate enough to be used for diagnostic purposes. Until now, the development of novel prediction tools has been fundamentally limited by the lack of three-dimensional structures for many of the proteins in the clinical genome.

A significant breakthrough came with the development of the artificial intelligence protein structure prediction algorithm AlphaFold2, which now enables the accurate prediction of the three-dimensional structure of virtually any protein.[7] This resulted in the generation of a wealth of novel structure-derived features, which have been proven highly promising and are now being integrated into the training of the next generation of missense variant classification algorithms.[8] Nonetheless, due to the absence of mutated protein structures, these efforts have been restricted to the information harbored by the native structure, potentially omitting essential information that could improve the classification performance of these tools.

To resolve this issue, we addressed the computationally massive challenge of obtaining the 3D structures of approximately 70,000 pathogenic and benign missense variants sourced from Humsavar, a well-curated human mutational database. The prediction of this large number of protein structures demands substantial computational resources that only a high-performance computing (HPC) cluster can offer.

Initially, we employed AlphaFold2 to predict both the native and the mutant structures on a large scale. This process involves two main steps: first, sequence database querying to construct the multiple sequence alignment of the target protein, which demands substantial CPU resources; second, the prediction of the protein structure using the deep learning model, which requires powerful graphical processing units (GPUs). Given the relatively long protein structures in our dataset, we quickly observed that AlphaFold2 struggled with a CPU bottleneck thus increasing the overall time required to generate the protein structures. To address this issue, we switched to ParaFold[9], an HPC-optimized implementation of AlphaFold2, for large-scale protein structure prediction, as it separates the above-mentioned steps into two distinct processes that can be individually submitted to specialized partitions. Using ParaFold, over a period of 14 months we have generated a comprehensive set of 68,000 protein structures encompassing 11,612 wild-type structures and 57,026 mutated ones. The dataset amassed a total size of 50 terabytes.

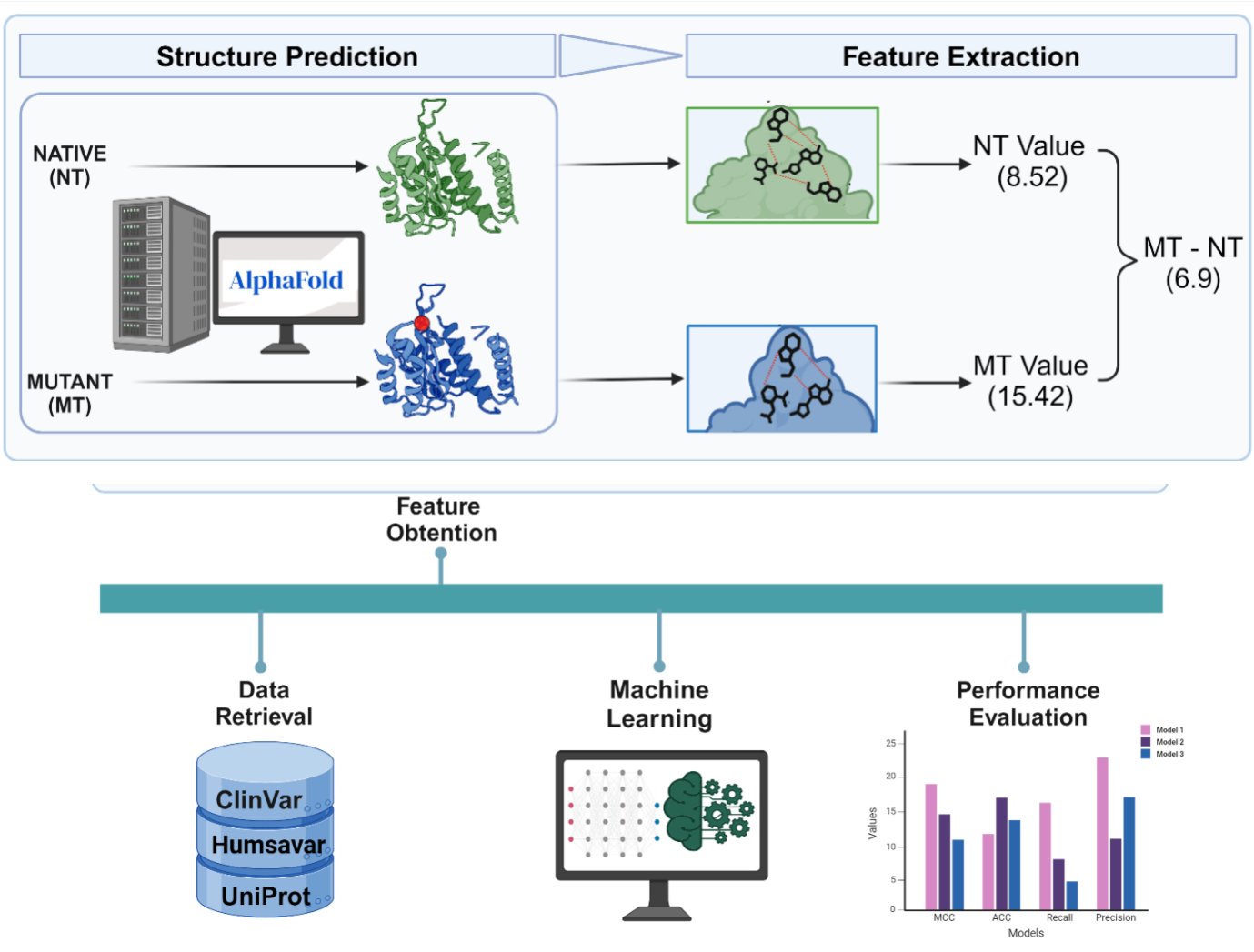


Figure 2: *outlines the workflow for developing a missense variant effect prediction algorithm, beginning with the data retrieval step and concluding with a thorough performance evaluation. Emphasis is placed on the structure prediction and feature extraction processes.*

Our academic group is interested in the process of carcinogenesis and has a close collaboration with Delta Bio 2000 Ltd. to aid the development of applied technologies in molecular cancer diagnostics. To advance our final goal of developing a machine learning and protein structure-based functional predictor software, we started a collaboration with the group of Prof. de la Cruz, at the Vall d'Hebron Research Institute, Barcelona, which is comprised of experts in machine learning-based variant classification. Within the frame of the collaboration, we have devised a comprehensive set of geometric and energy-based features to capture the differences between the native and mutated protein structures. Among the geometric-based features, we initially implemented an approach established by Weissenow et al., 2024, which utilized distance root mean square deviation (dRMSD) to quantify the structural distortion between native and mutant structures. This method was further enhanced by focusing more closely on the specific area of the mutation.

Furthermore, we developed a feature that measures the relative accessibility value of the native structure and utilizes this information to extract the atomic coordinates of the mutant. For the energy-based features, we implemented a coarse-grained version of the Miyazawa-Jernigan Potential, which is relatively straightforward to compute and provides insights into how protein stability changes.

We applied these features, along with a carefully selected set of sequence-based features, to train an XGBoost-based machine learning model. This model was tested on the protein Ornithine transcarbamylase (OTC, P00480) as a prototype to evaluate our features. The performance of the developed model was assessed using Matthew's Correlation Coefficient (MCC), accuracy, recall, and precision.

Our initial results indicate that the inclusion of structural-based features leads to a 20.69% increase in MCC and a 7.59% increase in accuracy. These findings suggest that incorporating these features can significantly enhance the performance of prediction algorithms. Therefore, our goal is to leverage these features to train machine learning models and investigate their impact on the algorithm's overall efficiency and performance.

1. American Association for Cancer Research (AACR). (n.d.). Understanding cancer development. In AACR Cancer Progress Report. Retrieved July 13, 2024, from https://cancerprogressreport.aacr.org/disparities/chd20-contents/chd20-understanding-cancer-development/#:~:text=Alterations%20in%20the%20DNA%20sequence,arise%20in%20the%20daughter%20cells.

2. Vears DF, Sénécal K, Borry P. Reporting practices for variants of uncertain significance from next generation sequencing technologies. Eur J Med Genet. 2017;60(10). doi:10.1016/j.ejmg.2017.07.016

3. Hoffman-Andrews L. The known unknown: The challenges of genetic variants of uncertain significance in clinical practice. J Law Biosci. 2017;4(3). doi:10.1093/jlb/lsx038

4. Ioannidis NM, Rothstein JH, Pejaver V, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am J Hum Genet. 2016;99(4):877-885. doi:10.1016/j.ajhg.2016.08.016

5. Qorri E, Takács B, Gráf A, et al. A Comprehensive Evaluation of the Performance of Prediction Algorithms on Clinically Relevant Missense Variants. Int J Mol Sci. 2022;23(14). doi:10.3390/ijms23147946

6. Gunning AC, Fryer V, Fasham J, et al. Assessing performance of pathogenicity predictors using clinically relevant variant datasets. J Med Genet. 2021;58(8):547-555. doi:10.1136/jmedgenet-2020-107003

7. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873). doi:10.1038/s41586-021-03819-2

8. Özkan S, Padilla N, de la Cruz X. QAFI: A Novel Method for Quantitative Estimation of Missense Variant Impact Using Protein-Specific Predictors and Ensemble Learning, 07 May 2024, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-4348948/v1]

9. Zhong B, Su X, Wen M, Zuo S, Hong L, Lin J. ParaFold: Paralleling AlphaFold for Large-Scale Predictions. In: ACM International Conference Proceeding Series. ; 2022. doi:10.1145/3503470.3503471

Axle load estimation based on data provided by strain gauges installed on a bridge



Bence Szinyéri is a PhD student at the Department of Automation and Applied Informatics of BME (Budapest University of Technology and Economics). His research topic is the application of deep learning algorithms.

Bence Kővári is an associate professor at the Department of Automation and Applied Informatics of BME. His main research areas are the application of artificial intelligence algorithms in the fields of automated signature verification, image processing and signal processing.

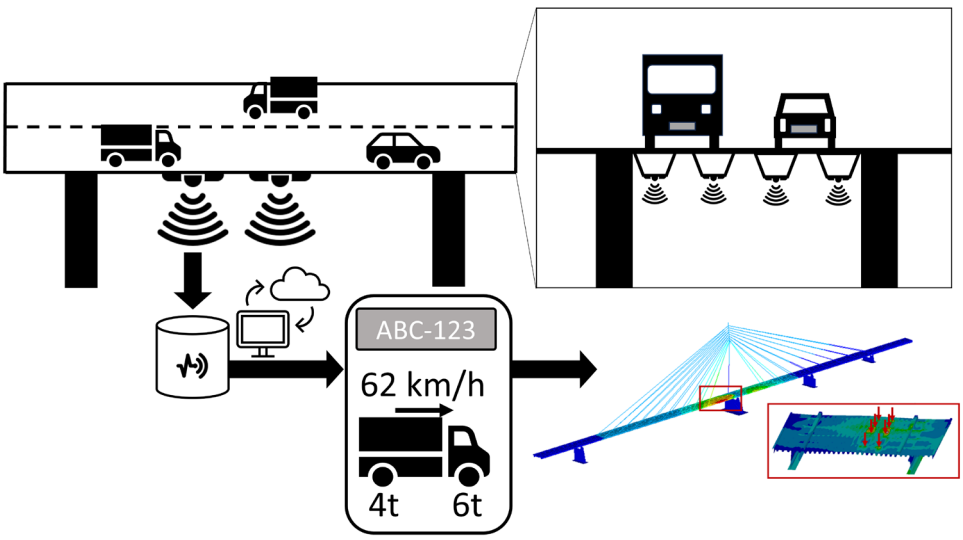
István Völgyi did his PhD at the Faculty of Civil Engineering of BME. He is currently an associate professor at the Department of Structural Engineering, head of laboratory development of the Structural Testing Laboratory. His research areas are reinforced concrete beam structures and structural diagnostics based on long-term support structure measurements.

Dénes Kollár is an assistant professor at the Department of Structural Engineering of BME. His main research areas are numerical simulation of welding processes and structural analysis based on numerical methods.

Attila László Joó is an associate professor at the Department of Structural Engineering of BME. Previously, he was the vice-dean of finance at the BME Faculty of Civil Engineering (ÉMK) and the director of the BME Centre for Innovation Management and Cooperation (FIEK). His main research areas are thin-walled steel structures, scaling for earthquakes, as well as digitalization in the construction industry including research of digital twins of supporting structures.

Monitoring the condition of a bridge is an important task for the engineers responsible for its operation. If we could track the various impacts on bridges, it would help in scheduling maintenance tasks and provide real-time information about the condition of the bridges. Our research group's goal is to create such a system, known as a digital twin. Among the impacts affecting the bridge, one of the most significant is the effects induced by traffic passing over the bridge. Heavy and overloaded vehicles and their axles play a major role in the fatigue of the bridge. Therefore, a subtask of creating the digital twin is to develop an axle load estimating system. We aim to create an axle load estimating system, also known as a Bridge Weigh-in-Motion (BWIM) system, based on the data provided by the strain gauges installed on the bridge. Such a system offers both short-term and long-term benefits. In the short-term, it can filter out overloaded vehicles, while in the long-term, it can provide data-supported reports and alerts about the condition of the bridge.

The main advantage of Bridge Weigh-in-Motion (BWIM) systems over axle load measurement systems embedded in asphalt is that the strain gauges installed on the bridge are placed under the bridge deck. It makes the installation simple, does not involve disrupting the road surface, allows for easy replacement in case of malfunction, and extends the expected lifespan of the system.



Processes Running in the Bridge Weigh-in-Motion System

The processes running within the Bridge Weigh-in-Motion system are outlined schematically in the attached diagram. The diagram clearly shows that strain gauges are typically installed at two cross-sections under the bridge deck. The signals provided by these strain gauges are stored in a database of the monitoring system. The collected data is processed by a software that outputs the axle loads of the vehicles, as

well as other optional parameters like speed or license plate number. Finally, the extracted data are integrated into the bridge's digital model, its digital twin.

Our research group formed by researchers from the Faculty of Civil Engineering and the Faculty of Electrical Engineering and Informatics at the Budapest University of Technology and Economics (BME) has developed a prototype of such an axle load estimation system on the Monostori Bridge in Komárom. The solution is based on a simulation process. If a vehicle with known parameters and a known trajectory crosses the bridge, the strain values measured in the bridge can be calculated using a Finite Element Method (FEM) based simulation. We need to solve the inverse of this task, that is, calculate the original axle loads from the strain data measured on the bridge. It must be done even when vehicles are traveling in multiple lanes. This is where artificial intelligence is taken into account. A deep learning-based algorithm is trained to solve this task. These training processes require computational capacities where the KIFÜ HPC infrastructure is used to accelerate training.

Field measurements have proven that the current system is capable of achieving the class A (best class) according to the COST 323 standard, which is widely used in the field of axle load estimations, based on errors in gross vehicle weight (GVW) and axle load estimation.

Future work should include integrating vehicle traffic data with the digital twin of the bridge. By knowing the vehicle traffic and performing the appropriate calculations, the condition of the bridge can then be monitored in real-time through its digital twin. Another goal is to generalize the solution so that the system can be easily installed on multiple types of bridges.

The research is carried out within the framework of the Artificial Intelligence National Laboratory (RRF-2.3.1-21-2022-00004) project supported by the European Union.

The Super Team: Supernovae Near Supermassive Black Holes



Dr. Barnabás Barna is a scientific associate at the University of Szeged. His primary research area is the origin of thermonuclear supernovae and their distance measurement. During his stay in Prague, he also investigated the long-term effects of supernovae on their environment, for the hydrodynamic simulations of which he first used Czech and then Hungarian supercomputers.

Supernovae are the most energetic events in the Universe. Supermassive black holes (SMBHs) are objects with the strongest gravitational fields in the Universe. What happens when these unstoppable forces encounter? Our research project, which began in 2020 during my first postdoctoral year at the Astronomical Institute of the Czech Academy of Sciences, aims to answer this question. SMBHs are found in the center of every large galaxy, including the Milky Way: "our" monster black hole called Sgr A* has a mass four million times that of our Sun – and yet it's still a lightweight in its own category! Since the discovery of SMBHs, it has been an open question how they could have grown so large, as current observations show relatively slow mass growth of black holes at the centers of galaxies.

One hypothesis of our team is that stellar explosions, called supernovae, may play a significant role in the development of SMBHs. In the central region of galaxies, not only the density of interstellar gas and dust is high, but there are also many massive stars , which leads to relatively frequent supernova explosions. The enormous kinetic energy released by supernovae sweeps up interstellar gases, and as it slows down, it can deliver them „in one package" to the gravitational well, i.e., directly near the black hole. In this way, the differences in mass growth among certain SMBHs could be explained, as well as the fact that Sgr A* currently does not show any significant activities.

Hundreds of thousands of years may pass since the supernova explosion near the center of the galaxy to the spiraling of the swept-up cloud of matter into the black hole, thus, the process cannot be studied through direct observations. Therefore, with my colleagues, we have developed hydrodynamic simulations that realistically model the physical processes occurring in the environment of SMBHs, but contain sufficiently simple approximations to be run according to dozens of different scenarios. Even so, highly detailed simulations (a 30x30x30 light-year volume, with a 0.1 light-year resolution) can only be managed in a supercomputer environment like the Komondor HPC.

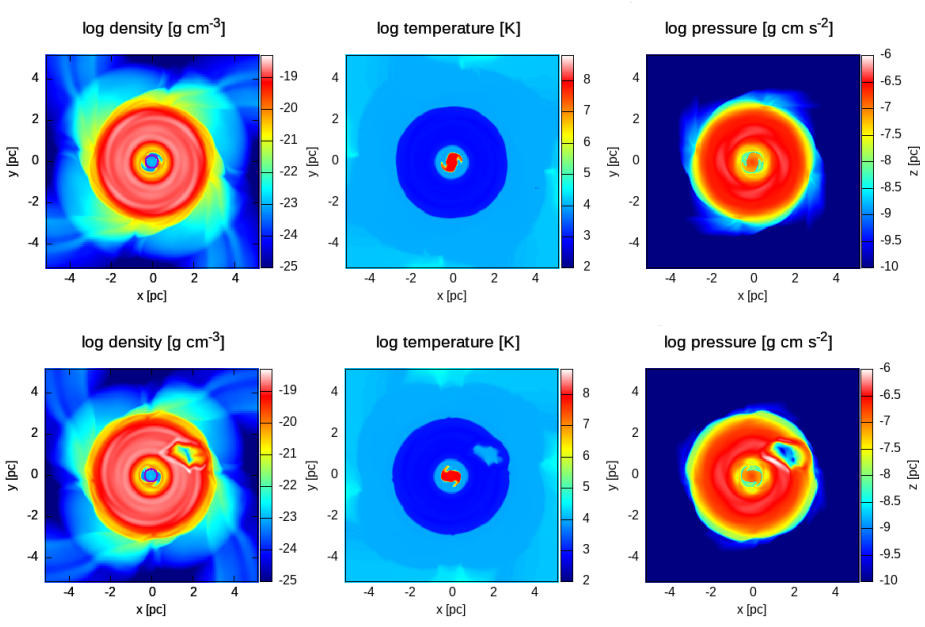


Figure 1: *Physical properties of the circumnuclear disk around the SMBH in the orbital plane before the explosion (top row) and 10 000 years after the explosion (bottom row). The expansion of the supernova remnant compresses the disk gas in all directions. Source: Barna et al. In prep.*

During our previous research, we showed that a single supernova can transport thousands of solar masses of gas to the central black hole assuming the dynamic properties of the Galactic center and a nearly homogeneous density environment. Later, by taking different

compositions of central star clusters, we added strong outflows called stellar winds to the models, which almost completely prevented the inward transfer of supernova remnants eventually. In our currently studied models, we have further improved our hydrodynamic simulations with the dense gas structure of several light-years in diameter around the SMBH, the so-called circumnuclear disk, which, despite the stellar wind from the center, is assumed to be able to maintain the spiraling of interstellar gas toward the SMBH, and in the case of a supernova, this process may even be significantly amplified. In the end, it seems a lot depends on where the supernova appears in the disk of matter. Some explosion sites particularly hinder the flow of matter towards the SMBH (as the explosion „blows away" a significant portion of the matter disk), while in other cases it can amplify it (by pushing several hundred solar masses of gas towards the center). Which is the dominant process will only be estimated based on statistics from multiple simulations – these calculations are running on Komondor as I write these lines. However, it seems more and more certain that stellar explosions play a significant role in the development of SMBHs, and the increased activity around black holes is likely due to the indirect effects of stellar explosions.

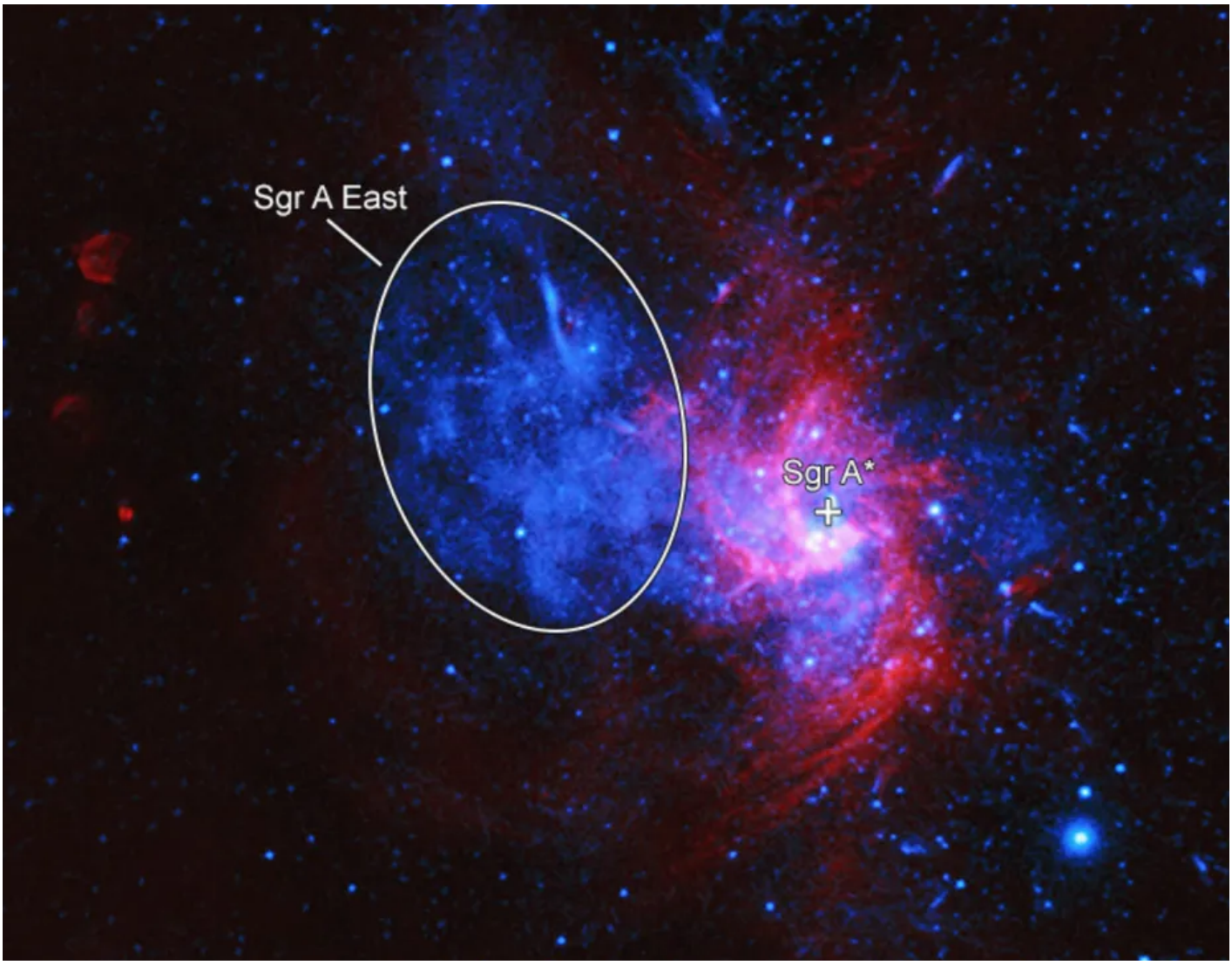
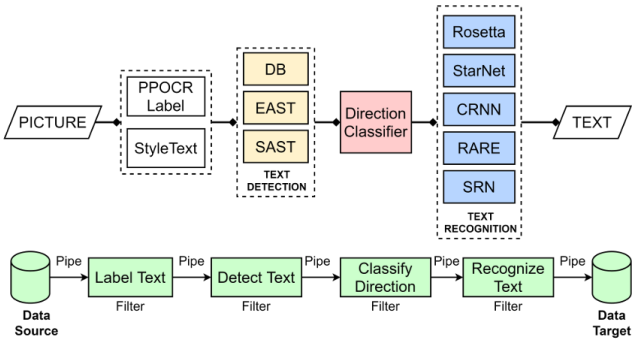


Figure 2: *The Sagittarius A East supernova remnant within the vicinity of the supermassive black hole Sagittarius A*, at the center of the Milky Way. Source: NASA / CXC / Nanjing University / Zhou et al. / NSF / NRAO / VLA*

Synthetic language training in an HPC environment

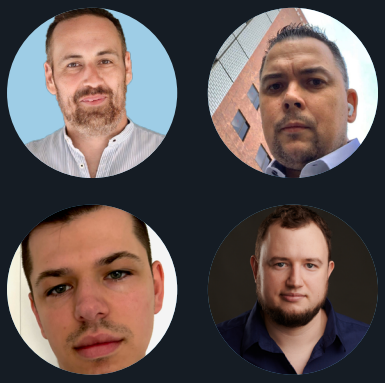
During the COVID-19 epidemic, ITware began developing a Japanese, English, and Hungarian hybrid language model with a Japanese partner. The goal was to establish a combined language to support our daily software development and other engineering fields via specific OCR-supported applications. Apart from accuracy, realism was also critical. For example, the capacity to translate technical documents written in Japanese while including English terms and diagrams on the fly. We began training in our dedicated AI development environments but overgrew it quickly due to the combination of Japanese, English, and Hungarian alphabets.

Our pilot project's purpose was to optimize and speed the learning of OCR (optical character recognition) models' (see applicable architecture) and the development of related data sets in a high-performance distributed computing environment known as the HPC infrastructure. We aimed to significantly increase the models learning speed and performance by using available resources, particularly GPUs (for training) and CPUs.



Paddle Architecture Diagram. In our case, text recognition was performed using CRNN.

During the two years before the synthetic (JP-EN-HU-IT) language learning project, we encountered several obstacles in our computer settings. The most challenging difficulty was that data production and model learning on our tools could have been faster. It took up to a week to compile a data collection of 50 million photos, and we spent several weeks teaching the models. Our development process and timetable needed to be established so that the training and data creation activities could occur at night and on weekends, reducing the time spent waiting for them. The available



ITware has been a leading player in software development in the region since 2001, with diverse experience related to large enterprise systems and AI developments. More than 60 employees create unique solutions in the field of artificial intelligence and full-stack developments.

Ármin Zentai: software architect, lead developer, expert in complex infractructures. He is „Hungary's most versatile developer", innowwide in the strict sense of the word. Multiple winner of the competition. He participated in the beta testing of Kommondor and in the meantime became a real HPC specialist during the OCR project.

Péter Bessenyei: software architect, lead developer, AI trainer. He has been working in custom software development for more than 20 years, during which he used a wide variety of technologies in many projects. He has a master's degree in electrical engineering from BME, as well as a Stanford Machine Learning and Certified Scrum PO certificate. He is the trainer of the AI4DEV (www.ai4dev.hu) program.

Martin Turcsányi: AI developer, test automation developer. He is an economic IT specialist, business data analyst, currently a master's student in data science at the Obuda University. His main areas of interest include various LLM models, computer vision related topics, NLP, data-related modelling, building various software architectures and automated testing.

Attila Bíró: CTO, business development manager, researcher. He has an MSc in electrical engineering, specializing in automation and industrial informatics. He has been leading complex domestic and international projects for more than 20 years, in addition, he deals with innovation management and Japanese / international business development. In connection with artificial intelligence, he is a researcher at several universities, a PhD student in the field of informatics and healthcare.

resources, such as CPU threads, RAM, and GPU capacity, needed to be increased to complete large-scale OCR tasks efficiently. The parallel usage of many GPUs, which is critical to attaining faster results, took more work. These considerations eventually led us to select Komondor HPC infrastructure for our work.

What did we understand about the HPC environment? Very little. How much experience did we have? Practically zero. We came from a solid professional background with even more energy and motivation. We now understand that various conditions must be met to work efficiently on an OCR machine vision project in an HPC environment. It is important to remember these to avoid unnecessary expenses and frustration.

First, you must grasp the operation and design of HPC systems, including parallel and distributed computing models, resource management, and cluster scheduling (for example, the Slurm scheduler). If you are unfamiliar with these, have someone who will continually support and/or push you to learn and master them. It is critical to understand the deep learning frameworks used, such as TensorFlow and PyTorch (in our case, PaddlePaddle, which is specialized for OCR), as well as how to optimize and scale them in environments with multiple GPUs and CPUs distributed across multiple physical machines. You must be able to manage and prepare massive data sets efficiently. This includes storing and loading data in the right format (e.g., LMDB or HDF5) and optimizing IO operations to improve system performance. To minimize file management issues, you should also understand data storage restrictions and quotas and how to archive and rebuild data as needed.

Managing software environments and dependencies is also essential. You must ensure that the libraries and frameworks you use are compatible with the versions supported by the HPC infrastructure, including CUDA and cuDNN. There is no way to install another version. You should also be familiar with containerization technologies used in HPC environments, such as Singularity, because your privileges are limited, and you can install new packages or upgrade old ones with these solutions. Debugging and performance improvement require monitoring running processes and evaluating system performance using the appropriate tools and metrics. Last, it is critical to properly prepare and test on smaller data sets and in local contexts before working with full-scale data and the entire HPC infrastructure. Even in local situations, running, testing, and optimizing on many physical computers with different GPUs and CPUs is worthwhile. This allows for early detection of possible problems and fine-tuning procedures, ultimately saving time and resources.

Operating well in an HPC setting necessitates a technical understanding of systems and tools, rigorous planning and preparation, and the flexibility and problem-solving abilities to deal with unexpected obstacles. The KIFÜ can connect you with professionals if you have a solid basis. We joined during the beta testing phase, so the learning curve was longer and more intriguing than normal. Still, we learned about the system's possibilities in greater depth than anticipated.

During the test project, we obtained considerable time savings and performance gains, including a dramatic reduction in data set production time and up to a 30-40% improvement in certain model outputs. These findings demonstrate that transitioning to an HPC environment was tremendously helpful regarding time and model quality. Overall, the decision to use an HPC environment was critical to the project's success and efficiency, as the issues we encountered in our systems were too restricting.

Parallel multi-GPU algorithm research for brain EEG data processing and analysis



Zoltan Juhasz is an Associate Professor at the University of Pannonia (Department of Electrical Engineering and Information Systems). He has several decades of experience in parallel and high-performance computing using PC clusters, GPUs, cloud systems and supercomputers. He was the principal investigator of several national and international research projects. His current interest is in parallel algorithm design and implementation, and performance optimisation in hard computational problems such as biomedical signal processing and plasma simulation.

Zeyu Wang is a Biomedical Engineer (MEng, 2021, Shenyang University of Technology, P.R. China)). He is final year PhD student at the University of Pannonia. His main research interests include GPU computing, EEG signal processing and data analysis, signal decomposition methods and machine learning algorithms. In his PhD research he studies EEG signal decomposition methods and their efficient, high performance single and multi-GPU implementations.

Iffah Syafiqah binti Suhaili Iffah Syafiqah holds an MSc degree in Neuroscience from Universiti Sains Malaysia (USM, 2021) and is currently a PhD candidate under the Stipendium Hungaricum Fellowship of The Tempus Public Foundation in University of Pannonia, Veszprem. Her doctoral research explores brain oscillations using advanced EEG techniques (Event Related Potential methods, Time-Frequency and Independent Component Analyses), focusing on phase and connectivity analysis. Her work aims to deepen our understanding of brain network dynamics and its role in cognitive processes, such as early visual processing, perception of visual art and neurolinguistics.

Balint Toth is a final year MSc (Computer Science) student in the Faculty of Information Technology of the University of Pannonia. His research interest include parallel computing, GPU algorithms and object-oriented programming. His work on EEG GPU algorithms was awarded with a 1st prize in the 2024 University Student Research Conference (TDK).

Supercomputers are indispensable tools of modern science and innovation. In order to strengthen Europe's global technological position, over the past few years, the EuroHPC programme[1] deployed several high-performance supercomputers in EU member states. These computers use hundreds to thousands of specialized accelerator cards (GPUs) to perform more than 1015 (million times billion) computational operations per second. This enormous performance, however, can only be harvested with special programs that are able to perform billions of calculation steps at the same time, in parallel. Our research group focuses on designing and implementing massively parallel algorithms of key EEG processing methods for multi-GPU supercomputers.

EEG (Electroencephalography) technology is a measurement technique capable of registering the bioelectric activities of the brain with sub-millisecond resolution. Not surprisingly, EEG is an important tool in cognitive neuroscience and psychology research, and in clinical applications such as epilepsy or sclerosis multiplex diagnosis. Data is collected via electrodes placed on the scalp with sensitive analogue-digital converters. To achieve sufficient spatial resolution, which provides sufficient details for inferring the anatomical locations (sources) of the measured activities, at least 64, but preferably 128 or more scalp electrodes should be used.

Besides handling large amount of data generated in EEG experiments, EEG researchers face several challenging computational problems. EEG data is regularly contaminated by noise and unwanted bioelectric (eye blink and eye movement, muscle noise, heart-related) artefacts that have to be removed before data analysis. In addition, due to the very good electrical conductivity of the brain, the electric field potential measured on the scalp might be the additive sum of the effect of several simultaneously active cortical sources at different locations. Removing the noise, artefacts and computing the original unmixed sources all require highly compute intensive and time-consuming algorithms.

The EEG research community relies mainly on MATLAB and (more recently) Python-based EEG processing software packages (eg. EEGLAB[2], Fieldtrip[3], Brainstorm[4], MNE Python[5]) that are optimised for ease-of-use but not necessarily for execution efficiency. Single subject data processing times can easily take several hours for high-density EEG datasets and when group studies are performed, execution times on a single computer can

easily reach several days. Unfortunately, these software tools provide only limited parallelism support since they were originally designed for traditional CPU architectures and cannot benefit from modern parallel computers, especially from highly parallel, GPU (graphics processor) accelerated supercomputers.

Our research group focuses on designing and implementing massively parallel algorithms of key EEG processing methods for multi-GPU supercomputers. Our ultimate goal is to create a complete GPU-only EEG processing pipeline that can run on hundreds or thousands of GPUs simultaneously, reducing group analysis times from several days to few minutes. In contrast to MATLAB-based method developments and efforts aiming to develop function-rich EEG processing software frameworks, our work is focusing on computational performance and perfect utilisation of supercomputers. This is in line with the aims of European HPC research. The number of GPU cards found in the pre-exascale supercomputers is in the range of tens of thousands resulting in a staggering number (106– 108) of compute cores available for calculations. Since most of today's algorithms were designed for orders of magnitude smaller systems, new algorithms and programming techniques must be developed that minimise the time spent on memory operations, data communication and synchronization, in order to maximally utilise these heterogeneous hardware systems. To succeed, we need to rethink our computing strategies; we need to restructure and redesign algorithms, and develop novel numerical methods that will be suitable for such large-scale systems. This is one of the grand challenges of modern computer science.

To date, we have developed a number of important algorithms used in EEG pre-processing including (i) convolution based signal filtering, (ii) power spectral density calculation, (iii) bad channel detection and interpolation, (iv) current source density estimation[6], (v) the Infomax Independent Component Analysis algorithm[7], (vi) computation of several statistical features and (vii) various versions of the Empirical Mode Decomposition[8] algorithm and (viii) code snippets used in functional connectivity (brain network) analysis. The implementations were developed in the C/CUDA programming language. The single and multi-GPU development, and the algorithm testing and performance optimisation work have been performed on the Komondor supercomputer, using the NVIDIA A100 GPU cards of the GPU partition. Figure 1 illustrates the data partitioning scheme used for the multi-GPU implementations.

In addition to the GPU algorithm development, we have also benefited from the CPU partition of the Komondor computer, on which we were able to run multiple MATLAB-based EEG analysis pipelines simultaneously, achieving 10-20x speedup when executing group EEG analysis tasks.

We are currently in the process of testing the scalability of the implemented algorithms and integrating them into a configurable pre-processing GPU pipeline. Figure 2 shows the scalability results for the Wavelet Transform implementation.

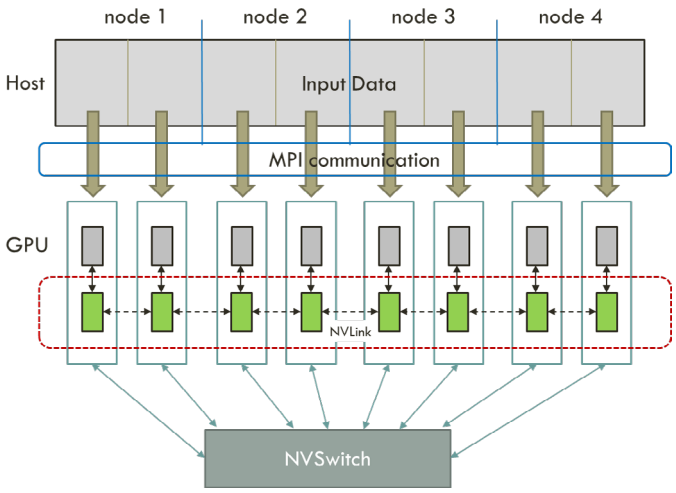


Figure 1: *. Data partitioning for parallel multi-GPU algorithm execution.*

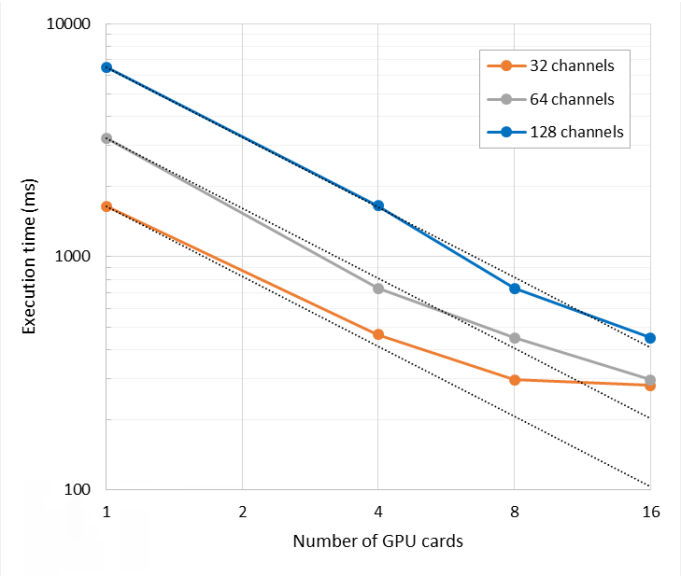


Figure 2: *Execution times of the Wavelet Transform on different number of GPU cards (data length = 65536, 64 wavelets per channel). Ideal scaling is depicted with the point line. Note the near perfect scaling in the case of 128 signal channels.*

1. "Homepage - EuroHPC JU." https://eurohpc-ju.europa.eu/index_en (accessed Aug. 17, 2024).

2. A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," J. Neurosci. Methods, vol. 134, no. 2004, pp. 9–21, 2004, doi: 10.1016/j.jneumeth.2003.10.009.

3. Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data," Comput. Intell. Neurosci., vol. 2011, 2011, doi: 10.1155/2011/156869.

4. F. Tadel, S. Baillet, J. C. Mosher, D. Pantazis, and R. M. Leahy, "Brainstorm: A user-friendly application for MEG/EEG analysis," Comput. Intell. Neurosci., vol. 2011, no. January 2004, 2011, doi: 10.1155/2011/879716.

5. A. Gramfort et al., "MNE software for processing MEG and EEG data," Neuroimage, vol. 86, pp. 446–460, 2014, doi: 10.1016/j.neuroimage.2013.10.027.

6. F. Perrin, J. Pernier, O. Bertrand, and J. F. F. Echallier, "Spherical splines for scalp potential and current density mapping.," Electroencephalogr. Clin. Neurophysiol., vol. 72, no. 2, pp. 184–7, Feb. 1989, Accessed: Nov. 18, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0013469489901806.

7. A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution.," Neural Comput., vol. 7, no. 6, pp. 1129–1159, 1995, doi: 10.1162/neco.1995.7.6.1129.

8. N. E. Huang et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," Proc. R. Soc. A Math. Phys. Eng. Sci., vol. 454, no. 1971, pp. 903–995, Mar. 1998, doi: 10.1098/rspa.1998.0193.

How can weather forecasts be more accurate?



Péter Kardos
Research and Development Engineer
MouldTech Systems Kft.

Péter Kardos has been working in the Meteorological Research and Development Division of MouldTech Systems Kft for 4 years. He has several decades of experience in the field of meteorological service development, of which for nearly 15 years he led the team providing aviation meteorology services at Budapest Ferenc Liszt International Airport. The department, which was once part of the HungaroControl organization, produced nowcasting forecasts supported by neural networks for the first time in the country, and uniquely, developed a real-time verification system integrated into a quality management system. In recent years, he has dealt with data assimilation studies of forecasting systems that use numerical methods, with particular regard to the use of drone measurements in this field.

The desire to discover the weather is an old one with mankind. Obviously, it is a great advantage to know what the weather is going to be like in the next few hours or days. Despite the fact that many people feel that meteorologists are still unable to provide accurate forecasts, the truth is that the area of weather forecasting, and within it numerical weather prediction, has made enormous progress in recent decades.

In the 1990s, the main challenge was to produce numerical forecasts with a time advantage, i.e. to make the virtual time running during the calculations spin faster than the real time. This obviously meant a strong trade-off between spatial resolution and speed, because the more detailed numerical models were, the more computationally intensive and consequently slower the process. Faster computers gradually opened up the possibility of higher resolution models, so the development of faster CPUs and numerical models evolved hand in hand. Progress in processors was so rapid that practically by the 2010s we reached the point where models were unable to evolve, not because of a lack of computing power, but because developers were faced with a different type of problem. Namely, the lack of data. Despite increasing the resolution, the models were no longer delivering more detailed information, because the measurement data that updated the initial state of the atmosphere turned out to be insufficient.

Obviously, to move forward, the data gap in numerical models needs to be addressed in some way. Several approaches have been attempted, such as the collection of meteorological data from aircraft, or data feed based on the atmospheric interaction of GNSS signals. These have delivered some improvements, but developers are looking for further ways to fill the data gap in the lower few kilometres of the atmosphere.

Our research team has been studying the meteorological applications of UAVs, more popularly known as drones, since 2011. Drones are capable of transmitting atmospheric data measured by their own sensors in real time, and even provide the possibility to provide additional information by actively measuring weather-dependent data gaps in specific locations.

We have developed over the past two years a meteorological measuring drone system capable of continuous meteorological profiling under operational conditions, measuring temperature, humidity, pressure and wind parameters up to 2500m above the surface.

However, many technical questions have been raised about the use of the data. Can we use these measurements to make forecasts more accurate? How can we integrate them into the existing data collecting system? How often do we need to measure? Up to what altitude? What is the ideal profiling network density?

In order to answer these questions, we have launched a specific project based on a numerical prediction system simulation environment that is capable of running a large number of simulations. This is necessary for several purposes. Firstly, in order to be able to start a meaningful study at all, we need to prepare a so-called error statistic for the forecast domain, which will later help us to consider how to take a measurement into account when we want to improve a real weather field based on it. Once the error statistics have been compiled, we can then investigate the spatial effects or choose the ideal way to process the measurement. Finally, of course, we can also answer the question of the impact of a single drone measurement in a given setting, and the role of networked, multi-drone measurements.

We developed our simulation environment based on container technology. Our LoA with KIFÜ, which was possibly one of the first SME collaborations with the Agency, gave us the opportunity to test and develop our system on several HPC platforms. The installation of Komondor, the country's highest-capacity computer, further expanded our capabilities, and its large CPU capacity ensured that results were delivered quickly despite the significant computational demand.

In our research so far, we have run more than 16,000 weather simulations, all of which consisted of a 6-6 hour forecast time window. Based on the results, it can be seen that the ideal density of the drone profiling network based on the current settings of our forecasting model is approximately 30 km, i.e. the whole country could be covered by 20-30 drone profiling station. The impact of drone measurements is of course weather-dependent and seems clearly changeable, but it is clear that a single measurement at a critical location and time could lead to significant improvement.



MouldTech Systems Ltd's meteorological prototype drones, the only one from Hungary to contribute in the campaign coordinated by the World Meteorological Organization.

Of course, our research doesn't finish here, we are working on testing multiple drone measurements at the same time, and there is plenty of engineering work to do in addition to meteorological developments.

Apparently, the work we do is quite rare in the world. At international level, the World Meteorological Organization (WMO) is coordinating a drone measurement campaign in the year 2024, focusing on precisely the questions mentioned above. If the positive expectations of drone measurements are confirmed internationally, we can expect to see an explosion of developments in this field in the next 5-10 years. We believe that this will revolutionise the accuracy of forecasts, further expand the possibilities of machine intelligence-assisted forecasting and contribute to the era of data-driven precision meteorology in atmospheric science, which will have a positive impact on all aspects of our life.

Simulation of urban air pollution, development of a general-purpose HPC portal and fluid dynamics solvers





László Környei graduated as a computer scientist and physicist from the University of Szeged and earned his PhD in 2010. He has worked in HPC environments at SZFKI and AUDI Hungária. Currently, he teaches at Széchenyi István University, with research focused on air pollution modeling.

Zoltán Horváth is a university professor who graduated as a mathematician from Eötvös Loránd University (ELTE) in 1989, where he also earned his PhD and habilitation in numerical mathematics. Since 2006, he has been leading the Department of Mathematics at Széchenyi István University and founded the MathSO research group in 2009. He has been the head of the HU-MATHS-IN network since 2013.

József Bakosi obtained a mechanical engineering degree from the University of Miskolc in 1999 and a PhD in numerical fluid dynamics from George Mason University in 2008. After 20 years of engineering and research work, he is now engaged in HPC software development at Széchenyi István University.

Mátyás Constans is a scientific mathematical programmer at Széchenyi István University, with over 8 years of experience in GPU programming in the fields of GPGPU and rendering. He is currently working on the implementation of Navier-Stokes equations on GPUs, using MPI and CUDA technologies, specializing in low-level programming.

Flow simulation modeling is among the computer applications with the highest computational demands. From the applications developed by our research group, we present one industrial (aeroacoustic) and one environmental (urban air quality monitoring) application. The code development, benchmarking, and execution performed on the Komondor supercomputer add significant value to these applications. The software development and applications presented in the article were largely realized through the HiDALGO2 EuroHPC project.[1]

High levels of urban air pollution are proven to be one of the most significant risk factors in the development of cardiovascular diseases, contributing to millions of deaths annually.[2] The EU has defined air pollution indicators, such as the number of days per year exceeding limit values, which member states are required to report. Within the EU, a dedicated working group, FAIRMODE[3], supports the work of teams using modeling to assist these reports, and our research group is an active member of this initiative.

Our research group has developed an HPC-supported procedure called UAP (Urban Air Pollution), with which we can generate 3D geometry of any city featured on OpenStreetMap and divide it into cells. For longer modeling times, we work with 10 million cells, while for validation and benchmarking, we use 30-70 million cells. The 10-million-cell model allows for a resolution of 2-3 meters at street level. Following this, we compute wind flow in the 3D geometry using boundary conditions derived from global simulations, solving the transient Navier-Stokes equations with turbulence modeling (k-ε or k-ω) using the UAP-FOAM software based on OpenFOAM. In addition to UAP-FOAM, we use RedSim, a completely in-house developed code natively using OpenMPI + CUDA[4], and Xyst, another proprietary code.[5] After calculating air flow, we solve the dispersion modeling by incorporating point, line, and area sources of pollutants, along with meteorological data such as atmospheric boundary layer height and solar radiation intensity, into the solution of the linear advection-diffusion 3D differential equation.

We applied our UAP model during the FAIRMODE hackathon, using data from Antwerp for the year 2016. As the only group utilizing HPC technology, we were able to run and evaluate the hackathon task on the Komondor supercomputer in two weeks, using transient simulations, which was unique among participants. The results were published in.[6]

Among industrial fluid dynamics tasks, aeroacoustic problems are considered classic and challenging. Based on industrial demand, our research group calculated the noise generated by pressure changes from gas flowing at high speed out of a pipe. The difficulty of the task is indicated by the need to accurately detect a pressure fluctuation of approximately 5 Pa, which required running

the simulation on at least 70 million cells. The calculations were performed using the RedSim software and validated against measurement results from company engineers and university researchers.[7]

When we ran the solution on the university's NVIDIA V100S graphics card, it took 5 days to complete the task – while in an industrial environment, this calculation took several weeks at the time of validation (3 years ago). With access to the Komondor supercomputer, we developed RedSim for multi-GPU use, and according to benchmarks, the code would run about four times faster on 8x A100 GPUs.

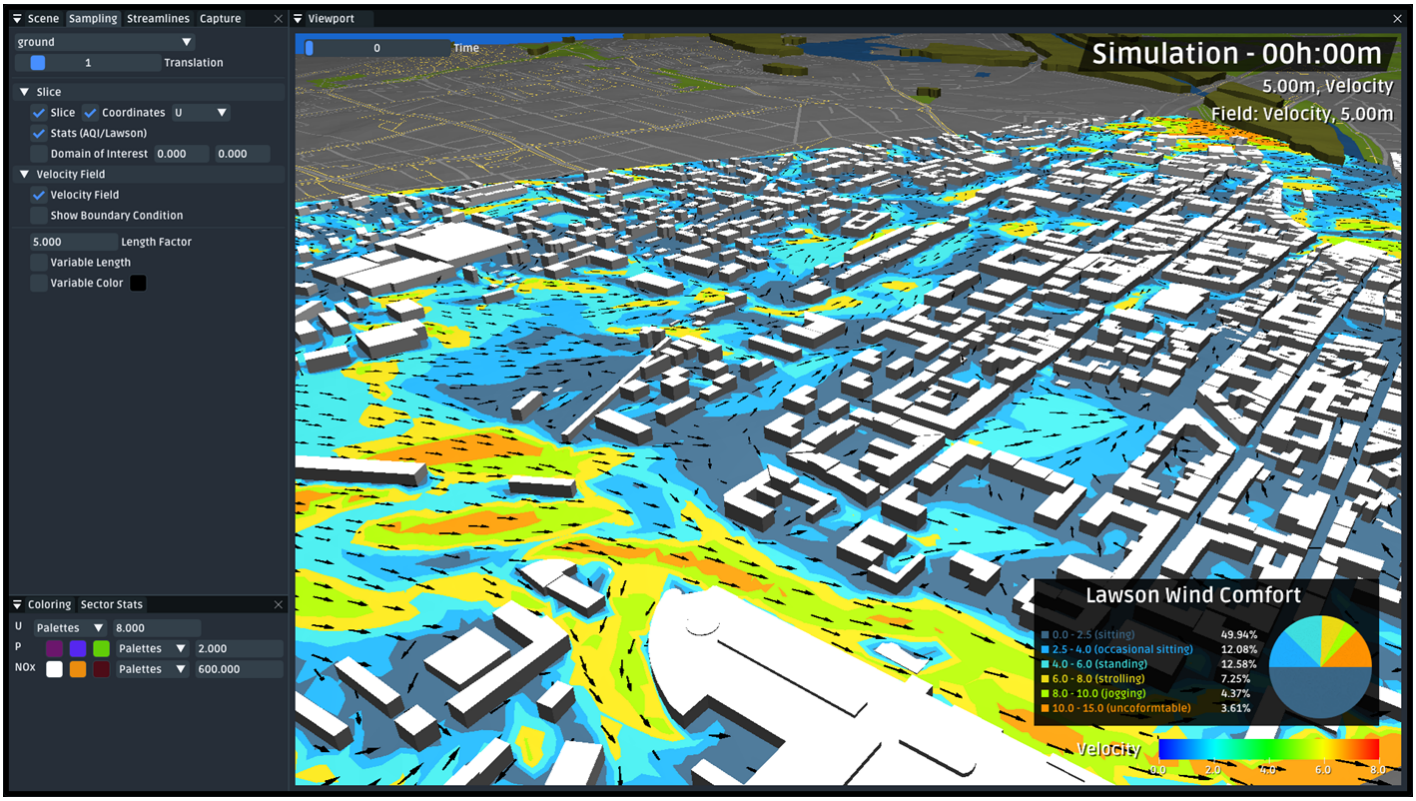


Figure 1: *Simulation of urban air pollution using UAP-FOAM on different HPC systems, with a 14-million-cell model. We present the runtime for one hour of simulated time on a log-log scale, based on the number of nodes, for the Komondor, HLRS Hawk, and three EuroHPC machines.*

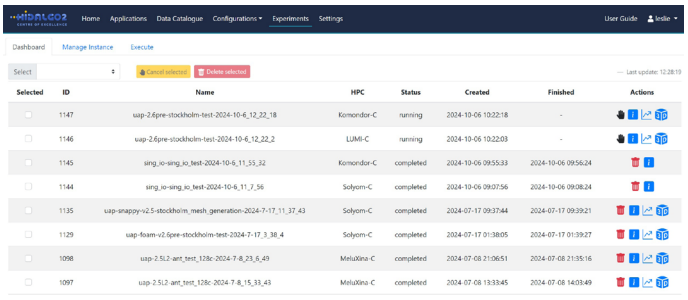


Figure 2: *Calculation of urban wind comfort using RedSim in the 3D model of the city of Győr, with Lawson wind comfort results depicted in a cross-section at two meters height. The HPC results are visualized directly in the browser using the CFDR viewer developed by the group, without the need for installing software.*

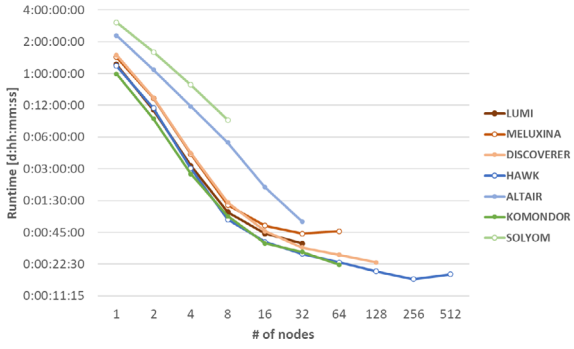
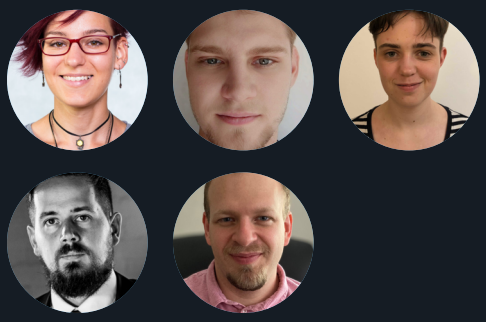


Figure 3: *List of simulations launched on the custom-developed MathSO portal. Using this portal, simulations are run and evaluated on Komondor and EuroHPC machines via containerization techniques. The portal supports two-factor authentication and VPN connections, and results can be displayed using the integrated CFDR viewer.*

1. https://www.hidalgo2.eu/
2. https://www.unicef.org/press-releases/air-pollution-accounted-81-million-deaths-globally-2021-becoming-second-leading-risk
3. https://fairmode.jrc.ec.europa.eu/
4. https://redsim.mathso.sze.hu/about/overview
5. József Bakosi et. al. „Complex-Geometry 3D Computational Fluid Dynamics with Automatic Load Balancing" in Fluids 8(5):147, DOI: 10.3390/fluids8050147
6. Fernando Martin et. el. „Using dispersion models at microscale to assess long-term air pollution in urban hot spots: A FAIRMODE joint intercomparison exercise for a case study in Antwerp" in Science of The Total Environment, Volume 925, 15 May 2024, 171761, DOI: https://doi.org/10.1016/j.scitotenv.2024.171761
7. Zoltán Horváth et. al. „Higly accura numerical solution of the Navier-Stokes equations in pipes." in Hu-Maths-In success stories of Mathematical short-term projects for Industry in 2017-2021 pp. 122

Advancing Microbiome Engineering with High-Performance Computing



Group members (Neural Bioinformatics Research Group (PPCU-FITB)):

Babett Bodnár
Bendegúz Filyó
Judit Juhász
Dániel Krizsán
Balázs Ligeti

Balázs Ligeti is an Associate Professor at Pázmány Péter Catholic University and leads the Neural Bioinformatics Research Group. The group focuses on bioinformatics, particularly in microbial genomics and developing genomic language models. They have been working on the ProkBERT model family, which addresses challenges in microbial sequence analysis, such as promoter prediction and phage detection. Their work combines insights from machine learning and biology to contribute to more effective bioinformatics tools.

In our project supported by KIFÜ, we trained and tested genomic language models to advance microbiome engineering.

Biomedical science generates vast amounts of complex data, necessitating high-performance computing for effective analysis. The challenge lies not only in the volume of data but also in its complexity, requiring sophisticated models to extract new insights. Our research focuses on analyzing genomic data using statistical and artificial intelligence (AI) methodologies.

The majority of this data comprises nucleotide sequences obtained through DNA or RNA sequencing. Sequencing enables us to infer chromosomal structures and identify uncultivable microorganisms. However, the sheer volume of sequencing data surpasses our current analytical capabilities, resulting in underutilized datasets.

Our research centers on the microbiome – the complex communities of microorganisms, including bacteria, viruses, fungi, protozoa, and microscopic algae. Despite their microscopic size, these organisms play crucial roles in health, disease, industrial processes, and biotechnology. They perform complex tasks through coordinated or competitive interactions. The microbiome exhibits immense diversity; many species remain undiscovered, and the functions of numerous proteins within known genomes are still unknown. Exploiting this diversity holds significant potential for combating pathogens, addressing antibiotic resistance, and engineering microbial communities for applications such as bioremediation.

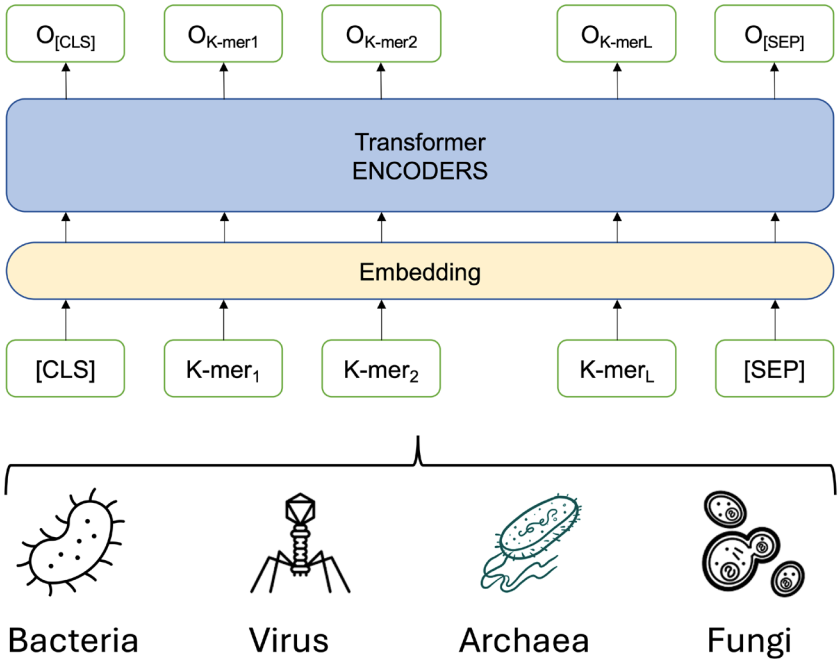
Traditional experimental approaches to develop new applications are often costly and challenging, particularly when culturing certain bacteria is infeasible. By applying AI, we can perform virtual experiments and optimize processes and genes, facilitating solutions to microbiome engineering challenges.

We address these challenges using neural networks, specifically genomic language models. Analyzing genomic data with AI requires computationally interpretable representations, which involve encoding sequences into vectors. These vector representations enable the application of algorithms to predict properties such as the pathogenicity of microorganisms (figure).[1] A genomic language model serves as the mechanism that transforms sequences into these vector embeddings.

Utilizing the Komondor supercomputer, we developed and evaluated the ProkBERT language model family.[1] These models are compact yet efficient and are readily accessible for use by the broader research community.

We have demonstrated the utility of ProkBERT in various tasks, notably bacteriophage detection. Bacteriophages are viruses that infect bacteria and have potential applications in phage therapy—using viruses to target and eliminate antibiotic-resistant pathogens. Identifying and characterizing phages within microbiomes is essential for this purpose. However, our knowledge of phages is limited; most remain uncharacterized, rendering traditional database comparison methods insufficient. These methods are complex, multi-step processes that are slow and difficult to scale to large datasets.

In contrast, ProkBERT models address tasks such as phage identification with higher accuracy and efficiency, and significantly faster than other larger neural network-based solutions like the Nucleotide Transformer (developed by InstaDeep, NVIDIA, and TUM) or DNABert-2.[2,3] Notably, ProkBERT performs well even on phage types not encountered during training, including viruses from deep-sea and other extreme environments.

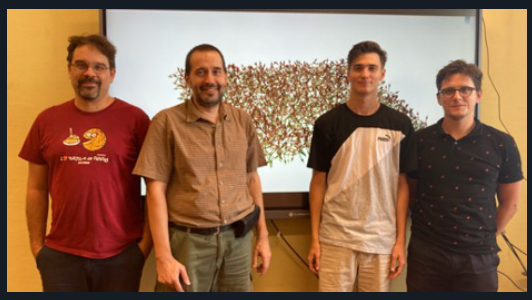


ProkBERT model. The model was developed for microbes and was pretrained on large corona of bacteria, viral, archaea and fungi sequences. A new k-mer based tokenization, the local-context-aware (LCA) tokenization approach has been developed for the genomic sequences. The model provides sequence representations (encoder type) and can be used for a wide range of tasks, such clustering, classification or regressions.

By advancing genomic language models through high-performance computing, we contribute to the field of microbiome engineering, enabling more effective utilization of the vast genomic data available. Our work facilitates the development of innovative solutions to complex biological problems, paving the way for new applications in medicine, biotechnology, and environmental science.

1. Ligeti, Balázs, et al. "ProkBERT family: genomic language models for microbiome applications." Frontiers in Microbiology 14 (2024): 1331233.
2. Zhou, Zhihan, et al. "Dnabert-2: Efficient foundation model and benchmark for multi-species genome." arXiv preprint arXiv:2306.15006 (2023).
3. Dalla-Torre, Hugo, et al. "The nucleotide transformer: Building and evaluating robust foundation models for human genomics." BioRxiv (2023): 2023-01.

The Magnetic Exchange – MagEx project



Some members of the MagEx research team from ELTE.

The main goal of the MagEx group's work is the in-depth study of magnetic van der Waals heterostructures and the development of novel magnetic memory architectures (MRAM) based on these structures. The researchers participating in the group: László Oroszlány (Assistant Professor, ELTE), János Koltai (Assistant Professor, ELTE), Zoltan Tajkov (Postdoctoral Researcher, ELTE), Assem Al-Assaf (PhD Student, ELTE), András Balogh (PhD Student, ELTE), Tibor Dániel Pozsár (PhD Student, ELTE), Balázs Marcell Sipos (BSc Student, ELTE), Levente Rozsa (Senior Research Fellow, ELTE)

Magnetism, a manifestation of a fundamental force intrinsic to nature, continues to captivate scientists and drive technological innovation. From the compass that guided early explorers to the hard drives storing our digital memories, magnetic materials have been pivotal in shaping our modern world. In the realm of information technology, these materials are the unsung heroes behind the scenes, enabling data storage, sensors, and even the speakers in our devices.

Recent years have seen a surge of interest in two-dimensional (2D) van der Waals materials. These incredibly thin materials, often just a few atoms thick, possess unique properties that differ from their bulk counterparts. Their potential applications span from ultra-efficient electronics to flexible displays and advanced sensors. The ability to stack these 2D materials like atomic-scale Lego bricks opens a world of possibilities for creating tailored materials with specific properties.

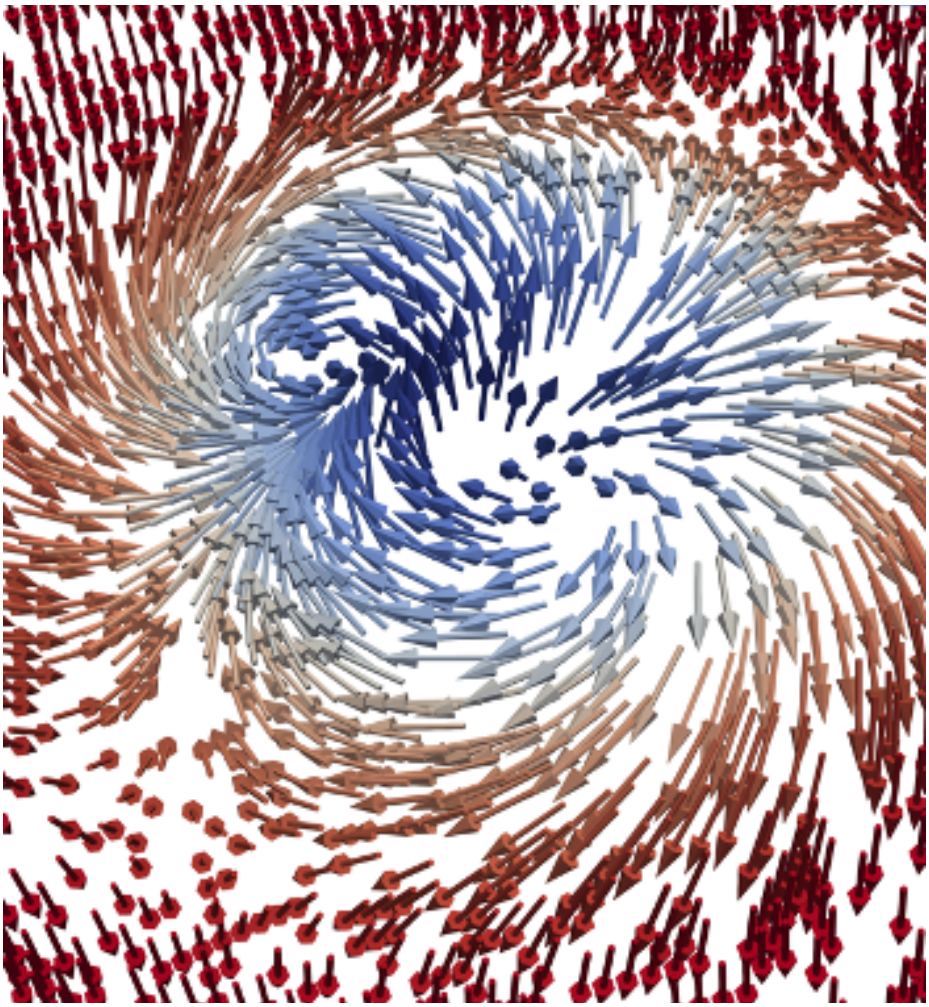
At the intersection of magnetism and 2D materials lies a new frontier: novel magnetic 2D materials like chromium iodide (CrI_3) and iron germanium telluride (Fe_3GeTe_2). These materials retain their magnetic properties even at the atomic scale, a feature that could revolutionize memory architectures. Imagine computer memory that is not only incredibly dense but also consumes far less power. These materials could pave the way for neuromorphic computing, mimicking the efficiency of the human brain, and usher in a new era of quantum computing. As we continue to explore these nanoscale marvels, we edge closer to a future where information processing is faster, more efficient, and more powerful than ever before.

Scientists involved in the MagEx are leading the charge in the field of condensed matter physics, especially when it comes to the theoretical study of modern low-dimensional materials. Contributors to MagEx have distinguished themselves by focusing on the electronic structure of these materials, an area of study that is both complex and crucial for advancing our understanding of quantum physics and material science. Low-dimensional materials, such as two-dimensional (2D) systems, exhibit unique properties that differ from their three-dimensional counterparts. MagEx's capability in modeling and analyzing these systems places it at the forefront of theoretical condensed matter physics, enabling groundbreaking research in this cutting-edge field.

In collaboration with international partners, particularly the University of Oviedo, coworkers in MagEx have recently developed a sophisticated computational framework that allows for the study of magnetism in low-dimensional systems. This framework is a significant advancement in condensed matter physics, as it offers a microscopic understanding of magnetic interactions at the atomic level. Such collaborations are vital in pushing the boundaries of what can be achieved theoretically and experimentally, as they bring together diverse expertise and resources to tackle complex scientific challenges.

The magnetic phases in low-dimensional systems can be far more intricate than the simple ferromagnets familiar from everyday life, such as those used in refrigerator magnets. These systems can exhibit a range of magnetic behaviors, including antiferromagnetism, where adjacent spins align in opposite directions, and more exotic states like skyrmions – tiny, swirling magnetic configurations (depicted in the figure) that are topologically protected. Understanding these complex magnetic phases is not only a theoretical challenge but also an experimental one, as these novel structures have promising applications in next-generation technologies, such as high-density data storage and quantum computing.

To conduct this research, MagEx project leverages the Komondor cluster, the high-performance computing resource provided by the Governmental Agency for IT



Magnetic moments forming a skyrmion.

Development. This powerful computational tool enables us to perform large-scale ab initio calculations and atomistic magneto-dynamics simulations, which are essential for exploring how microscopic magnetic interactions translate into observable properties, such as critical temperatures and magnetic textures. MagEx's research is further strengthened through partnerships with experimental experts from Helmholtz-Zentrum Berlin and the University of Valencia, ensuring that the theoretical insights gained are grounded in experimental reality. Together, these efforts place MagEx and its partners at the cutting edge of condensed matter physics, with the potential to revolutionize our understanding of magnetism in low-dimensional systems.

IMPRINT

Responsible Publisher:
Governmental Agency for Information Technology Development
Professional Proofreader: Orsolya Feró, Gyula Ujlaki
Copy Editor: Orsolya Feró
Graphic Design: Teodóra Matisz
Address: Government Information Technology Development Agency
1134 Budapest, Váci út 35
www.kifu.gov.hu
E-mail: info@kifu.gov.hu


HPC @hu
Competence Centre
KIFÜ
Funded by
the European Union
EURO²
The EuroCC 2 project is supported by the European Union's Horizon 2020 research and innovation framework program in accordance with Grant Agreement No. 101101903
hpc.kifu.hu
hpc@kifu.gov.hu