50 years 1973/2023 CYFRONET

Academic Computer Centre

CYFRONET AGH

2023

ACC CYFRONET AGH is a leading unit empowered by the Committee for Scientific Research to develop and manage the High-Performance Computers (HPCs) and Cracow Metropolitan Area Network (MAN). CYFRONET is the coordinator of the PLGrid Program and is recognized by the National Centre for Research and Development as a Centre of Excellence in the area of grid and cloud services.

Dear Readers!

In 2023, Cyfronet celebrates its 50th anniversary. For these fifty years, we have served Polish science by providing computing and network resources and assistance within the scope of our competences. It was a time of extremely great technological breakthroughs that made it possible to conduct scientific research in an increasingly advanced manner. The role of computers has changed: from machines available only in selected specialized units, they have become widely available and ubiquitous IT tools. Many extremely important changes have also taken place in Cyfronet. However, the essence of our activity, our mission, which we have fulfilled over the years, has remained the same. We support science, innovative economy and the information society by providing computing power, data storage systems, network services, software, as well as helping and transferring specialist knowledge. Treating the need for change as a natural sequence of events not only in science, we are constantly working on modifying the services offered by using the latest technological solutions, and those in IT are a continuous series of radical changes.

The years 2021-2022 brought several significant fruits of our constant efforts to provide Polish science with the best computing possibilities. We launched two new supercomputers: Ares and Athena, which became the fastest supercomputer in Poland. Athena's configuration includes servers with AMD EPYC processors and NVIDIA cards with A100 GPGPUs. Thanks to these resources, Athena's theoretical computing power is 7.7 PetaFlops, and for the purposes of AI calculations, it is the computing power of almost 240 PetaFlops!

In June 2022, for the first time in history, 3 machines from one Polish computing center were included in the TOP500 list of the fastest supercomputers. These are working in Cyfronet: Athena (105th position), Ares (290th) and Prometheus (475th). Cyfronet supercomputers also took high places on the Green500 list of the most energy-efficient supercomputers – headed by Athena on the prestigious 9th place. At the same time, together with partners forming the LUMI consortium led by Finland, we celebrated the fact that the LUMI supercomputer was ranked third in the world and first in Europe on the TOP500 list. This is an extremely important achievement, as Polish scientists have been able to perform calculations on a supercomputer from the "world's top shelf". LUMI is accessed via the PL-Grid program portal, which is coordinated by Cyfronet.

In addition, the European High Performance Computing Joint Undertaking (EuroHPC JU) has selected Poland as a site for the installation of a next-generation supercomputing system as part of a pan-European data processing infrastructure. A *mid-range* supercomputer with a power at least several times greater than that of Athena will be installed by the end of 2023. The operator will be ACC Cyfronet AGH, which confirms the recognition of our competences in the field of HPC (High Performance Computing) in the international arena.

What challenges will the world face in the next 50 years? Some of them, such as energy issues, those related to environmental protection and anthropogenic threats, can already be outlined. Others will appear unexpectedly. However, the experience of half a century of our activity shows that regardless of the prevailing conditions, cooperation based on trust and mutual understanding is key. Therefore, on behalf of Cyfronet, I would like to thank you for the last 50 years of fruitful cooperation and invite you to further joint activities!

Yours sincerely,

Prof. Kazimierz Wiatr
Director of ACC Cyfronet AGH

DIRECTOR
Kazimierz Wiatr, Prof.
phone: +48 12 633 34 26
e-mail: Kazimierz.Wiatr@cyfronet.pl

Deputy Director for IT Infrastructure
Karol Krawentek, MSc Eng.
phone: +48 12 633 34 26
e-mail: Karol.Krawentek@cyfronet.pl

Deputy Director for High Performance Computers
Marek Magryś, MSc
phone: +48 12 633 34 26
e-mail: Marek.Magrys@cyfronet.pl

Deputy Director for Administrative Affairs
Agnieszka Szymańska, MSc
phone: +48 12 633 34 26
e-mail: Agnieszka.Szymanska@cyfronet.pl

Deputy Director for Financial Affairs
Angelika Zaleska-Walterbach, MSc
phone: +48 12 633 80 53
e-mail: Angelika.Zaleska@cyfronet.pl

Computer Networks Department – Krzysztof Gaweł, MSc (manager)
Phone: +48 12 634 10 25, e-mail: Krzysztof.Gawel@cyfronet.pl

High Performance Computers Department – Patryk Lasoń, MSc Eng. (manager)
Phone: +48 12 632 33 55, e-mail: Patryk.Lason@cyfronet.pl

Data Security Department – Grzegorz Sułkowski, PhD Eng. (manager)
Phone: +48 12 632 33 55, e-mail: Grzegorz.Sulkowski@cyfronet.pl

Data Storage Department – Adrian Marszalik, MSc Eng. (manager)
Phone: +48 12 632 33 55, e-mail: Adrian.Marszalik@cyfronet.pl

HPC Software Department – Łukasz Flis, MSc (manager)
Phone: +48 12 632 33 55, e-mail: Lukasz.Flis@cyfronet.pl

Operational Center Department – Andrzej Zemła, PhD (manager)
Phone +48 12 632 33 55, e-mail: Andrzej.Zemla@cyfronet.pl

Users Department – Robert Pająk, MSc Eng. (manager)
Phone: +48 12 632 33 55, e-mail: Robert.Pajak@cyfronet.pl

Technical Department Nawojki – Damian Trela, MSc Eng. (manager)
Phone: +48 12 632 33 55, e-mail: Damian.Trela@cyfronet.pl

Technical Department Podole – Mariusz Kula, MSc Eng. (manager)
Phone: +48 12 632 33 55, e-mail: Mariusz.Kula@cyfronet.pl

Administrative Department – Agnieszka Szymańska, MSc (manager)
Phone: +48 12 633 34 26, e-mail: Agnieszka.Szymanska@cyfronet.pl

Financial Department – Angelika Zaleska-Walterbach, MSc (manager)
Phone: +48 12 633 80 53, e-mail: Angelika.Zaleska@cyfronet.pl

Projects Department – Aleksandra Pałuk, MSc (manager)
Phone: +48 12 632 33 55, e-mail: Aleksandra.Paluk@cyfronet.pl

Laboratory of the PLGrid Program – Jacek Kitowski, Prof. (manager)
Phone: +48 12 633 34 26, e-mail: kito@agh.edu.pl

Laboratory of Computing Acceleration and Artificial Intelligence
Paweł Russek, PhD DSc Eng., Associate Professor (manager)
Phone: +48 12 633 34 26, e-mail: Pawel.Russek@cyfronet.pl

Laboratory of Cloud Technologies – Łukasz Dutka, PhD (manager)
Phone: +48 12 633 34 26, e-mail: Lukasz.Dutka@cyfronet.pl

Laboratory of Data Processing – Roksana Wilk, Eng. (manager)
Phone: +48 12 632 33 55, e-mail: Roksana.Wilk@cyfronet.pl

Laboratory of Interdisciplinary Scientific Computing – Joanna Kocot, MSc Eng. (manager)
Phone: +48 12 632 33 55, e-mail: Joanna.Kocot@cyfronet.pl

Laboratory of Quantum Computing – Mariusz Sterzel, PhD (manager)
Phone: +48 12 632 33 55, e-mail: Mariusz.Sterzel@cyfronet.pl

Laboratory of Informatics Methods in Medicine – Marian Bubak, PhD Eng. (manager)
Phone: +48 12 633 34 26, e-mail: bubak@agh.edu.pl

Laboratory of Parallel Algorithms – Bogusław Cyganek, Prof. (manager)
Phone: +48 12 632 33 55, e-mail: cyganek@agh.edu.pl

Laboratory of Applications of Computational Techniques
Łukasz Rauch, PhD DSc Eng., Associate Professor (manager)
Phone: +48 12 632 33 55, e-mail: lrauch@agh.edu.pl

Laboratory of Visual Techniques – Jacek Przybylski, MSc (manager)
Phone: +48 12 632 33 55, e-mail: Jacek.Przybylski@cyfronet.pl


ATHENA
moc dla polskiej nauki


ATHENA


Cyfronet a zadania ery sztucznej inteligencji i komputerów kwantowych
ATHENA

Term of office: 2021-2024

Marek Jarnicki, Prof. *The Chairman*	Jagiellonian University
Tadeusz Lesiak, Prof. *Deputy Chairman*	Institute of Nuclear Physics Polish Academy of Sciences
Jakub Badowski, MSc	Oil and Gas Institute National Research Institute
Marek Gorgoń, Prof.	AGH University of Science and Technology
Jarosław Górski, MSc Eng.	Academy of Physical Education
Robert Gryboś, PhD	Jerzy Haber Institute of Catalysis and Surface Chemistry Polish Academy of Sciences
Tomasz Hachaj, Prof.	Pedagogical University in Cracow
Marek Kisiel-Dorohinicki, Prof.	AGH University of Science and Technology
Mariusz Kwinta-Pudełko, Msc Eng.	Agricultural University in Cracow
Wojciech Majka, MSc Eng.	Jagiellonian University Medical College
Tomasz Masier, MSc Eng.	Institute of Forensic Expertise
Roman Mazur SDB, PhD	Pontifical University of John Paul II in Cracow
Paweł Mikłaszewski, MSc	Academy of Music Academy of Fine Arts Academy of Theatre Arts
Wojciech Palacz, PhD	Jagiellonian University
Dariusz Put, Prof.	Cracow University of Economics
Elżbieta Tomczyk, MSc	Łukasiewicz Research Network Cracow Technological Institute
Dariusz Żelasko, MSc Eng.	Cracow University of Technology
Jacek Barski, Deputy Director	Marshal Office of the Małopolska District
Paweł Schmidt, Director	Cracow City Hall


CYFRONET AGH

Athena – strong support for scientific calculations

Athena achieves the theoretical computing power of over **7.7 PFlops** (7709 TFlops), which gave the machine 105th place on the TOP500 list in June 2022 and makes it the fastest supercomputer in Poland at present. The new, accelerated computing system, installed in Cyfronet in 2021, is to provide the Polish scientific community and innovative economy with the most modern computing resources based on the latest generation GPGPU processors and accelerators along with the necessary data storage subsystem based on very fast flash memories.

Athena in numbers	
Number of computing cores	6 144
Number of GPGPUs	384
Computing power	7.7 PFlops
TOP500 – the list of the world's fastest computers (June 2022 edition)	105th position

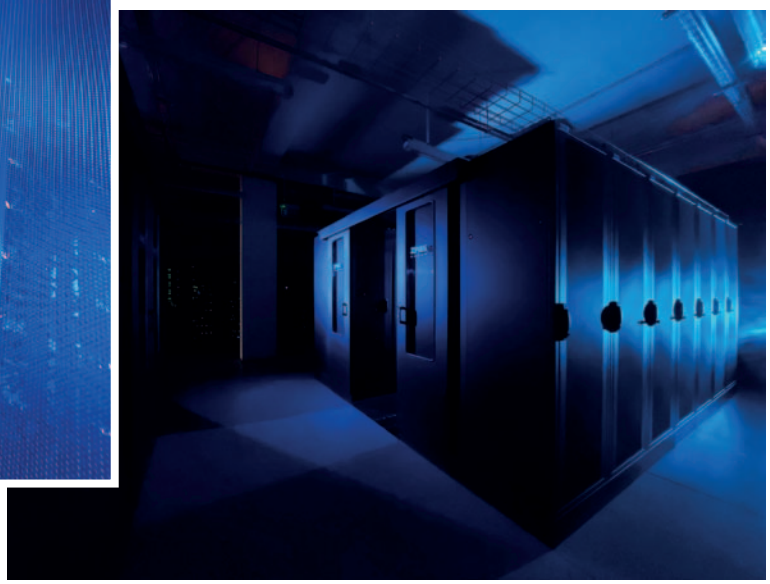
Athena's configuration includes: 48 servers with AMD EPYC processors and 1 TB of RAM (6,144 CPU compute cores in total) as well as 384 NVIDIA A100 GPGPU cards.

The indispensable element enabling the use of such high computing power in an effective way is the provision of a high-performance internal network of a supercomputer (Infiniband HDR with 4 x 200 Gb/s bandwidth per server) and a very fast disk subsystem. It is built on the basis of the Lustre open source software, currently used in Ares and Prometheus supercomputers, and dedicated disk servers equipped with flash memory in the NVMe standard. The system was installed in the existing Cyfronet data center and integrated with the PLGrid infrastructure.

This type of infrastructure meets the needs of users of Cyfronet supercomputers, who use the computing infrastructure both to perform standard high-performance scientific simulations (HPC) and to apply artificial intelligence (AI) and machine learning (ML) methods in order to conduct research in the field of medicine, pharmacology, biology, chemistry, physics and many other fields of science. **Athena's computing power for AI computing is almost 240 PetaFlops!**

The expected effect of delivering specialized computational resources of Athena will be the extension of the scope of research works, the possibility of undertaking advanced simulations and analyses, and increasing the possibilities of processing continuously flowing data from laboratories around the world. The direct expected results of the work will be articles and scientific studies, patents, and in the longer term, innovative solutions that may be the basis for the development of new solutions in the economy.


ATHENA
ATHENA
ZPAS
GROUP
AGH



Ares – towards shorter computation time

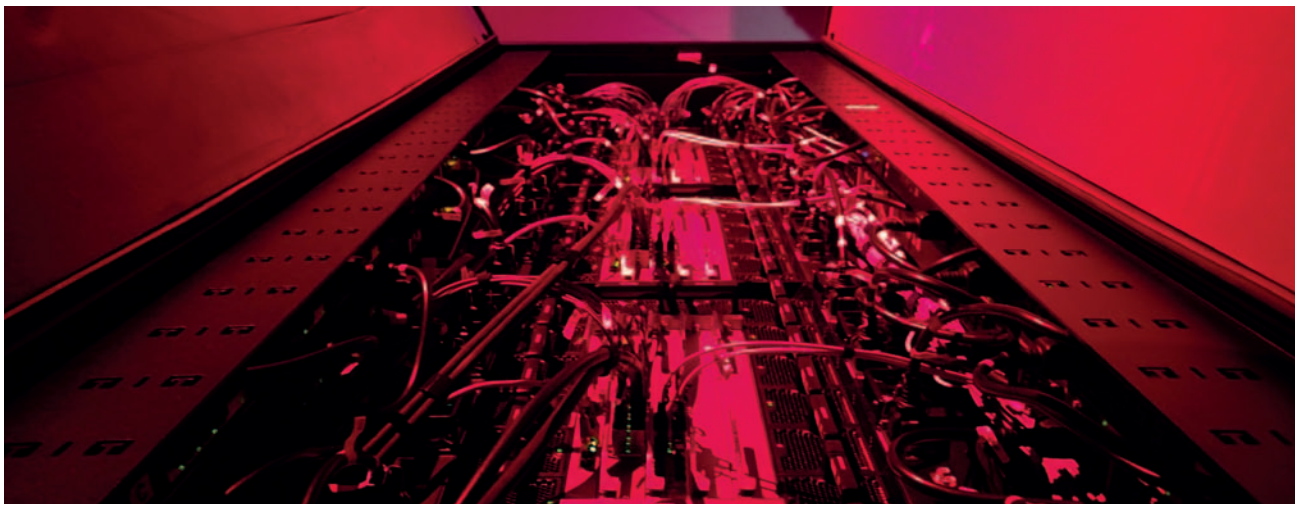
In 2021, the Ares supercomputer was launched in Cyfronet. It is built of computing servers with Intel Xeon Platinum processors, divided into three groups:

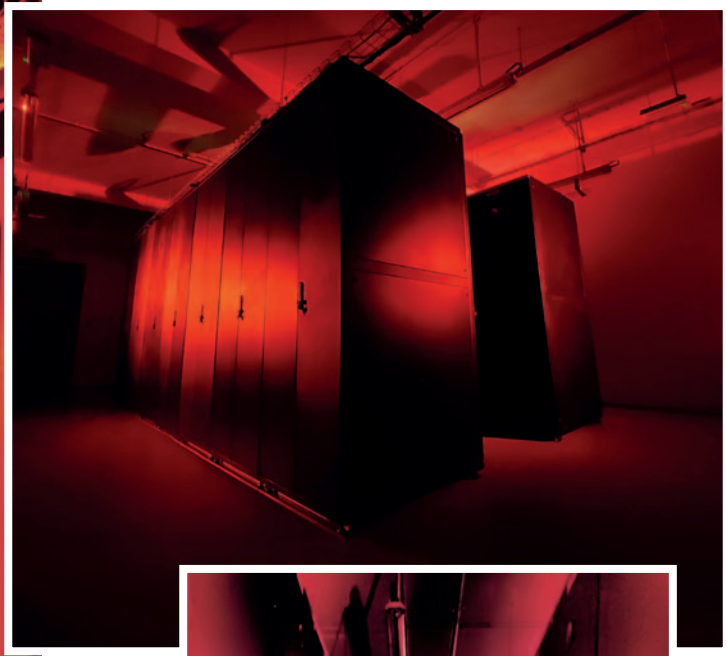
- 532 servers equipped with 192 GB of RAM,
- 256 servers with 384 GB of RAM,
- 9 servers, each with 8 NVIDIA Tesla V100 cards.

Ares in numbers	
Number of computing cores	37 824
RAM	147.7 TB
Number of GPGPUs	72
Computing power	4 PFlops
TOP500 – the list of the world's fastest computers (June 2022 edition)	290th position

The total theoretical performance of the CPU parts is over 3.5 PFlops, and of the GPU part is over 500 TFlops. Ares is also supported by a disk system with a capacity of over 11 PB. An InfiniBand EDR network is used for data transfer. The supercomputer has 37,824 computing cores and 147.7 TB of RAM. It is also equipped with a liquid cooling system.

Ares complements Cyfronet's computing resources by providing a newer generation of processors and servers with more memory. This will enable shortening the computation time of scientific tasks and addressing problems that so far could not be run on a large scale due to insufficient memory. In addition, placing Ares in a geographically different location than Prometheus guarantees the continuity of the provision of computing services.

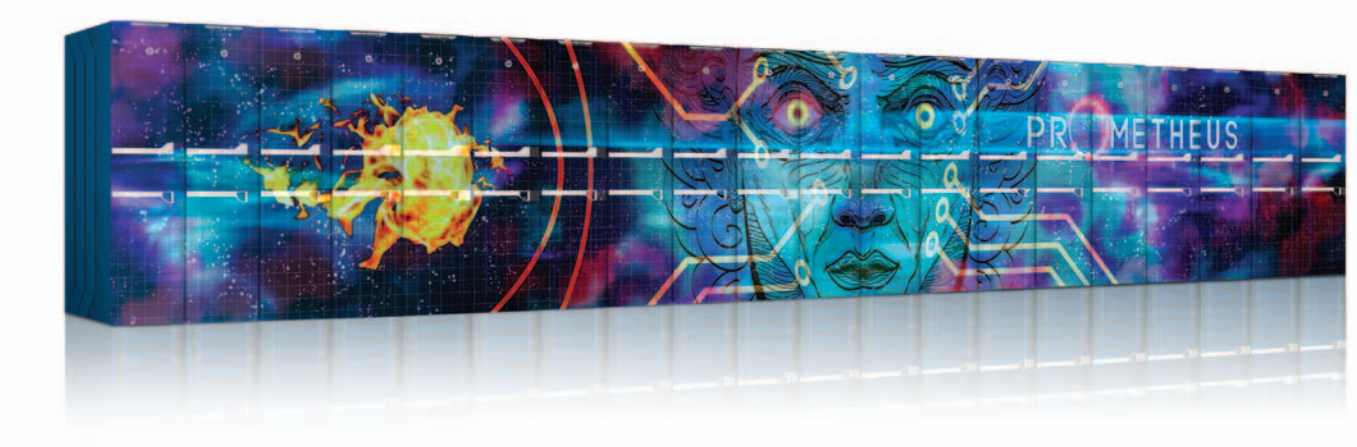




Prometheus – PetaFlops computing power

Changes in the world of science follow very quickly and affect the speed of development of IT facilities, which Cyfronet offers to scientists. Researchers' growing demands for computing power and data storage are clearly visible from the disciplines almost traditionally associated with high-performance computers: chemistry, physics, astronomy, life sciences and fields related to them. Astronomy, astrophysics and space physics are based on the one hand on data acquisition and analysis, and on the other on complex computer simulations. Biological, chemical and medical sciences as well as those mentioned above are characterized by rapid development and introduction of new, increasingly sophisticated research methods, e.g. molecular techniques based on high-performance DNA sequencing. Medicine, as a multidisciplinary field, deals with a number of time-consuming analyses, e.g., the human genome. It results in increased demand for automated collection, storage and analysis of biomedical signals and images, what in turn leads to necessity of use of the supercomputing resources in order to implement these processes. The possibility of linking together multiple unique data, i.e. the clinical, genetic as well as environmental and social data, brings many benefits, but also in this case the dedicated services are needed that can be offered only by supercomputing centers.

These are the tasks Prometheus – the most powerful Polish supercomputer – deals with. As the successor of Zeus, it has become a part of the PLGrid infrastructure and serves scientists, also within international research projects. Prometheus is used for: data results analysis, numerical simulations, (big) data processing, and advanced visualisations provision.

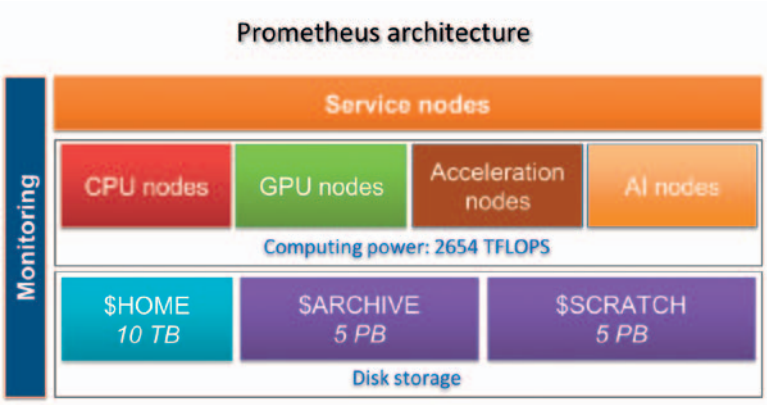


Prometheus consists of more than 2,239 servers based on the HP Apollo 8000 platform, combined with the super-fast InfiniBand FDR network with 56 Gbit/s capacity. Its energy saving and high-performance Intel Haswell and Intel Skylake processors offer 53,748 cores. These are accompanied by 283.5 TB of DDR4 RAM and by two storage file systems of 10 PB total capacity, and 180 GB/s access speed. Prometheus has also been equipped with 144 NVIDIA Tesla K40 XL and 32 NVIDIA Tesla V100 GPGPUs. The theoretical performance of Prometheus is 2.7 PFlops (PetaFlops)!

Due to the innovative technology of direct liquid cooling of processors and RAM modules, Prometheus is also one of the most energy-efficient computers in its class in the world. This was achieved by using the cooling water having a temperature of 28°C. To cool down the water to such a temperature in our climate it is enough to use cheap in use dry-coolers, instead

of ice water generators, consuming relatively large amounts of electricity. With use of water cooling, electronic components operate at temperatures lower than normal, what positively affects not only the failure, but also allows to reach efficiency more than 5% higher than for a similar installation based on the classic air cooling. Furthermore, liquid cooling allowed for extremely high installation density of 144 computing servers in one rack, therefore Prometheus, weighing of more than 40 tons, covers 18 m^2 area and is placed on 20 racks only. This also has a significant impact on internal data transmission, because distances of connections are critical here.

Prometheus in numbers	
Number of computing cores	53 748
RAM	283.5 TB
Number of GPGPUs	144
Computing power	2.7 PFlops
TOP500 – the list of the world's fastest computers (June 2022 edition)	475th position



Prometheus has been installed in a high-tech computing room, exclusively adapted for its operation. The supercomputer's proper functioning is additionally supported by the accompanying infrastructure, including such systems as guaranteed power supply with an additional generator, modern air-conditioning and gas extinguishing.

Prometheus once again has been listed on the TOP500 list of the world's fastest computers (June 2022 edition) and took the **475th position**, as one of five supercomputers from Poland.

Division into parts with diverse functionality, applied in the Zeus supercomputer, has been very well used by its users. Due to this fact, the Prometheus architecture is also a composite of several classes of nodes, varying in terms of architecture of computing resources and functionality:

- classical cluster of computing servers with highly efficient CPU nodes equipped with two Intel Xeon processors,
- cluster of servers equipped with graphic accelerators GPGPU NVIDIA Tesla K40 XL,
- acceleration partition with a set of devices supporting the Prometheus configuration with several types of accelerators (including GPGPU NVIDIA K80, Intel Xeon Phi 7120P, and Nallatech FPGA cards),
- a partition dedicated to calculations related to artificial intelligence, equipped with GPGPU NVIDIA Tesla V100 graphics accelerators. It is worth mentioning that **this partition is a system with computing power over 4 PFlops for tensor operations and 256 TFlops for standard calculations performed on double precision numbers, which makes it the fastest dedicated solution for artificial intelligence available for the needs of science in Poland.**

Thanks to Prometheus users have received more than seven times greater opportunities compared to the previously used Zeus. Much more efficient processors, faster network of internal connections, and a greater amount of memory of Prometheus enable to perform calculations on a scale impossible to achieve using previous Cyfronet's resources.

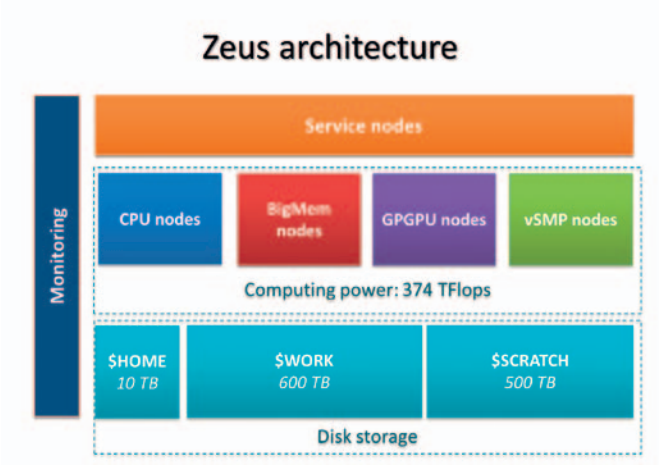
ZEUS – over 60 000 CPU-years

Cyfronet operates one of the fastest supercomputing systems in Poland, named *Zeus*. It currently provides 374 TFlops of theoretical performance, 25,468 CPU cores and over 200 GPGPUs. All this, equipped with 60 TB of RAM and 2.3 PB of disk storage supports the computations of scientific communities.

The Zeus supercomputer was launched in 2008. Since that time, it has been continually noted (12 times) on the TOP500 – the list of the world's fastest computers. Four of these locations were on **TOP100** subset, with **80 – the highest noted spot**. Zeus was 10 times the fastest in Poland.

The architecture

Zeus is a heterogeneous computing cluster. It constitutes of four classes of nodes, varying in terms of architecture of computing resources, specifically tailored to the requirements of the scientific communities. The Zeus architecture is a composite of four partitions:



- **classical cluster** of computing servers with highly efficient CPU nodes equipped with two Intel Xeon processors and 16-24 GB of memory per node,
- cluster of servers with large amount of memory – "**fat nodes**" with four AMD Opteron processors and 256 GB of memory per node,
- set of servers equipped with **GPGPU** accelerators (Intel Xeon processors as well as NVIDIA M2050 and NVIDIA M2090 cards) and **FPGA** accelerators (Pico Computing M-503 modules with Xilinx Virtex-6 LX240T),
- **"virtual" SMP** computer with large, shared memory, using vSMP software of the ScaleMP company – the nodes with Intel Xeon processors connected with a specialized virtual machine hypervisor, which allows for booting up the machines up to 768 cores and 6 TB of memory.

Diversification of the node types gives a possibility to fit users' applications to the hardware, which matches at best their characteristics and special requirements. For example, the classic CPU node group is dominated by serial and parallel (MPI) jobs, while the second one is great for large memory jobs. The GPU nodes allow some applications to benefit from GPGPU accelerators and the

vSMP nodes give a possibility to run huge memory jobs or scale applications, which do not use any inter-node communication library, like MPI, for parallelism. It is worth noting that Zeus-vSMP was the Europe's biggest installation of this type when launched!

The users

Since it was launched, the Zeus cluster has been serving the whole scientific community from Poland. In 2021 alone Zeus performed over **1.6 million tasks** serving scientists from many universities and research institutes. Majority of tasks were executed on more than one processor and this trend is constantly increasing. Typical user computations request even **several thousand of cores just for one job!**

Three supercomputers from Cyfronet together on the TOP500 list of the fastest supercomputers in the world

Supercomputers from Cyfronet on the TOP500 list

Zeus supercomputer

- 2010 - VI, 161st place, 55 TF
- 2010 - XI, 85th place, 105 TF
- 2011 - VI, **80th** place, 124 TF
- 2011 - XI, 88th place, 162 TF
- 2012 - VI, 89th place, 271 TF
- 2012 - XI, 106th place, 357 TF
- 2013 - VI, 114th place, 374 TF
- 2013 - XI, 146th place, 374 TF
- 2014 - VI, 176th place, 374 TF
- 2014 - XI, 211th place, 374 TF

Zeus and Prometheus supercomputers

- 2015 - VII, 269th and 49th place, 374 + 1659 TF
- 2015 - XI, 387th and **38th** place, 374 + 2399 TF

Prometheus supercomputer

- 2016 - VI, 49th place, 2399 TF
- 2016 - XI, 60th place, 2399 TF
- 2017 - VI, 72nd place, 2399 TF
- 2017 - XI, 78th place, 2399 TF
- 2018 - VI, 103rd place, 2399 TF
- 2018 - XI, 131st place, 2399 TF
- 2019 - VI, 174th place, 2399 TF
- 2019 - XI, 241st place, 2399 TF
- 2020 - VI, 288th place, 2399 TF
- 2020 - XI, 324th place, 2399 TF

Prometheus and Ares supercomputers

- 2021 - VI, 373rd and **216th** place, 2399 + 3510 TF
- 2021 - XI, 440th and 267th place, 2399 + 3510 TF

Prometheus, Ares and Athena supercomputers

- 2022 - VI, 475th, 290th and **105th** place, 2399 + 3510 + 7709 TF


TOP500 CERTIFICATE

The List.

Athena - FormatServer THOR ERG21, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100 SXM4 40 GB, Infiniband HDR

Cyfronet, Poland

is ranked

among the World's TOP500 Supercomputers

in the 59th TOP500 List published at the ISC22 Conference on June 01, 2022.

Congratulations from the TOP500 Editors

Erich Strohmaier NERSC/Berkeley Lab · Jack Dongarra University of Tennessee · Horst Simon NERSC/Berkeley Lab · Martin Meuer Prometeus

On May 30, 2022, during the ISC High-Performance conference in Hamburg, the latest TOP500 list of supercomputers with the highest computing power in the world was announced. **For the first time in history, the list simultaneously included three supercomputers from one Polish computing centre. These are those operating in Cyfronet: Athena (105th place), Ares (290th), and Prometheus (475th).**

The story of the machines installed in Cyfronet that were on the TOP500 list began in 1996 when SPP1200/XA-32 computer took 408th place. After a long time without records, the Zeus supercomputer appeared on the list in 2010 and remained there until 2015. From then on, Cyfronet has been represented by Prometheus, which was joined by Ares in June 2021 and by Athena in June 2022.

A list of places that Zeus, Prometheus, Ares and Athena supercomputers have taken on the TOP500 list in recent years is presented in the adjacent column.

The presence of Cyfronet supercomputers on the TOP500 list is an important confirmation of competences in the following areas: availability of resources, quality of services, users' trust, reliable operation and greater access to EU programs.

The world's top energy efficiency

All three Cyfronet's supercomputers that took places on the TOP500 list have also been ranked on the Green500 list of the most ecological supercomputers. The main criterion (energy efficiency) is calculated as the ratio of the number of floating-point operations per second (computing power of a supercomputer) to energy consumption: Gflops/W. Athena's 9th place on the Green500 list of the world's most energy-efficient supercomputers is a particular success. This position proves the excellent ratio of the provided computing power to the electricity consumption. Ares and Prometheus were also on the Green500 list. Ares took 83rd place, and Prometheus ranked 162nd position.


The GREEN 500 CERTIFICATE

Athena - FormatServer THOR ERG21, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100 SXM4 40 GB, Infiniband HDR

Cyfronet, Poland

is ranked

among the World's TOP500 Supercomputers

in the Green500 List published at the ISC22 Conference on June 01, 2022.

Congratulations from the Green500 Editors

Wu-chun Feng Virginia Tech

Kirk Cameron Virginia Tech

Supercomputers usage

Supercomputers in Cyfronet are part of the European cloud and grid infrastructure under the European Grid Infrastructure (EGI). At the same time, they are also important supercomputers in the PLGrid nationwide computing infrastructure – the platform for conducting *in silico* research and enabling calculations with use of high-performance computers, also within the cloud and grid architecture.

Via the PLGrid infrastructure scientists can get access to the supercomputers' resources. Dedicated computing environments, so-called domain grids, and specialised IT platforms enable conduction of increasingly complex research problems. The research portfolio carried out with the help of Cyfronet supercomputers is quite reach. It includes:

- planning multi-stage calculations with the use of quantum accelerators,
- testing the physicochemical properties of macromolecules,
- modeling of radical polymerization processes,
- research on new superconducting materials,
- numerical analyzes of the combustion process in the chamber of a hybrid rocket engine,
- prediction of gravitational waves using machine learning,
- microphysical processes in space plasma,
- machine learning in veterinary medicine,
- designing fluorescent probes used in medical imaging.

A wide range of research topics is evidence of constantly increasing number of scientists, who are aware of advantages of supercomputers. With their help one can get the final results of huge simulations many, many times faster, compared to the case of an ordinary, desktop computer. Supercomputers enable to significantly reduce time of computations that using a single computer would often take many years (in specific cases more than 150, 700 or even 1000 years). Here they may be usually performed within a few days. What is important, Cyfronet users can benefit from the professional support – starting from full documentation, through training, to individual consultations with experts.

In addition to individual scientists and small research groups, even international consortia carry out calculations from many different scientific disciplines with the help of supercomputers – of course with the participation of Polish scientists. Scientific computations do not include simulations only. Computing power is utilised by Polish researchers also within international scientific projects, including experiments like CTA, LOFAR, EPOS, Large Hadron Collider in CERN and the recently discovered gravitational waves in LIGO and VIRGO detectors.

Year	No. of jobs	CPU time in years
Zeus supercomputer		
2008	603 525	207
2009	2 227 804	876
2010	4 009 049	990
2011	7 557 817	5 052
2012	8 126 522	7 923
2013	7 932 978	11 016
2014	7 694 224	12 980
Zeus and Prometheus altogether		
2015	7 505 763	15 952
2016	7 748 677	24 653
2017	9 066 892	39 232
2018	8 342 686	42 436
2019	4 993 639	44 027
2020	5 696 919	41 761
Zeus, Prometheus and Ares altogether		
2021	5 549 582	43 409

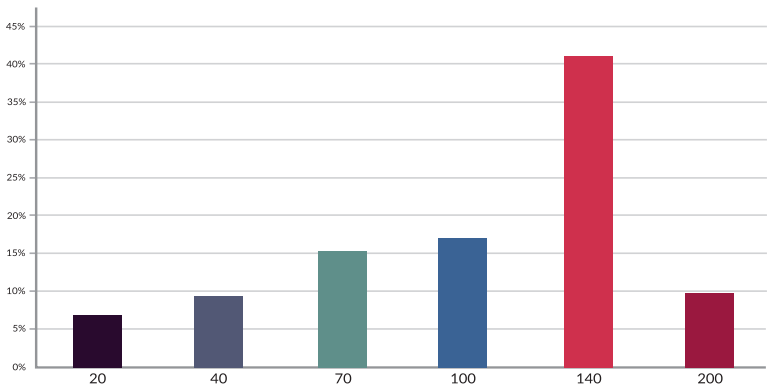
Obviously, even the highest positions in the TOP500 list, or the latest technologies used to build high-performance computers do not fully reflect the importance of this kind of computing resources for the Polish scientific community. The usefulness of supercomputers provided by ACC Cyfronet AGH as a tool for conducting research is best evidenced by statistical data on their use.

The table presents the aggregated key data on the number of computational tasks and their duration, performed by Cyfronet for other units.

It is worth mentioning that huge users' demands for computing power and space for data storage would not be fulfilled without continuous extension of computing resources and disk storage. Therefore, we carefully analyse users' suggestions and statistical data related to carried out computations together with world's trends in computing.

The scientific level of the tasks carried out with the use of the infrastructure provided by ACC Cyfronet AGH is very high. This is evidenced by the results of scientific and research works carried out in 2021 using this infrastructure, which were presented in many publications.

Ministry of Education and Science marks of articles published in 2021 by Cyfronet Users in scientific journals



Comprehensive infrastructure of efficient and safe storage of digital data

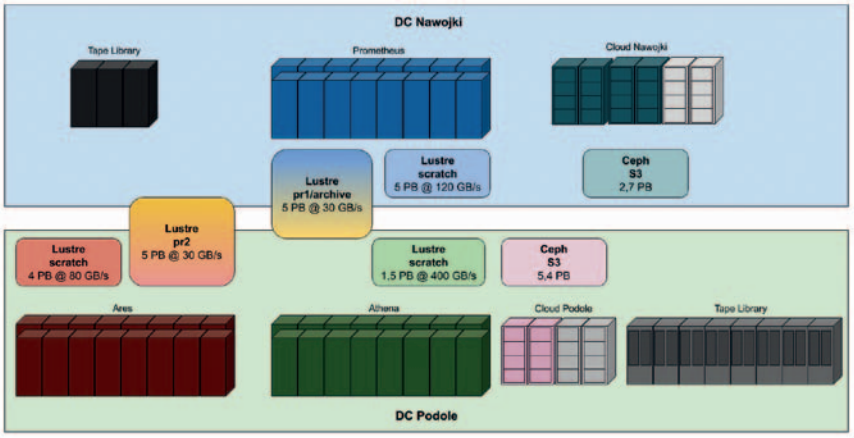
The currently observed phenomenon of the rapidly growing amount of digital information also applies to the scientific community. Access to very efficient supercomputers enables performing analyses of large-scale research problems, what results in generation of huge data sets. They require a completely new approach to information processing and storage. This problem, being currently one of the most important challenges of the modern digital world, is described by the concept of BigData. Also in ACC Cyfronet AGH there is clearly visible correlation between the growing expectations concerning available capacity, speed and additional functionalities of storage resources, caused by offering more efficient computing systems. The architecture of the Cyfronet Data Storage System, the main mass storage platform for High-Performance Computers, is composed of following elements:

- the SAN network – the efficient and highly available network dedicated to communication among devices within the Data Storage System, and clients using shared resources or services,
- disk arrays and servers of various types, offering the storage space for the users' data – starting from fast, but expensive and less capacious solutions, and ending with the devices with large storage capacity and relatively cheap, but with limited efficiency,
- service servers, with specialised tools and virtualisation software, providing users with functionalities such as automatic backup and archival, hierarchical data storage systems, high-performance hardware file platforms or distributed network file systems,
- tape libraries and specialised software used to store critical user data on magnetic media,
- additional infrastructure, including Ethernet, Infiniband as well as solutions supporting management of the IT infrastructure and enabling secure storage of magnetic media.

At present, the total storage capacity of Cyfronet disk and tape resources is **70 PB**.

Mass storage for supercomputers

The proper teaming of computing infrastructure with the right selection of storage solutions can assure the best quality of services provided to scientific users. The scale of problems in this area grows with the complexity and efficiency of the supercomputers used. Currently, data storage systems attached to Cyfronet supercomputers store billions of files up to terabytes. The broad thematic scope of research on the resources provided by Cyfronet is reflected in the variety of configurations of the Centre's key supercomputers and thus also in the structure of dedicated storage resources. The resources of the Cyfronet data storage system are located in two locations.



Supercomputers use, among others, efficient temporary space, the so-called scratch. The critical element here is the speed of operation, which is why it is based on a high-speed distributed file system architecture – Lustre. The advantage of Lustre is the ability to scale the capacity and efficiency of the disk space. By combining the capacity of multiple servers, I/O bandwidth is aggregated and scales with additional servers. Moreover, bandwidth and/or capacity can be easily increased by dynamically adding more servers without interrupting users' computations. Currently, all supercomputers in Cyfronet use the scratch space implemented by Lustre. In the case of Prometheus, this space has the capacity of 5 PB and the speed of 120 GB/s. Ares has the space with the total of 4 PB and the speed of 80 GB/s. In both of these computers, the scratch space is realized with the help of mechanical disks. In the case of Athena, user data is stored on solid-state drives. The use of this type of solution significantly increases the efficiency of the system. The planned capacity of this type of space for Athena is to be 1.5 PB and achieve the bandwidth of 400 GB/s.

Most of Cyfronet's disk memory resources are dedicated to the needs of users of domain services developed in the PLGrid program. The PLGrid infrastructure offers a dedicated workspace for groups using domain services – the functionality necessary to enable collaboration between scientists working in geographically dispersed locations. This functionality is implemented using the Lustre file system. The maximum capacity of the /pr1 resource in the Prometheus supercomputer is 5 PB, and the total speed of reading and writing operations reaches 30 GB/s. In the case of Ares and Athena, the /pr2 resource has the capacity of 5 PB and the speed of 30 GB/s.

The object-oriented data storage system is an additional resource for storing users and projects' data resources in Cyfronet. It is based on the CEPH software. The data in this system is available through the S3 protocol based on the REST API and is stored in globally unique containers (buckets) in which users store their data in the form of objects.

A particular case of mass storage are resources for large projects and international collaborations in which Cyfronet participates, such as WLCG (Worldwide LHC Computing Grid), analyzing data from the LHC detector at CERN, or CTA (Cherenkov Telescope Array), studying gamma radiation using a network of radio telescopes. These projects require substantial disk resources, often available using unusual protocols such as SRM, xroot, or GridFTP. Cyfronet provides this type of disk space using several instances of dedicated DPM software (Disk Pool Manager) and using dedicated networks such as LHCone. The total capacity of DPM systems in Cyfronet exceeds 2 PB.

Currently, the total available disk capacity used by ACC Cyfronet AGH is approximately **45 PB.**

Backup-archiving services in detail

ACC Cyfronet AGH provides its users with a wide portfolio of services related to securing information stored in a digital form. In addition to advanced technological solutions such as communication networks dedicated to storage systems, modern disk arrays or hardware file servers, the Centre also performs conventional backup-archiving services, based on magnetic media. Contrary to the expectations of the inevitable end of solutions using data storage on magnetic tapes, this technology is constantly evolving, and offers in successive generations not only the increasing capacity of the media, but also significantly better capacities and mechanisms supporting the safety and effectiveness of the information storage (e.g. data encrypting and compressing algorithms, which are embedded in the tape drives).

Cyfronet has currently two tape libraries having in total over 5 thousand slots for LTO magnetic tape drives and 25 drives of the V and VI generation. A single LTO-7 magnetic medium has a physical capacity of 6 TB and allows recording at the speed up to 300 MB/s, which theoretically allows the storage of almost 30 PB of uncompressed data in tape libraries. Described resources are used for performing current backup and archive of important information resources of the Centre's users.

Backup is performed on the active data – that might be currently in use – through a replication process from the source location to a separate, isolated destination. The ideal backup procedure ensures consistency of the source and backup data, both at the level of a single object (a file located on a hard drive), and in the case of complex IT systems, such as database or mail servers as well as virtual environments. Physically, the cloning process is usually done by copying the source data from the backup client disk to disk/tape resources of the target backup server, using dedicated or shared access medium, such as Ethernet or SAN. The purpose of an archive is to ensure security of unused data and to release occupied storage resources. In contrast to the backup, the archive is performed once, by the migration of the data from the source location to the destination.

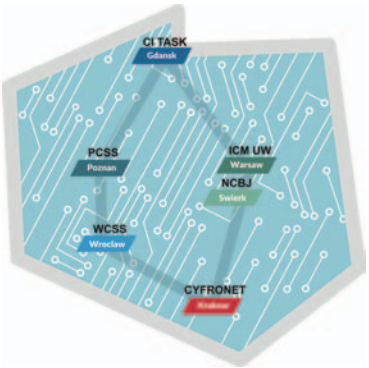
ACC Cyfronet AGH provides a wide range of backup services, addressed directly to users, and operating without their interaction. Among those at the disposal of users, there are ones based on FTP, NFS and SCP network protocols, acting within the dedicated backup servers. These machines provide backup solutions for users, allowing them to direct access to the backup data. It is up to users to decide which data they treat as a backup and which as archives.

For the special cases Cyfronet offers users a dedicated backup-archive service called the Universal Archiving. Within this service, the user is given a dedicated disk space, protected at many levels. User's data in this case is protected by a distributed disk array equipped with disk resources protected by RAID-6 level functionality and additionally secured by the HSM system of hierarchical data storage. To advance the data safety even more, users' backup data are additionally protected by geographical data replication to the associated units. Last but not least, users of Universal Archiving system can further increase their data safety by encrypting their data with use of certificates. At present, the total storage capacity of Cyfronet tape resources exceeds **25 PB**.

PLGrid Program

The infrastructure and the projects

The idea of the PLGrid Program has been invented by Cyfronet as a result of vast knowledge and experience gained in variety of national and EU projects. In 2007, it constituted formally as the PLGrid Consortium. At present, the Consortium consists of the following institutions: Academic Computer Centre in Gdansk, Interdisciplinary Centre for Mathematical and Computational Modelling in Warsaw, National Centre for Nuclear Research in Otwock-Świerk, Poznan Supercomputing and Networking Center, Wroclaw Centre for Networking and Supercomputing, and **Academic Computer Centre Cyfronet AGH as the initiator and leader of the PLGrid Program and Consortium**. The work carried out by Consortium partners led to the full-fledged distributed infrastructure for scientific computing. This infrastructure comprises not only high performance computing hardware, but also mass storage and dedicated tools for deployment of scientific applications on the available resources.



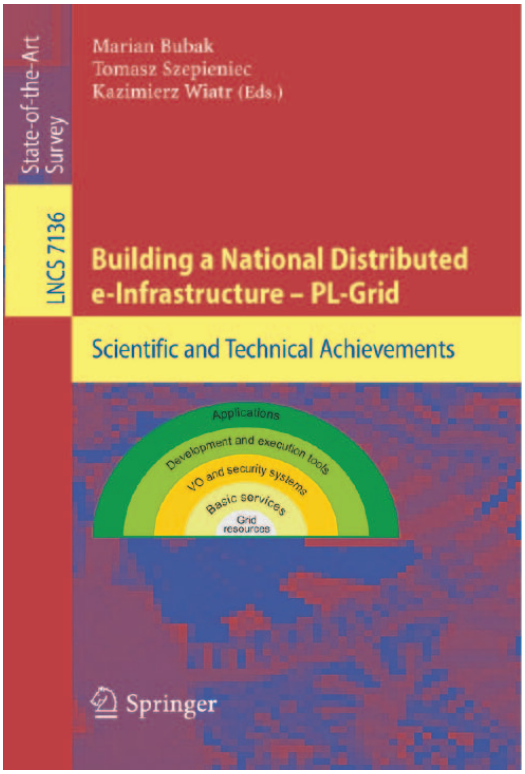
The design and construction of the PLGrid infrastructure started in the framework of the PL-Grid project (Polish Infrastructure for Supporting Computational Science in the European Research Space), in response to science needs, in which computers become more and more important. The main goal of the built infrastructure was to support scientists' investigations by integrating experimental data and results of advanced computer simulations carried out by geographically distributed research teams with use of supercomputers localised in High Performance Computing Centres. This aim was accomplished, among others, by extending the amount of computational resources in all PLGrid Consortium institutions. What is more, thanks to the PL-Grid project, in fall 2011 all Consortium partners have been spotted on TOP500 – the list of fastest world supercomputers. The same year Zeus supercomputer in Cyfronet has been located at 80th position – what gave it the first place among Polish supercomputers.

The next step of the PLGrid Program was to provide the researchers with necessary IT support through preparation of the specific computing environments, i.e., services and software as well as helping users in planning, running and analysing complex scientific experiments. Preparation of dedicated computing environments, so called domain grids, tailored to the needs of 13 different groups of scientists, was the most important task of PL-Grid follow-up – implemented within the PLGrid Plus project (Domain-oriented services and resources of Polish Infrastructure for Supporting Computational Science in European Research Space).

Adaptation of the infrastructure to the needs of scientists brought by domain grids was a great success of the PLGrid Plus project. Therefore, these activities have been further extended by the "New generation domain-specific services in the PL-Grid Infrastructure for Polish Science" project. In the PLGrid NG project, the domain-specific grids were developed for several other groups of scientists, representing fourteen research fields **(in total, in the two projects, IT support tools were built for 27 scientific disciplines)**.

However, the PLGrid Program did not stop on development of domain-oriented solutions only. Thanks to longstanding involvement in the development of grid computing infrastructures, Cyfronet is now recognized as a Centre of Excellence in the area of cloud and grid services – an achievement reflected by the new large-scale scientific grant named Distributed Computer and Data Infrastructure Centre of Excellence – PLGrid Core. This grant represented the next step in the development of the PLGrid Program and extension of the infrastructure towards Cloud Computing and handling big data calculations. It aimed not only at extension of hardware and software portfolio, but also dedicated accompanying facilities. One of them – a new backup Data Center built in separate geographical location highly increased security of scientific data sets.

It is worth noting that on the November 2015 edition of TOP500 the **Prometheus supercomputer, deployed at Cyfronet in 2015 in the framework of PLGrid Core, took the 38th position, the highest so far for supercomputers deployed in Poland!**

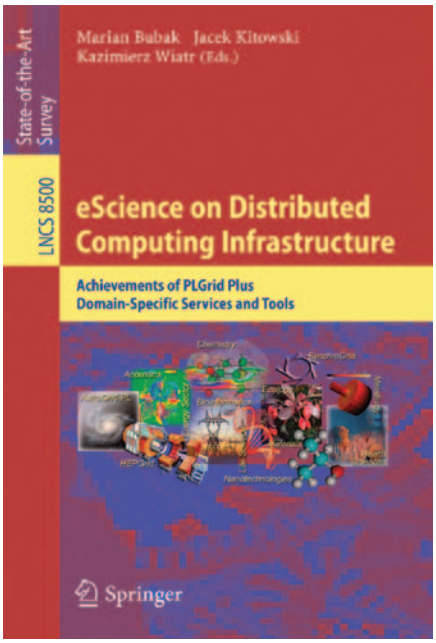


At present, enormous computing power and mass storage resources are available within the infrastructure. In addition, many tools supporting organization of computational experiments, designing and running applications, computationally supporting research and results' visualization were implemented in the infrastructure. Furthermore, the Consortium introduced a new service – Cloud Computing.

All the projects of the PLGrid Program have been co-funded by the European Regional Development Fund as part of the Innovative Economy program. ACC Cyfronet AGH has the honour to be their responsible coordinator. Vast range of services contributes to increase of cooperation between Polish scientists and international groups of specialists from many different scientific domains – also humanities and social sciences. The essential fact is that anyone who is performing scientific research can be the user of the infrastructure. Access to huge computational power, large storage resources and sophisticated services on a global level is free to Polish researchers and all those engaged in scientific activities associated with the university or research institute in Poland. All one has to do is to create an account via the PLGrid Portal.

Domain-specific grids in the PLGrid infrastructure

The PLGrid infrastructure, established through the Cyfronet's initiative, offers a uniform access to resources of all five Polish High-Performance Computing centres. Unification takes place at many levels, ranging from a user's single login and password across the infrastructure, to the access to scientific applications. Sometimes, however, the use of modern computing systems, services and tools of the e-infrastructure becomes relatively difficult for researchers. Basic infrastructure services are often insufficient to conduct scientific research, particularly in the context of large international consortia. In such situations, users need both assistance and close collaboration with service providers.



Therefore, within the PLGrid Plus project (2011-2015), the PLGrid infrastructure has been extended with specific environments, solutions and services, developed according to the identified needs of 13 pilot groups of scientists. The main aim of the project was to lower the barriers required for researchers to use the infrastructure, and, thus, attract new communities of users, who need the computational power and large disk space of supercomputers, but have no or little skills in using it. To enable and facilitate development of domain--specific environments, the project relied on a broad cooperation with representatives of various disciplines, often grouped in domain consortia.

The dedicated services hide the complexity of the underlying infrastructure and, at the same time, expose the actual functions that are important to researchers of the given domain. In this way, users are provided with exactly the functionality they need. What is more, it is exposed to them in their domain-specific manner to achieve maximum intuitiveness and usefulness.

Scientific and technical achievements of PLGrid Plus were presented in a book published in the Springer Publisher, in September 2014. The book is an important source of information for researchers, developers and system administrators, who use grid and cloud environments in their research. The book contains 36 chapters and is divided into three parts: the first one (chapters 1 to 8) provides a general overview of the work carried out in the project and a description of the current state of the PLGrid infrastructure, including new solutions in the field of security and middleware.

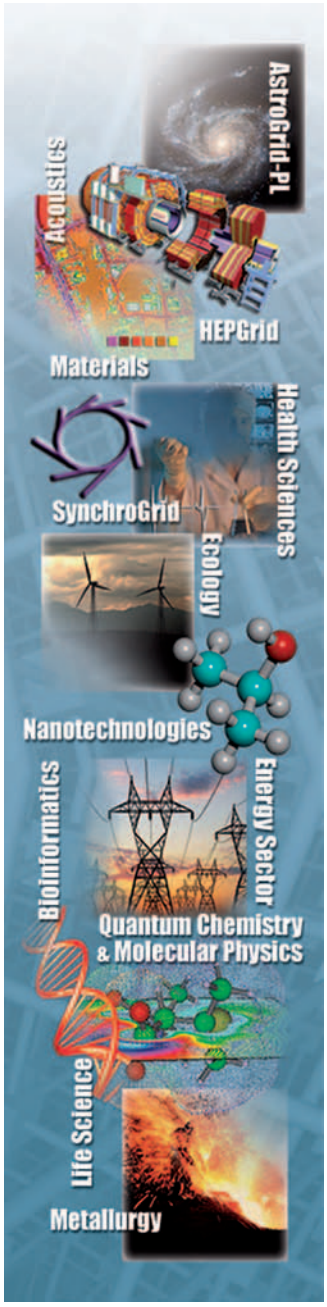
The second part (chapters 9 to 13) presents new environments and IT services that can be used by all of the previously mentioned groups of scientists. The third part (chapters 14 to 36) describes how specific environments, tools and services, prepared within the PLGrid Plus project, are used in advanced computations and computer simulations performed by different groups of researchers. These chapters present computational models, new algorithms and methods of their implementation using available tools and services.

Success of the PLGrid Plus project, in particular, the growing popularity of specialized tools and platforms prepared for the members of the first 13 strategic areas of science, led to a rapid increase in demand for related services to researchers in other fields. Therefore, the PLGrid Consortium launched the PLGrid NG project (2014-2015), whose primary objective was to implement, within the PLGrid infrastructure, several additional computing services for groups of scientists representing 14 new research fields.

New domain-specific services covered a wide range of activities: including provision of the specialized software, mechanisms of data storage and modern platforms integrated with a new type of tools and dedicated databases, which sped up research conduction as well as streamlined and automated the work of research groups.

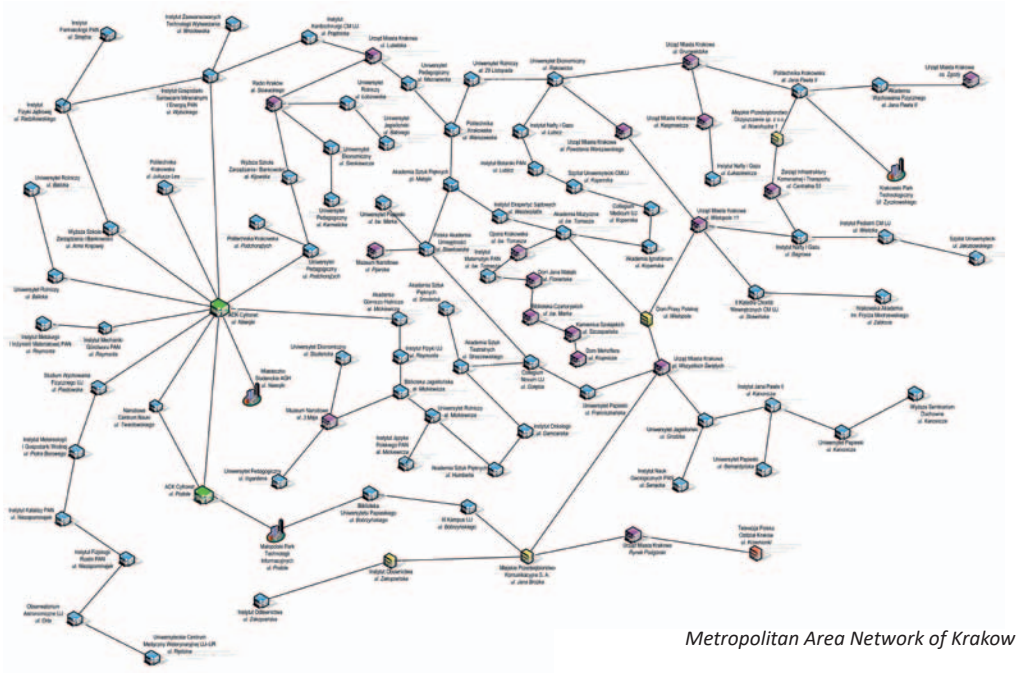
Preparation and implementation of a set of domain-specific services fit very well with the need of development of an advanced IT infrastructure designed for the implementation of modern scientific research. The well-tailored PLGrid e-infrastructure does not only fulfil researchers' needs for suitable computational resources and services, but also enables Polish scientific units collaboration with international research organizations.

Expansion of the existing computational infrastructure towards domain-specific solutions for research teams allowed more effective research conduction.



Metropolitan Area Network

One of the major characteristics of the present science is complexity of research challenges, including their multidisciplinary character, use of heterogeneous models, resources and massive amount of data produced by a variety of sources. Research is not performed by a small group of scientists anymore, but by international consortia. In order to bind those usually geographically distributed resources together, fast and reliable network connectivity is essential. Therefore, one of the principal tasks of the ACC Cyfronet AGH is development and maintenance of the Metropolitan Area Network (MAN) to achieve its availability 24/7.



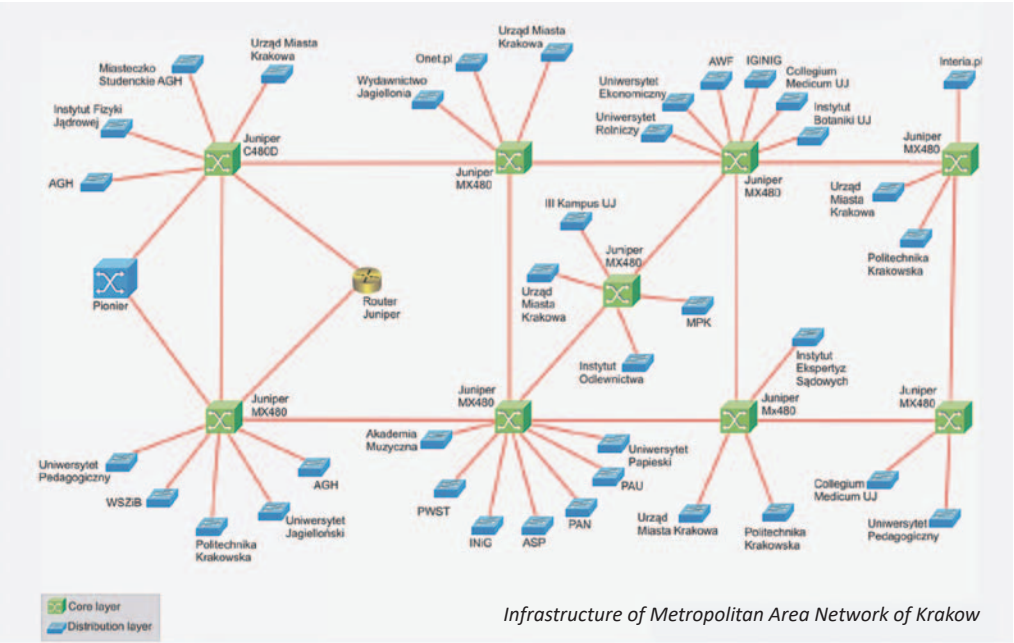
Metropolitan Area Network of Krakow

Main characteristics of MAN

It is not possible to attain high network availability without its constant development and adjustment to the needs of users. The length of dedicated fiber-optic links reached this year over 200 km. The core links of the network are located in the Old Town area and reach the academic campus of AGH University of Science and Technology. Furthermore, the network covers also Bronowice, Krowodrza, Czyżyny and Nowa Huta zones. Recent expansion of the network included such distant research centres like Prokocim, Borek Fałęcki and the 3rd campus of the Jagiellonian University in Pychowice. Development of the core backbone includes also other directions, up to the borders of Kraków.

The fiber-optic infrastructure is the basis of the MAN operation. ACC Cyfronet AGH takes efforts to include in it the largest possible number of university facilities and research institutions. At the same time, due to the ever-growing role of modern communication means, in everyday work it is very important that fiber-optic infrastructure, in addition to high bandwidth, could also ensure secure communication. It is realised through the use of backup links, which allow to maintain the continuity of operation in situations when primary routes are broken.

The core data link layers are implemented using top quality equipment with 1 and 10 Gb Ethernet technologies, while 100 Gb interfaces are gradually being introduced. Each of the backbone networks switches is connected with at least two and sometimes even three neighbours for automatic and transparent recovery in case of a failure of any network device or link. Our users can obtain fiber-optic connectivity to the network via 10/100/1000 Mbps or 1 Gbps Ethernet cables as well as through traditional modem uplinks.



Infrastructure of Metropolitan Area Network of Krakow

The Metropolitan Area Network is directly connected to Warsaw, Katowice, Bielsko-Biała and Rzeszów through the PIONIER network. Currently the links can serve up to 2x10 Gbps capacity. High Performance Computing centres in Poland (Gdańsk, Kraków, Poznań, Warsaw and Wrocław) are integrated with links of 2x100 Gbps capacity. The PIONIER network enables also communication with major national and foreign computing centres. International connectivity is achieved through the GEANT scientific network with 100 Mbps capacity. In addition, the reserve connection with 5 Gbps capacity is established to the Telia Carier Poland and Lumen Technologies networks.

Network services provided to the users

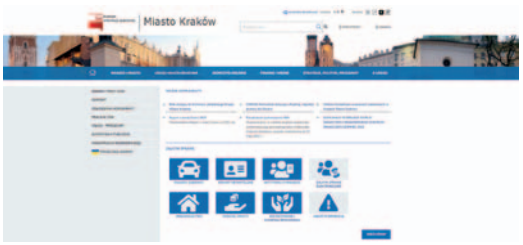
From the beginning of the Polish Internet (mid 1991) ACC Cyfronet AGH has been actively participating in the development of the telecommunications infrastructure and, what is very important, the wide range of Web services. Those include:

- **e-mail** accessed via SMTP protocol or web interface *http://poczta.cyfronet.pl*,
- **www**: CYFRONET operates a set of web sites, which in addition to news from the world of science, present information on the culture, sights and many other fields,
- **news**: discussion groups covering all areas of interest – from highly specialized scientific to general-purpose boards,
- **dns**: domain name system servers – performing translations of network domain names to IP addresses for users of the Krakow MAN,
- **ftp**: CYFRONET mirrors major international software archives, providing shareware and freeware applications for MS Windows and UNIX systems. The establishment of this service has significantly reduced the traffic on CYFRONET's international links while at the same time enabling faster downloads of software for users of the Krakow MAN,
- **eduroam**: provides the academic network access at all locations on eduroam on the world with a single authorized account, providing at all locations the same way as access to the network at the parent unit,
- **box**: a network drive (*http://box.cyfronet.pl*) allowing file exchange and synchronisation. The drive can be also accessed from mobile devices via dedicated application.

Network services in numbers in 2021	
Number of e-mails	> 17 000 000
Number of e-mail server sesions	> 50 000 000

Portals and mobile applications

The Centre does not limit its activities to the scientific areas only – it also contributes to the development of the information society. The Web server at ACC Cyfronet AGH serves as an Internet hub for the entire Kraków scientific community. The Centre continues to develop and extend its Web portal, which has gained substantial popularity over the years. In 2021, a new version of the portal was published, including a version for mobile devices.

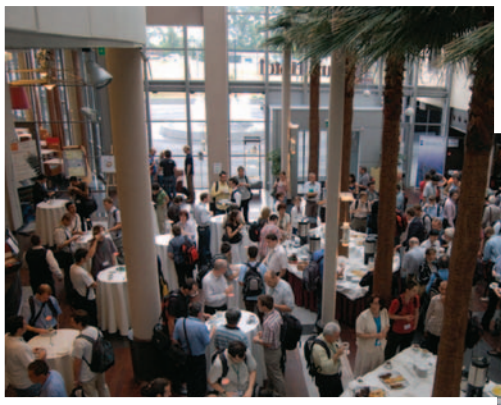


Cooperation with Kraków authorities is of particular importance for the Centre. The agreement between the Municipality of Kraków and CYFRONET, regarding the promotion of the City has resulted in the creation of an up-to-date portal. Aside scientific information the portal introduces its readers to the culture, historic sites, tourism, local transit and many other aspects of life in Kraków.

In collaboration with the City Hall, the Centre has been developing and running the Internet Bulletin for Public Information in the Kraków Region. In 2005 this collaboration was extended in order to provide content services for municipal units, libraries, schools, etc.

In 2007, the "Magical Kraków" web portal – *www.krakow.pl* has been nominated for the World Summit Award as the best e-Government service in Poland. The mobile version of the portal was awarded at the conference Mobile Trends, Mobile in 2012 as the best city mobile web site in Poland.

Cooperation with the City Hall explores also the area of mobile devices. CYFRONET has developed – among others – a mobile application "Kraków.pl". The app can be used as a Kraków city guide, a source of important information like phone numbers, info points, consulates or pharmacies. The most important part of this app is the ability to check all those places on an offline map. Our app is available in few languages.




www.CYFRONET.kra

The KSAF archive

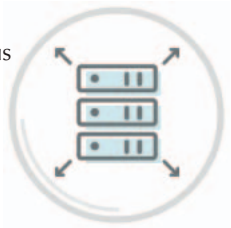
The KSAF archive

Computational resources

ACC Cyfronet AGH provides a mature computing infrastructure for Polish science based on five main pillars. Furthermore, complex support and training are available for the users.

Computational resources

Athena, Ares and Prometheus supercomputers provide:
15 PFlops
100 000+ cores
600 GPGPUs
950+ TB RAM.



Storage

45 PB of disk and 25 PB of tape storage space and fast scratch Lustre filesystems enable big data processing and analyses.

Scientific software

Vast portfolio of tools, libraries and scientific applications for research in various fields of science.

Computational cloud

Cyfronet's PaaS based on OpenStack provides elastic solution for computational environment which can be easily adapted to researchers' needs.

Tools for scientific collaboration

Tools and services such as Bitbucket Git repositories server and JIRA issue & project tracking solution ease scientific projects coordination and communication between researchers.

Advanced computing platforms and domain-specific services

Among the scientists conducting research with use of high-performance computers and large storage resources there is a need for different types of interaction with a computer or with the infrastructure. To address these needs Cyfronet provides a number of advanced IT platforms and dedicated services that hide the complexity of the underlying IT infrastructure and, at the same time, provide the functionalities important from the point of view of scientists from the particular field, precisely tailored to their needs.

Together with computing infrastructure we provide a selection of tools, which enable researchers to perform complex, large-scale experiments and manage their results in an easy way. The efficiency of the performed analyses and the safety of their associated data are guaranteed by appropriate IT solutions, benefitting from the extensive experience of Cyfronet's developers. The platforms have been successfully applied in the PLGrid Program for domain specific grids. As we mentioned before we have prepared more than 70 tools, platforms and services gathered into 27 scientific domains dedicated for important scientific topics and strategic fields of Polish science. All those services are provisioned in the framework of the PLGrid infrastructure, allowing Polish scientists and their foreign collaborators to access it in a convenient manner.

Among others, at the Centre we offer advanced tools and graphical interfaces that enable construction of dedicated environments for scientific research, building application portals, conducting virtual experiments, visualization of calculations' results, executing complex scenarios with parallel tasks, as well as supporting uniform and efficient access to data. All of these services are important support for researchers, as they have an impact on improving and, where possible, automating the work of research groups, what greatly accelerates obtaining research results. On subsequent pages we will learn about capabilities of selected services.

Invitation to cooperate

We are looking for people interested in development of domain-specific services. We also offer support in scientific research.

We encourage scientists to send us their program codes for the compilation by the experts at the Centre. After installation, we provide assistance in their effective use. We also enable the use of scientific software licenses held by research groups.

The Cloud Computing in PLGrid

The PLGrid infrastructure has been designed with particular focus on scientists and their needs. Its character allows for easy adaptation to, even sophisticated, research challenges performed by different groups of scientists – from small research teams up to international consortia of researchers. To fulfil their requirements, in addition to typical computing and storage platforms, we offer the PLGrid Cloud Computing Platform.

– *Up to now, the PLGrid infrastructure has been providing a set of well-defined environments with computing and storage resources. The cloud platform is not just an extension of them. We foresee it as a new quality level of conducting research* – says Kazimierz Wiatr, the Centre's Director. – *A user can easily connect to a requested set of virtual machines (VM), with full access rights to the operating system. To achieve high security, all the VMs operate in a dedicated, local area network. Particular services can be accessed from all over the world, easing cooperation between scientists* – adds Director.

There are several advantages of the cloud computing we would focus in particular:

- The Cloud increases elasticity of research, as scientists can tune the virtual machines to their specific needs. Up to now, to set-up a "virtual laboratory" solving some specific scientific problem, some help of PLGrid experts was needed. Now, each scientist can create and easily extend such virtual laboratory alone.
- The catalogue of VMs offered by PLGrid contains many OSes. Thanks to this, users can run their software applications with Operating Systems other than Scientific Linux, including Windows or other Linux OSes.
- With Cloud, it is easy to build and put in operation a test environment. This feature is very convenient for scientists developing their own software. Any test task can be then easily performed and its results analysed.
- It is possible to maintain a communication with already executed computing job. In addition, every virtual machine can be easily duplicated, even in thousands of copies or more. A start of a new VM takes just around 30 seconds.
- The Cloud platform is also the best and in many cases the only solution for running jobs with legacy software packages. In a secure LAN environment even old, deprecated operating systems can be used. This feature is also a solution for dispersed international groups using variety of different packages for their research. Every group can run their own computations and easily share their results with others.

– *The Cloud Computing in PLGrid and Cyfronet is an innovative solution on a European scale. We have a strong belief it will bring a new quality level for research conducted by our users* – concludes Prof. Wiatr.

Currently **200+ various types of VMs** are utilised on Cyfronet resources.

ONEDATA

Onedata is a global data management system, which provides transparent access to data stored on distributed storage resource managed by multiple providers. Onedata can scale to meet the needs of small user communities or large federations of users and storage providers, making it a perfect solution for large research initiatives, long-tail of science as well as for commercial purposes. Onedata allows users to rely on a single solution for managing their personal as well as research data sets and access them efficiently on any machine, from personal laptop as well as from a Cloud virtual machine.

Onedata provides a unique federation system based on zones, which enables storage providers to organize into trusted federations and allows users to easily request storage resources from providers within a zone.

Features for users

- Unified access to data stored on heterogeneous storage systems distributed across the infrastructure. With Onedata, users can access their data from anywhere, as the system automatically replicates and transfers necessary blocks on demand.
- All data is organized into *space*, which can be regarded as virtual folders or volumes, accessible from any client machine via POSIX protocol.
- Easy to use web based Graphical User Interface for data access, discovery and management.
- Support for easy data sharing and collaboration with other users, while ensuring security through custom Access Control Lists and creation and management of user groups.

- Open data publishing functionality integrated into the user interface, enabling publication of prepared datasets, registration of DOI identifiers and indexing in open access portals.

Features for administrators

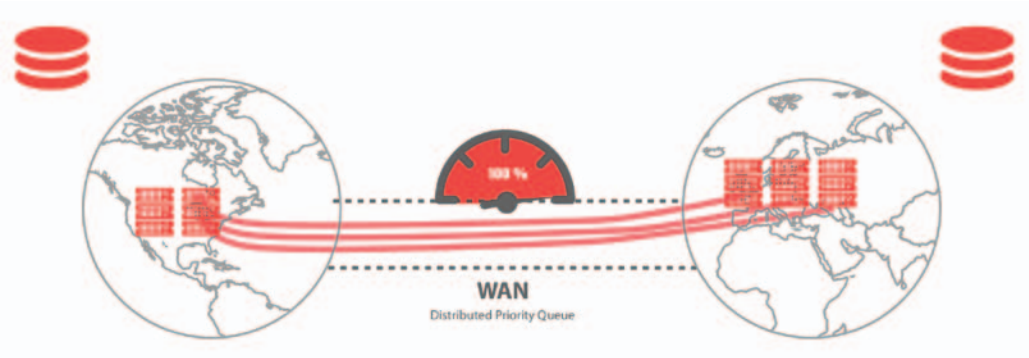
- Simple deployment based on Docker containers using a friendly command line client.
- Easy storage support for user requests based on secure tokens.
- Complex monitoring information available on all aspects of the system, accessible through REST API or directly visualized in the administration panel of the Graphical User Interface.
- Support for multiple storage backends including POSIX based storage (e.g. Lustre), Amazon S3, Ceph, OpenStack SWIFT, and GlusterFS.

Features for developers

- Easy integration with Onedata services using REST API and CDMI protocols.
- Flexible authentication and authorization of requests based on Macaroon tokens.
- Complete reference documentation of the REST API including sample clients for several programming environments.

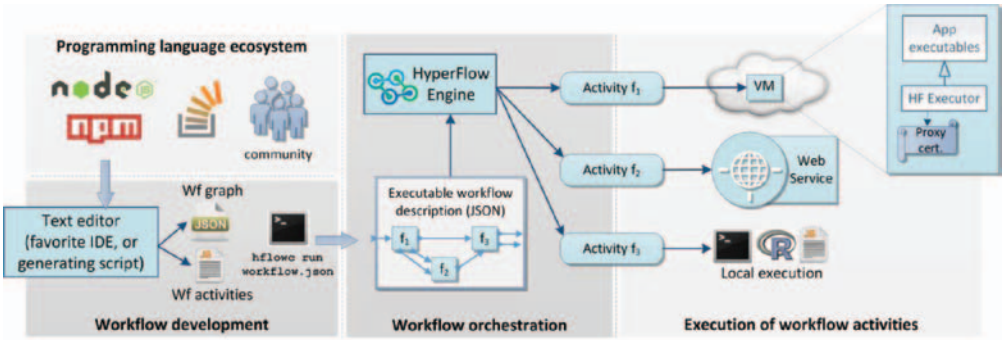
Onedata users

Onedata is currently deployed and evaluated in several initiatives in Europe including Polish National Grid infrastructure PLGrid, INDIGO-DataCloud, EGI DataHub, Human Brain Project and Helix Nebula Science Cloud. In HBP it has proven to meet the users' hard requirements of real-time brain visualization use case.



More information: *https://onedata.org*

HyperFlow is a lightweight tool that enables orchestration of scientific applications into complex pipelines or *scientific workflows*. HyperFlow aids users in composing their applications into workflows, deploying them in the cloud, and executing them.



Workflow programming

A workflow in HyperFlow is described as a graph of its activities (called *processes*) using a simple JSON-based data structure. Workflow activities perform the actual scientific procedures – steps in the scientific pipeline. In HyperFlow, workflow activities can either be implemented in JavaScript or mapped to executable programs. The JavaScript code is executed by the HyperFlow engine in the context of the `Node.js` runtime. An experienced workflow developer can thus take advantage of a mainstream programming ecosystem – large community, advanced tools, thousands of libraries and other resources – instead of using a proprietary development environment. Consequently, workflow activities can easily be programmed to invoke external Web Services, or execute local commands as part of the scientific pipeline defined by the workflow.

In the second option, the workflow developer can choose not to implement any JavaScript code, only associate each workflow activity with a previously prepared Virtual Machine image where appropriate programs are installed, and specify commands that are to be executed when a given workflow activity is triggered.

The availability of these two programming approaches makes HyperFlow equally suitable for experienced programmers / software engineers who desire low-level programming capabilities and high productivity, and domain scientists who are not experts in IT technologies and only wish to construct scientific pipelines out of existing modules.

Workflow deployment

HyperFlow automates workflow deployment in the cloud. The user only needs to prepare a configuration file specifying the mapping of workflow activities onto available Virtual Machine images, while the HyperFlow tool takes care of the rest. The user invokes a simple command *hflowc setup* which results in creation of appropriate VM instances in the cloud. These VM instances contain the workflow runtime environment and the scientific applications invoked from the workflow.

Workflow execution

After the workflow instance has been created in the cloud, the user executes the workflow simply by invoking *hflowc run <workflow_directory>*. Every workflow runs with its own instance of the HyperFlow runtime environment. Consequently, different workflow runs are isolated from each other which increases security and reliability.

The HyperFlow cloud runtime environment (called *HyperFlow Executor*) automatically takes care of transferring input data from the user directory to Virtual Machine instances, invokes the application executables and uploads output data back to the user directory. A variety of data transfer options are available, including a network file system, secure gridftp, and Amazon S3.

Applications

HyperFlow has become a part of several larger systems where it has been used for a number of applications. In the PLGrid infrastructure, HyperFlow serves as a workflow management system that enables the users to run scientific workflows in the cloud. An example application is a workflow-based solver for finite element meshes which can be applied to diverse problems. HyperFlow is also being integrated with the PaaSage middleware (*http://www.paasage.eu*) as an execution engine for scientific applications deployed in a multi-cloud environment. In the ISMOP project (*http://www.ismop.edu.pl*), HyperFlow is a component of a flood decision support system used to orchestrate flood threat assessment workflows. Hyperflow will also be used in parametric computing and workflow processing, as a replacement for Scalarm technology.

Contact

HyperFlow is developed and maintained by the DICE team (*http://dice.cyfronet.pl*). Please feel free to contact us in case of any questions or suggestions.

HyperFlow source code and manuals are available at *https://github.com/hyperflow-wms/hyperflow*.

Model Execution Environment

The Model Execution Environment (MEE) is a software stack which facilitates the execution of computational workflows on high-performance computing infrastructures, including those available at ACC Cyfronet AGH. The platform's goal is to ensure that computations can be executed straightforwardly by domain scientists, i.e. researchers who do not possess intimate knowledge of the specifics of interaction with computing clusters and other large-scale computing systems.

Pipelines, steps, and models

Within MEE, computational workflows are represented by the so-called pipelines, i.e., collections of computations (each of which is called a step) where the outcome of one computation provides input for another. MEE provides a wide range of facilities enabling users to design steps, arrange them into pipelines, and execute these pipelines on the available computational resources.

Each step is based upon a collection of computational artifacts (executable code) stored in a GitLab repository. This is referred to as the model. When a step is called for execution, MEE automatically uploads the requested model to the HPC infrastructure, and monitors its performance on the input data provided. Users are able to select a specific version of the given model when launching the pipeline (based on Git versioning mechanisms); thus facilitating traceability and repeatability of computations.

In addition, pipelines can be executed in either automatic or manual mode. An automated pipeline will be executed in its entirety, while a manual pipeline contains a breakpoint at the end of each step, asking the user to manually request the processing of any subsequent steps. This enables users to

download and review interim results and potentially cancel the execution of pipelines which are not expected to yield useful output, thus preserving computational resources.

All MEE features can be accessed via a user-friendly web-based UI. Furthermore, MEE provides programmatic access, which enables integration with higher-level software tools.

Research data management

As MEE schedules and monitors the execution of computational pipelines on HPC resources, care must be taken to manage the associated research data, ensuring that the appropriate input is made available to the underlying models, and that results can be retrieved from the HPC infrastructure. To this end, MEE provides a set of data management interfaces where users of the infrastructure can upload input files and download results. The platform itself manages HPC data storage resources and provides automatic stage-in and stage-out capabilities for research data, along with a set of top-level UI interfaces for its users.

Security

The Model Execution Environment is integrated with PLGrid authentication and authorization mechanisms. All users of the PLGrid infrastructure can use their login to authenticate themselves with MEE, and subsequently schedule and run computations using their PLGrid accounts. Moreover, MEE makes use of PLGrid computational grants assigned to researchers. Pending computations are executed in the context of specific computational grants, which can be predefined within the platform.

Organizations

Externally, MEE provides a set of distinct workspaces dedicated to individual research teams. These are referred to as organizations. Each organization has a distinct entry point to MEE (i.e., a distinct URL) and can define its own pipelines and pipeline steps, as well as manage its own set of research data. MEE implements compartmentalization, where each organization can be managed separately, providing access to a distinct group of users.

Applications

To-date, the Model Execution Environment has been applied in multiple research projects and provides services to various research groups. The list includes the EurValve, PRIMAGE, and InSilicoWorld European projects, the POLVAS consortium, and the Sano Centre for Computational Medicine, along with several ad-hoc research collaborations for which individual MEE organizations have been defined.

https://mee.cyfronet.pl

photo: Adam Frączek, KSAF AGH

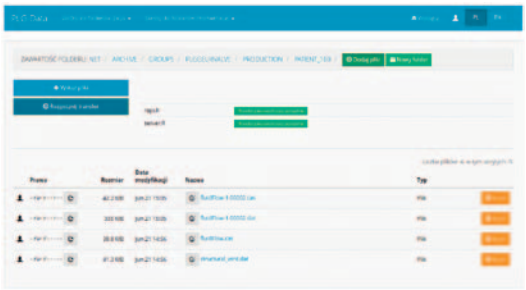

photo: Krzysztof Mastalski, KSAF AGH

Simple tool for file management on a computing cluster

PLG-Data

PLG-Data is a tool for management of data stored in the PLGrid infrastructure. It comes with a user-friendly web interface, and allows to upload, download, browse, delete and rename files and folders. It also helps with management of access rights for members of a research group, or external collaborators. It is currently integrated with the Prometheus cluster, and in the future it most probably will be integrated with new Cyfronet supercomputers.

The set of functionalities built in the tool includes, among others, the following:



- downloading files from the cluster to disk,
- adding new files and folders, and removing existing ones,
- renaming files or folders, and changing access rights to them,
- quick navigation to home, scratch and group folders, through a handy pull down menu,
- easy preview of image files without downloading them to a local disk.

Thanks to a specific construction of the URL address to particular files, the tool enables easy sharing of file location with other people, for instance through copying the browser address bar's content to an e-mail message or an IM communicator. The receiving person will be able to download a file, or view the contents of a folder, with one click – as long as that person is allowed access to the specific resource.

The service is secured with encrypted HTTPS protocol (between the user's computer and the PLG-Data portal) and the specialised GridFTP protocol (between the service's portal and the computing cluster). Application of such techniques allows the user to manage their files in a secure way. A person who uses PLG-Data does not receive any additional rights of access to files stored in the computing cluster, apart from the rights that the person already has.

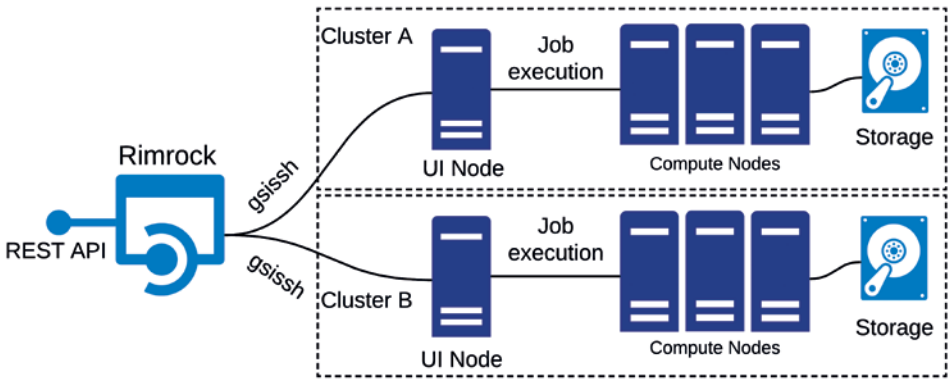
Logging in to the tool is done using either the PLGrid user-password pair, or a p12 certificate installed in one's browser. The service is available either in Polish or English. An advanced programmer's interface (API) helps developers to integrate their platforms, tools and services with the file storage inside the PLGrid infrastructure.

PLG-Data service address is: *https://data.plgrid.pl*

Rimrock, one of the services of the PLGrid infrastructure, enables management of scientific computation and result handling with the use of modern interfaces based on REST (Representational State Transfer). REST is a well-established programming pattern often used in applications with distributed architectures. By using REST, access to services, applications and advanced scripts deployed on the infrastructure becomes straightforward.

Readiness for various applications

Applying REST principles in the implementation of the rimrock service allows to use its functionalities independently of any programming language. It is therefore possible to create web and desktop applications as well as prepare advanced computation scripts (e.g. with the use of *Bash* and the *curl* command). An interesting approach also supported by the service is the ability to develop web applications, which can be run solely in the user web browser, minimizing the role of server-side software.



Support for several job management systems

The rimrock service uses Slurm job management system, what ensures support for its unique features. It allows for easy integration of legacy applications in newly developed systems. Access to computation results is facilitated by hiding the internal file transfer protocol (*GridFTP*) and by grouping the results according to the executed jobs.

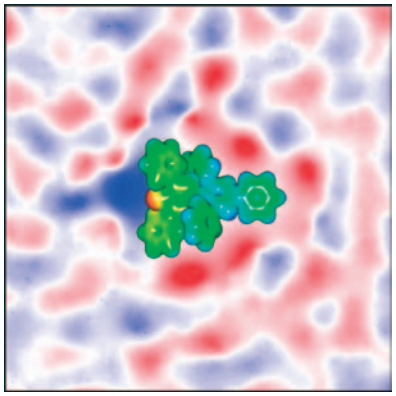
Data security

Data exchanged with the rimrock service is transferred with secure HTTPS connections and for user authorization a temporary user certificate (so called *proxy*) is used.

https://submit.plgrid.pl

Chemistry and Biology – electronic structure and molecular dynamics software

Modern computational chemistry requires constantly increasing resources. More and more computational power is needed to make large systems (especially current challenges of nanotechnology or biological sciences) tractable and improve the accuracy of obtained results. Fortunately, constant progress in computer technology and specialised software offered by Cyfronet meet this demand and enable various chemical computations.



A. Eilmes, P. Kubisiak: Electrostatic potential of an ionic liquid around the solvated dye molecule

Cyfronet clusters' nodes provide up to 1.5 TB of RAM and 48 cores per physical node, which enables quantum chemical computations that require a large amount of memory or a high number of cores with shared memory. Moreover, the fast InfiniBand interface allows good speed-up of calculations if distributed over many nodes. Various quantum chemistry codes also need fast and broad I/O for storage systems. The parallel-distributed Lustre scratch file system and the possibility to use RAMDisk on selected nodes enable that.

Efficient quantum chemistry computations also rely on efficient installation of scientific software and its proper usage. Our administrators' team has the necessary skills, knowledge and experience in installing various applications and efficient running computations. Our portfolio of software used in chemistry contains many packages. Among them, there are:

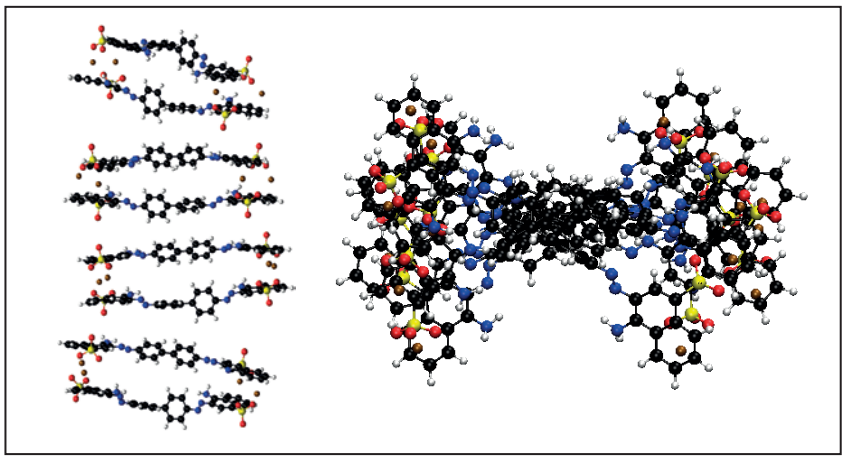
- Versatile and widespread used quantum chemistry codes such as **Gaussian**, **GAMESS US**, **NWChem**, **Schrödinger**, **Q-Chem, Psi4, ORCA** and **TURBOMOLE**, which are capable of calculating electronic structure and various properties of diverse molecular systems using both *ab initio*, density functional theory and semi-empirical methods.

- **Molpro**, **CFOUR** and **Dalton** suites to analyse chemical systems with great accuracy using sophisticated methods such as CC (up to CCSD(T)) and MCSCF.

- Amsterdam Modeling Suite **(AMS, DFTB, MOPAC, COSMO-RS)** provides methods to examine various properties (especially spectroscopic, such as NMR and ESR spectra) of molecular systems with reliable relativistic ZORA approach, COSMO-RS method and all-electron basis sets for the whole periodic table. With addition of versatile and well-constructed GUI of AMS (**AMSInput**, **AMSViev**, etc.) ADF package is used by many of our users.

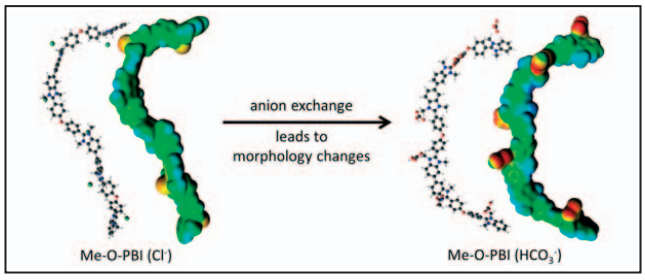
- Several packages, which could be used for solid-state systems. Among them **BAND**, **Quantum ESPRESSO** and **SIESTA** are worth mentioning.

- **AlphaFold** using machine learning to analyze the geometric structures of proteins.
- **Desmond**, **Gromacs**, **Amber**, **LAMMPS**, **NAMD**, **Tinker-HP**, **CPMD**, **CP2K** and **Terachem** suites for molecular mechanics and molecular dynamics simulations of systems containing hundreds of thousands and more atoms.



O. Klimas: Optimized stack of eight Congo Red molecules seen from different perspectives

Nowadays general-purpose computing on graphics processing units (**GPGPU**s) in many scientific domains provides great speed-up of calculations (up to several orders of magnitude). In our computing Centre, some of nodes provide possibility of such calculations on **CUDA** enabled **GPGPU**s (up to eight cards per node). Among software prepared to run on graphical processors our administrators' team prepared quantum chemical packages such as **GAMESS**, **Terachem**, **NAMD**, and **Quantum ESPRESSO, Tinker-HP**. Our experts extensively collaborate with several, mentioned above, packages developer teams. The Cyfronet team prepares and helps with adjusting the dedicated computing environment for our users.



Electrostatic potential of molecules in anion exchange membrane. Published by W. Germer, J. Leppin, C. Kirchner, H. Cho, H. Kim, D. Henkensmeier, K. Lee, M. Brela, A. Michalak and A. Dyck in Macromol. Mater. Eng. 2015, 300, 497–509

Machine learning (ML) and artificial intelligence (AI)

AI-accelerated data analysis is making great strides in many research domains, including materials as well as life science, linguistics and social science. The ability of neural networks to learn from complex data may significantly improve data analysis, classification and pattern detection, with potential applications in many systems, including image recognition, language processing and optimization.

The Cyfronet supercomputing centre faces up to these challenges and prepares several packages:

PyTorch is a package, specifically a machine learning library for the Python programming language, based on the Torch library. It enables implementation of complex Deep Learning algorithms from the Natural Language Processing, video and images processing and many other areas. It can be used for modeling new architectures in the field of machine learning with focus on experiments.

TensorFlow allows, like Pytorch, to implement models based on the tensor flow paradigm. Due to its character and static representation graph, it allows for efficient optimization of models training and inferences with respect to the computing platform.

Keras is a library used for designing neural models. It is an external API for engines based on TensorFlow, Microsoft Cognitive Toolkit, Theano, or PlaidML. It has been designed to enable fast experimentation with deep neural networks. It focuses on being user-friendly, modular, and extensible.

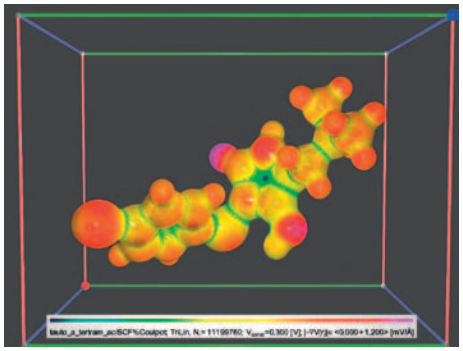
Scikit-learn is a software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN. It has been designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

SchNet is a deep learning architecture that allows for spatially and chemically resolved insights into quantum-mechanical observables of atomistic systems.

Horovod is a distributed training framework for TensorFlow, Keras, PyTorch, and MXNet. The main goal of Horovod is to make distributed Deep Learning fast and easy to use.

Data Visualization, POVRay/ScPovPlot3D

Data visualization enables analysis and understanding of the results of even very complex numerical calculations, especially multidimensional or time-dependent. Most applications for numerical calculations have a module that generates their visualization. Python has a matplotlib or VTK+ module, while Matlab or R also have graphic libraries. The situation is similar with regard to geovisualization programs (GIS) or chemical calculation programs. Unfortunately, no matter how much these programs are refined, the result of their operation is limited by the Cartesian product of available (and compatible) options.

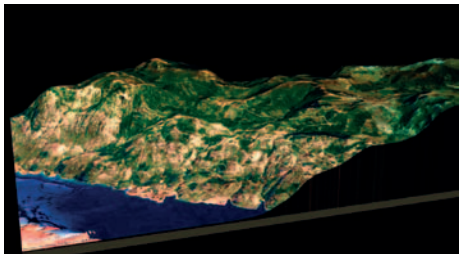


Janusz Opiła: Electrostatic field configuration around the polymer molecule. An equipotential surface with a trilinear approximation is shown, color encodes the electric field intensity module

Overcoming of this limitation, at least for the purpose of creating a prototype of visualization style for later implementation in a dedicated package, is possible, but requires using a general purpose graphics program, for example 3DMax, Blender or POVRay. However, only the latter is equipped with a scripting language (*Scene Description Language* – SDL), which allows for programmatic, non-interactive creation of visualizations, so is useful for mainframes. As the use of countless SDL language options requires quite persistent studies, a dedicated API was written in the form of a set of specialized modules named the "ScPovPlot3D". This is not a completed project as further extensions are still being added. Thus it may be called a beta version, but mature and working. Currently, the project is in version 4.0 and is hosted on GitHub (URL: *https://github.com/JustJanush/Plot3Dv4*) – the multiplatform API requires POVRay at least in version 3.7.

The most important modules are:

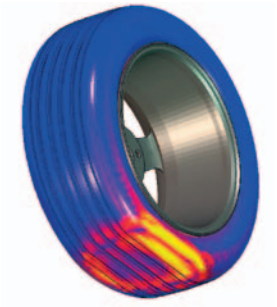
- VectorField.inc – hybrid vector field visualization using widgets and/or field stream tubes,
- Potential.inc – hybrid visualization of scalar fields, on regular and irregular meshes with trilinear or centripetal Catmull-Rom cubic interpolation,
- BPatchSurf.inc – hybrid surface visualization based on data on regular or irregular grids with implemented simple kriging (KDE),
- Mesh2Surf.inc – hybrid visualization of data defined on regular 2D grids (z=f (x, y)),
- TextExt.inc – extended 3D text formatting, oriented to the presentation of mathematical formulas.



Janusz Opiła: Terrain visualization based on altitude data collected on an irregular grid and textures obtained from the Google Earth Pro application vicinity of Karlobag, Croatia). Own study: DOI: 10.23919/MIPRO.2018.8400037

If necessary, the package's developer provides technical support. Contact information: *https://skos.agh.edu.pl/osoba/janusz-opila-2390.html*

CAD/CAE applications



Computer-Aided Design and Computer-Aided Engineering applications are essential tools in developing and building almost everything – from car parts to buildings. Through computer simulations, engineers can check the durability of constructs and devices; perform linear and non-linear structural analyses of contact phenomena, plasticity, recoil, etc. CAD/CAE software provides analysis of thermal conductivity, radiation and phase shifts. Significant for science are also fluids simulations: velocity fields, pressure fields, heat distribution, chemical reactions, etc.

Cyfronet's users can resolve all these tasks thanks to CAD/CAE packages of ANSYS, ABAQUS, FLUENT, MARC and OPERA.

ANSYS is a complex structural simulation package with an intuitive graphical user interface, supporting scientists from nearly any area of science or business. Results are calculated with high precision and may be presented by plots or tables, for example, isosurface diagrams and deformations. Computational capabilities of ANSYS are very high and involve: harmonic and spectral analysis, statistics and dynamics.

ABAQUS is devoted to solving problems in the industry using finite-elements analysis. A user can prepare a combination of finite-elements, materials, procedures of analysis and sequences of loads, according to individual requirements, to simulate vehicle loads, dynamic vibrations, multibody systems, impacts, crashes and much more.

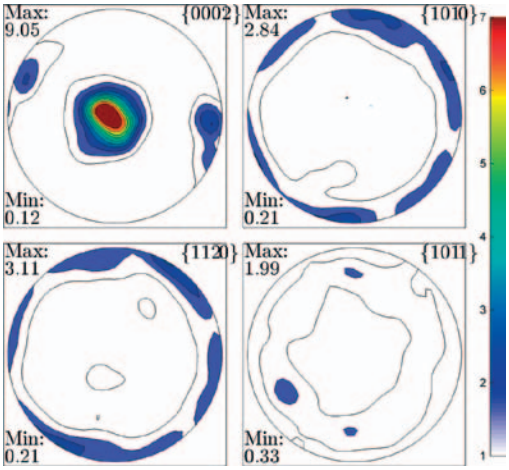
FLUENT software offers the broad physical modeling capabilities needed to model flow, turbulence, heat transfer and reactions for industrial applications ranging from air or liquid flow to semiconductor manufacturing. FLUENT can be used in numerous science domains, including chemistry, metallurgy, biomedicine, electronics, material design and many others.

MARC is a general-purpose, non-linear finite element analysis solution to accurately simulate the product behavior under static, dynamic and multi-physics loading scenarios. It can simulate all kinds of non-linearities, namely geometric, material and boundary condition non-linearity, including contact. It is also the solution that has robust manufacturing simulation and product testing simulation capabilities, with the ability to predict damage, failure and crack propagation. All that can be combined with its multi-physics capabilities that helps couple thermal, electrical, magnetic and structural analyses.

OPERA is a finite element software suite for design and optimization of electromagnetic devices in 2D/3D. It gives accurate numerical solutions for problems from multiple areas of science, including electrostatics, magnetostatics, low and high frequency electromagnetics. The software gives an ability to design and optimize many types of electrical devices: transformers, motors, switches, micro-machines, MRI scanners and X-ray tubes. It is a powerful virtual prototyping facility to accelerate the design process.

Symbolic math applications

Mathematical applications enable to conduct in reasonable amount of time even very complex and complicated calculations. Users of ACC Cyfronet AGH have access to software that supports calculations in the field of algebra, analysis, combinatorial math, statistics, theory of numbers, geometry or other math areas. Running calculations like integration, differentiation, symbolic processing, matrix operations, approximation and interpolation, Fourier and Laplace Transforms, digital signal processing, etc. is a lot easier. Results can be visualized with appropriate tools. Some of the applications can create interactive 2D and 3D plots. In scientific work, preparation of precise model that most accurately describes analysed issues, is essential.

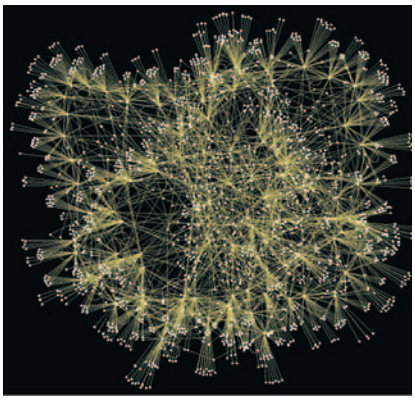


Bartosz Sułkowski: Results of texture simulations by visco-plastic self-consistent model of Zn after hydrostatic extrusion at 250 °C

A good example of software environment, which can be applied in above-mentioned issues, is **MATLAB**. Its modules (Toolboxes) allow performing computations in the field of financial modelling, partial differential equations, linear and non-linear optimization and much more. It is also possible to use Simulink – the environment oriented for simulations and visualizations from blocks, without the need for traditional programming.

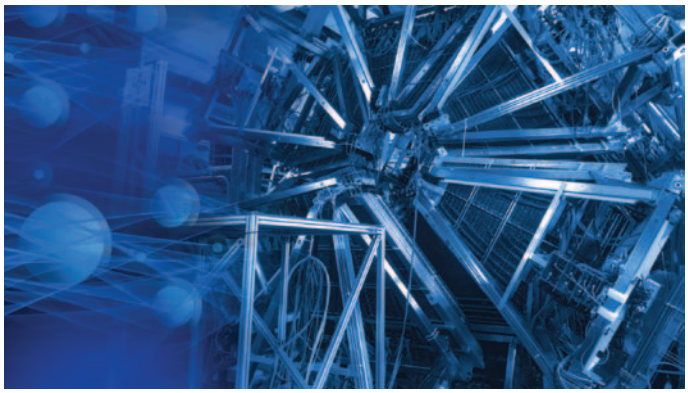
Apart of that environment, users can find in our software a useful application, **MATHEMATICA**, which allows parallel computations with defined precision, dedicated for symbolical and numerical calculations. An advantage of MATHEMATICA is, among other things, a tool for fixing mistakes.

Another example of universal and interactive mathematical software is **MAPLE.** It can be used for simplification of expressions and symbolic processing. It offers databases, enables code generation in other programming languages, creating slideshows with user commands and communication with MATLAB and CAD systems.



Rafał Rak: One minute price returns network for KGHM (the Polish stock company)

Dynamically developing scientific research requires more and more advanced tools nowadays. Among them, IT tools play a huge role, supporting the effective research from the moment of its design to the development of results. Cyfronet, by following the latest solutions and creating its own studies, tries to fulfill an important area of its mission to support science. Dedicated laboratories were established for these needs.



Laboratory of Quantum Computing

The Laboratory was established to conduct research on the use of quantum computers in calculations and to support classical calculations with quantum accelerators.

One of the key tasks is to follow the development of quantum computing technologies and available quantum accelerator platforms in order to use them in dedicated services offered by Cyfronet. We cooperate with other research entities and industrial partners, both as consortium members of joint initiatives and carrying out commissioned works.

Based on our own competences and the exchange of expert knowledge with a network of partners, our team is working on solving the problems that prevent us from wider and more effective using the quantum accelerators in calculations for the benefit of science and economy.

We also act to popularize calculations using quantum accelerators and provide substantive user support. In this regard, we prepare the necessary documentation and materials, conduct training and publish the results of research work.

Contact: Mariusz Sterzel, m.sterzel [at] cyfronet.pl

Laboratory of Parallel Algorithms

The Laboratory focuses on the broadly understood computational aspect of parallel algorithms, with particular emphasis on machine learning algorithms, and the possibility of their effective use on large computing clusters.

Due to the increasing amount of data available every day, traditional machine learning algorithms are becoming insufficient and the serial processing paradigm is computationally inefficient. In order to meet the new challenges related to the growing amount of data, in many cases it is necessary to use large computing clusters, which forces the adaptation of the algorithms used to work in parallel mode. In the Laboratory of Parallel Algorithms, we consider both theoretical and practical aspects related to this task. In particular, we focus on the following areas: computer vision, tensor computing, deep networks, low quality image processing, underwater image recognition, hyperspectral data classification, histopathological data classification, data unbalance.

Within the considered domains, we develop new algorithms that use parallel computations, in particular, we study the theoretical and practical aspects related to this phenomenon.

Contact: Bogusław Cyganek, cyganek [at] agh.edu.pl

Laboratory of Information Methods in Medicine

The main tasks of the Laboratory focus on two spheres. The first is research activity, which includes a thorough analysis and verification of available and potential answers to the challenges found at the border of medicine and information technology. The second one covers the design, development and subsequent operation of dedicated applications and platforms for medical applications. This scope also covers the monitoring of the security status of the developed software as well as data storage and processing mechanisms.

Thanks to the comprehensive approach to the processes: from the identification of the research problem, through the analysis of users' needs, to the final implementation, the Laboratory effectively implements its mission to support the scientific and medical community. As part of the dissemination of expert knowledge, members of the Team publish research results in scientific journals, participate in the preparation of information materials and conduct consultations for users.

Laboratory employees establish cooperation with renowned domestic and foreign research institutes and medical IT centers. The effects of this cooperation are, among others, ongoing and already implemented projects with significant participation of Team members:

- Sano: Centre for New Methods in Computational Diagnostics and Personalised Therapy,
- PRIMAGE: PRedictive In-silico Multiscale Analytics to support cancer personalized diaGnosis and prognosis, Empowered by imaging biomarkers,

- Virolab: A Virtual Laboratory for Decision Support in Viral Disease Treatment,
- Gliomed: Diagnostics of gliomas based on the slowly circulating DNA of the tumor,
- Eurvalve: Personalised Decision Support for Heart Valve Disease,
- CECM: A Centre for New Methods in Computational Diagnostics and Personalised Therapy.

The previous activity of the current Laboratory team is presented in detail on the following website: *http://dice.cyfronet.pl*.

Contact: Marian Bubak, bubak [at] agh.edu.pl

Laboratory of Data Processing

The Laboratory designs and implements dedicated applications and software platforms for applications in various fields of science. The Laboratory consists of specialists in the field of software architectures, Front-end and Back-end programming, user interface and user experience design, DevOps, testing, and requirements analytics. The team specializes mainly in:

- development of innovative methods of acquiring knowledge from available data,
- development of technologies supporting open data processing,
- integration of data and knowledge processing systems with existing repositories and e-infrastructures.

The Laboratory establishes cooperation with renowned scientific and research units as part of Polish and international projects. The effects of cooperation include:

- **Construction of the Sat4Envi Portal** (*https://dane.sat4envi.imgw.pl)*, providing satellite data from the Copernicus program. The portal enables searching, viewing, ordering and downloading satellite data and their derivative products using only a web browser.
- **Development and maintenance of the EOSC Portal** (*https://eosc-portal.eu/*) as part of a series of projects related to the European Open Science Cloud (EOSC). The portal provides access to the resources of many European e-infrastructures and research infrastructures through a unified user authentication system. EOSC activities focus on the implementation of the Open Science paradigm.

The Laboratory team was also involved in the creation of the **PLGrid Portal** (*https://portal.plgrid.pl*), which provides scientists with many software packages, libraries and scientific tools. In 2022, the new Laboratory of Interdisciplinary Scientific Computing was separated from the Laboratory of Data Processing.

Contact: Roksana Wilk, r.wilk [at] cyfronet.pl

Laboratory of Interdisciplinary Scientific Computing

The Laboratory conducts research and development work on the processes of conducting and supporting scientific calculations and the organization of scientific data. The Science Gateways portals elaborated by the Laboratory employees create the possibility of establishing cooperation with external entities: Polish and foreign.

The Laboratory tasks are also focused on:

- implementation of research grants and industrial orders,
- using expert knowledge to solve problems requiring the use of various computing resources, e.g. machine learning technology,
- popularization of calculations using the tools created by the Laboratory among users,
- essential support for users, realized among others by monitoring needs, developing documentation, and conducting training.

The team of the Laboratory, previously co-creating the Laboratory of Data Processing, developed the proprietary InSilicoLab programming environment, which includes a set of advanced tools and programming libraries that allow for the construction and development of dedicated research portals. Portals based on InSilicoLab are designed in such a way as to gather in one place all the tools that researchers need for *in silico* calculations. The main advantages are:

- easy running of the user experiments, even if they are complex, long and require many calculations,
- the ability to conveniently describe, categorize and search for input or output data.

The InSilicoLab technology is distinguished by striving for the greatest possible usability of the tools built with the help of the environment. This sphere includes both the usefulness for solving scientific problems in a given field, as well as the user-friendliness of the portal for its end-user.

The effects of the Laboratory team cooperation with renowned scientific and research units as part of Polish and international projects include the development of the **EPISODES Platform** (*https://tcs.ah-epos.eu/)* as part of a series of projects related to the European Plate Observing System - EPOS (*https://www.epos-eu.org/*). The portal and the tools organized around it are focused on the study and analysis of seismicity and other phenomena caused by human activity (e.g. exploitation of resources within a mine, creation of artificial water reservoirs). The portal is integrated with the European EPOS infrastructure.

Contact: Joanna Kocot, j.kocot [at] cyfronet.pl

Laboratory of Cloud Technologies

The Laboratory deals with the design and operation of the cloud for science, as well as tools for its effective use. The team is developing comprehensive environments for access to distributed data, taking into account both the issues of secure data storage and processing in the cloud, as well as convenient access interfaces (portals, applications) for the end user.

Bearing in mind the dynamic development of new technologies for processing and storing data in the cloud, the Laboratory constantly conducts research and publishes the results. Using the team's expert knowledge, it actively supports scientific initiatives, including international projects and e-infrastructures.

The flagship product of the Laboratory is **Onedata**: a globally scalable data management system, unifying access to data stored in distributed systems. Onedata responds equally well to the needs of both small user groups and large international research communities. The system enables users to use a homogeneous data management system for both personal and work-related data storage, such as research results, and enables accessing it efficiently from any device.

More information at: *https://onedata.org*.

Contact: Łukasz Dutka, l.dutka [at] cyfronet.pl

Laboratory of Applications of Computational Techniques

The key aspect of the Laboratory's operation is the development of numerical models and computer simulations based on supercomputer architectures, with particular emphasis on practical use in industry. The Laboratory conducts its own research on computational techniques, as well as tracks and analyzes new solutions in their application.

The Laboratory team establishes cooperation within Polish and international consortia, and together with partners it implements projects aimed at developing or improving the production processes of metal products. Among the most important effects of these activities are:

- **VirtROLL - Virtual strip rolling mill**: the main goal of the project was to create a computer system supporting the flexible design of rolling technology for flat products based on the calculation results of dedicated numerical simulation modules.
- **PROTEUS-RS - Long products manufacturing processes optimization strategies to improve the finished product quality by minimizing residual stresses**: the project provides for the production of models and a series of numerical simulations based on high performance computer architectures for the design of the long metal fabrication process.

Contact: Łukasz Rauch, lrauch [at] agh.edu.pl

Laboratory of Visual Techniques

Laboratory of Visual Techniques (Media Lab 5.0) is a film and television studio whose main task is to produce materials documenting and promoting the achievements of ACC Cyfronet AGH.

The popular science and training films and TV programs that we make show the results of research work carried out by the users of Cyfronet's computing infrastructure.

The Laboratory hosts workshops on journalism, the art of cinematography, sound production, editing, computer graphics and television studio operation. It is also a place of research in vision technologies, acoustics and online transmission techniques.

The Laboratory is a film and television studio with an area of approximately 100 sq.m., equipped with a grate with lighting, sound system, with an artificial horizon and scenery. The studio is connected with an auditorium for about 50 people. The studio is also connected with the control rooms of: VISION and LIGHT (with stations of: adjustment of camera tracks, vision realization, recording and playback, computer graphics and light), SOUND with the Yamaha LS 9 console and LECTOR. There are two ASSEMBLY ROOMS with computers equipped with Adobe CS6 software in the Laboratory. The Laboratory has Panasonic AG-HPX 371 cameras (3 pcs.) and other film and television equipment for image, light and sound production in studio and outdoor productions.

The studio cooperates with professional creators and producers as well as students of Krakow's universities.

The films made by MediaLab 5.0 can be viewed, among others:

- on the YouTube channel of ACC Cyfronet AGH (*https://www.youtube.com/user/CyfronetAGH/videos*),
- on the Pionier.tv platform (*https://pionier.tv/wideo-tag/cyfronet/*).

Contact: Jacek Przybylski, j.przybylski [at] cyfronet.pl

Laboratory of Acceleration of Computing and Artificial Intelligence

Our mission

Thanks to our expertise in the field of the Artificial Intelligence (AI) algorithms, and our knowledge of the modern computational methods, we can support scientific community in their AI-based research. Our knowledge and experience allows us to implement machine learning algorithms and dedicated to neural networks effectively using the AI-dedicated partition of Prometheus' supercomputer available at ACC Cyfronet AGH. The AI partition of Prometheus was built based on the four efficient computing servers; each equipped with eight nVidia Volta V100 GPGPU cards. The total computing power of the partition is over 4 PetaFlops, which is over four quadrillions ($4x10^{15}$) of AI-dedicated computations per second.

Domain specializations

Selected applications of Artificial Intelligence in the areas of natural language processing (NLP), image processing and time series analysis for various research problems are presented below.

Natural language processing

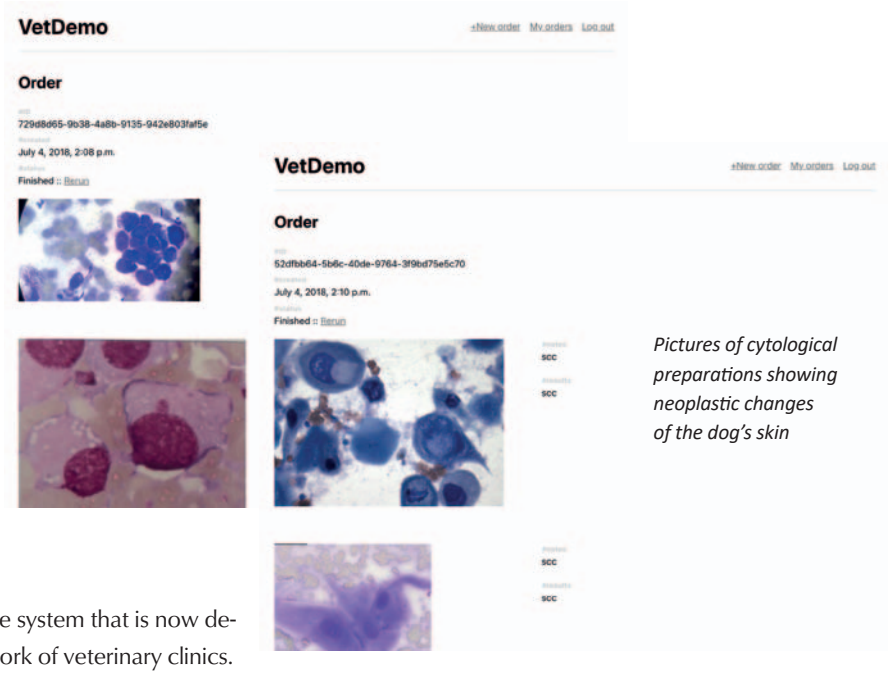
The Laboratory is proud of many years of experience in natural language processing. As an example, we have created and developed the tool that allows the users to search, compare and classify text documents. The result of this work is the web service called Scholar that is available at ACC Cyfronet AGH on the PLGrid platform. One of the important research problems is the analysis of impact of the methods for reducing the accuracy of textual data representation on the effectiveness of the NLP algorithms. We have managed to develop alteration of the methods that allow for a 10-fold reduction in computing energy consumption, if compared to the original implementation, with no significant loss of accuracy.

The emergence of neural networks-based solutions has revolutionized the NLP field. We research on the compression and hardware implementation of the sentiment assessment network, which showed that it is possible to reduce the accuracy of the network coefficients precision to 8 or even 4 bits while maintaining the network efficiency almost unchanged. Additionally, we examine the area of semi-supervised learning, where the amount of available tagged data is very limited and the output categories changed during the operation of the system. Our research has shown that it is possible to develop a solution with an accuracy of up to 98.9%.

Image processing

In the field of image processing using neural networks, the Laboratory's work focuses on the recognition and detection of objects for the needs of medical applications. During the tests, we developed

a system for the classification of neoplastic changes in the samples from the cytological examination of the animal tissue. The proposed system achieved an efficiency of approximately 96% for the three selected types of cancer. We used deep network models based on Resnet-50 and Resnet-152 nets. Also, to allow for the selection of the perception area in classification operations, we have developed a special training scheme that is based on genetic algorithms. One of the latest developments in the area is the Yolo3-based detection, which achieves a very good value 0.86 of mAP for the specially prepared images that are made with the use of many low-quality cytological preparations. The system that is now developed is to be ultimately used in the daily work of veterinary clinics.

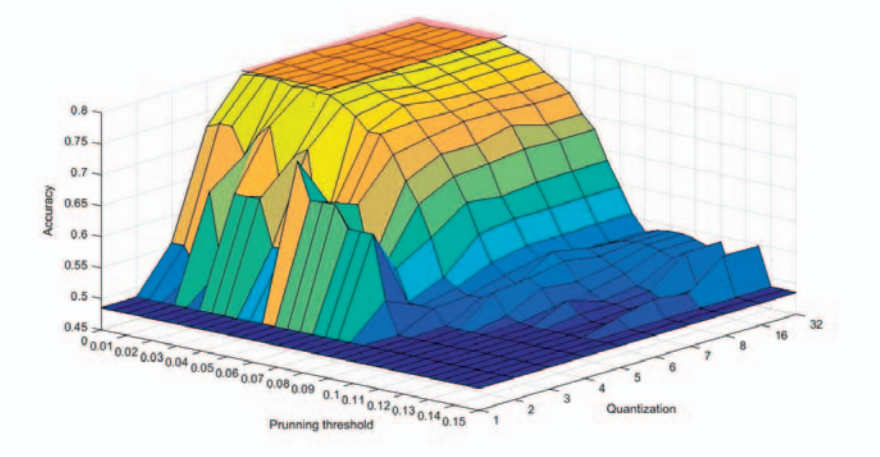


Pictures of cytological preparations showing neoplastic changes of the dog's skin

Time series analysis

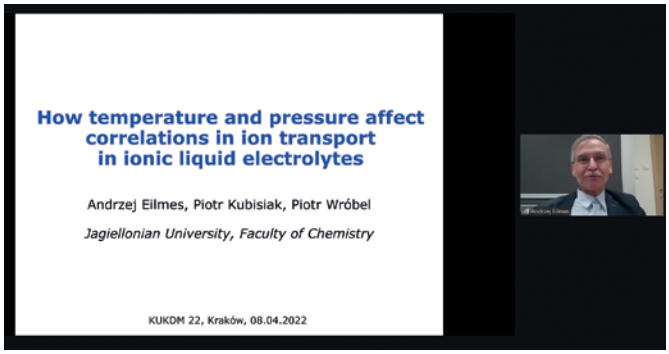
The works of the Laboratory also concern the modelling of time series for the detection of unusual situations. Detailed work includes practical applications such as anomaly prediction to avoid catastrophic damage to the magnets and other associated devices of the Large Hadron Collider (LHC) at CERN. By means of GRU and LSTM neural networks and dedicated post-processing, we have developed a system for quench detection in superconducting magnets, which additionally enabled the classification of detected anomalies. Moreover, to allow the user for very low response latency, the developed algorithm was implemented on the computing platform that is based on the Xilinx Zynq UltraScale+ MPSoC 285 FPGA architecture.

In the field of medicine, we address the problem of prediction of fainting of hospitalized patients, who are confined to a hospital bed for a long time. Thus, the analyzes concern the well-known problem of the loss of the leg muscles support for the cardiovascular system. This research is carried out with the cooperation of the Medical University of Graz.



HPC Users' Conference (KU KDM)

The basic premise of the HPC Users' Conference was to initiate annual scientific meetings devoted to users performing computations in ACC Cyfronet AGH with use of high-performance computers, computing clusters and installed software.



The conference was launched in 2008 and included several presentations by Cyfronet employees – describing the resources available in the Centre, as well as numerous lectures of researchers – presenting the scientific results achieved using Cyfronet hardware and software. In addition, two invited speeches were given – by Norbert Attig from Jülich Supercomputing Centre and Jaap A. Kaandorp from University of Amsterdam.

The first edition of the conference attracted much attention and increased the interest of users in Cyfronet resources. It proved that this type of event was much awaited and needed.

Nowadays, the HPC Users' Conference focuses on the large-scale computations and simulations, novel algorithms in computer science, tools and techniques relevant to high-performance computing, teaching in computer science, databases. However, the main aim of the conference is the overview of research results carried out using the computer resources of Cyfronet. It is also an opportunity to familiarize the users with the Centre and its resources, including the PLGrid infrastructure.

The conference includes a series of talks by scientists who perform research using Cyfronet resources and can present the role of these resources, typical usage scenarios and performance aspects. The event is an important opportunity for Cyfronet representatives to meet with these scientists and acquire the knowledge necessary to take the proper actions in order to adapt the computing infrastructure to scientists' needs and fulfil their requirements. On the other hand, the conference also gives a possibility for researchers representing various disciplines to exchange experience and become familiar with the new technologies and domain-specific services currently being deployed at the Centre.

The crucial parts of the conference are meetings with suppliers of Cyfronet hardware and software, as well as the panel discussion on efficient use of these resources. The latter is always attended by users – researchers, who use the chance to get familiar with news regarding the computing infrastructure in the Centre and to inform Cyfronet experts about issues encountered while interacting with this infrastructure.

In 2022, due to the ongoing coronavirus (COVID-19) pandemic, the conference was held online.

Contributed papers elaborated on the basis of the best conference talks were published in one of two well-regarded IT journals: Computing and Informatics (CAI) (*http://www.cai.sk*) or Computer Science (CSCI) (*www.csci.agh.edu.pl*).

https://www.cyfronet.pl/kukdm22

Sat4Envi – space technologies, widespread use

By providing satellite data from the Copernicus program, the Sat4Envi project has opened a new chapter for optimal decision-making in spatial planning and rapid response to environmental threats.

Satellite data at your fingertips

The core achievement of the project is available at *dane.sat4envi.imgw.pl*. The Customer Service System is a web browser operated tool for accessing the Sat4Envi satellite data warehouse. Thanks to this platform, the end-user gains the ability to search, view, order and download satellite data and its derivatives. The system is primarily used to monitor the earth's surface, the atmosphere and the weather, and identify meteorological threats. Access to all data is completely free of charge and does not depend on affiliation or profession. Moreover, it does not require registration in the basic version. After logging in, however, the user has the option of using additional functionalities, such as saving a set of the most frequently selected measurements or accessing the metadata with which the presented satellite images are described.

By the possibility to use WMS (Web Map Service) layers, which include borders of administrative units, the tool is functional for public administration units or regional institutions. The operation of the Customer Service System is intuitive and allows to quickly reach the desired types of data.

The project and collaboration

A system for collecting, sharing and promoting digital satellite information about the environment – Sat4Envi – aimed at providing wide and easy access to satellite data from the Copernicus program and data from other environmental or meteorological satellites. Behind the success of Sat4Envi stands the involvement of four closely cooperating units. Those are: Institute of Meteorology and Water Management – National Research Institute (IMGW, the project leader), Polish Academy of Science Space Research Centre (CBK), Polish Space Agency (PAK) and Academic Computer Centre CYFRONET AGH-UST.

IMGW was primarily responsible for the development of the station for receiving and processing data from polar satellites (the station was established in Krakow) and played the role of the coordinator. The joint effort gave the possibility to efficiently implement the other assumptions of the project, including:

- building a modern satellite data archive (IMGW, CYFRONET),
- the training center for new satellite technologies (CBK, PAK, IMGW),
- the center for providing scientific information to users (CYFRONET, IMGW, CBK).

The project, with a total value of PLN 17,903,900, was co-financed by the European Regional Development Fund under the Digital Poland Operational Program.

Technological challenges

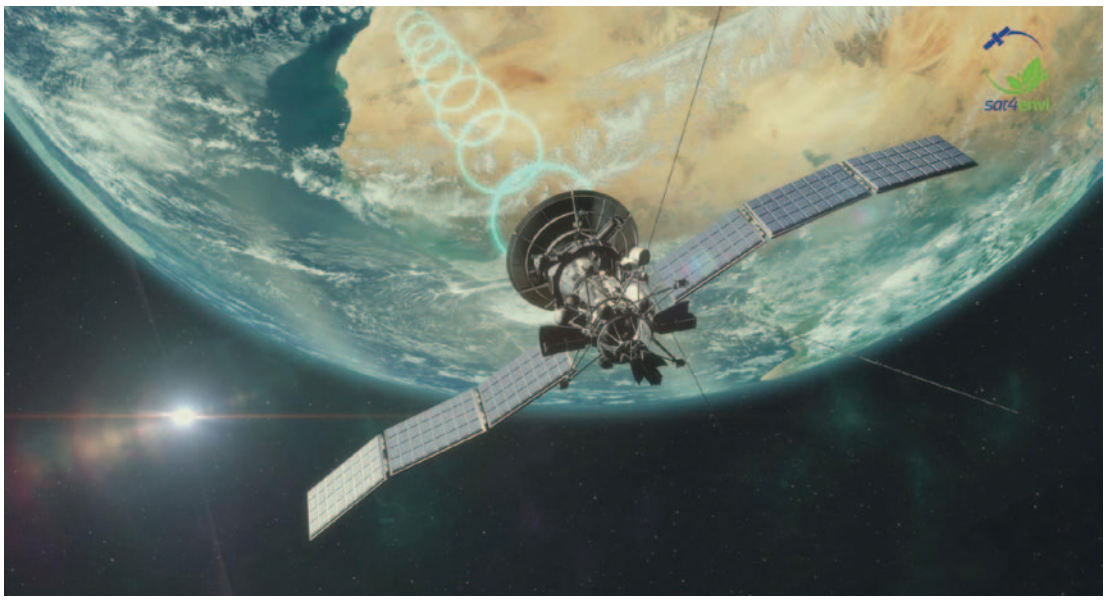
Building a portal with an access interface required a very careful and multi-faceted design, taking into account the ease of data search or the speed of finding specific functions. With the goal of creating a highly functional product also for "non-advanced" users, the development process was started with an extensive requirements assessment. Use cases were analyzed in various target user groups and multiple verifications were carried out using graphical mockups and prototype implementations. When working on the interface, user experience specialists also used the technique of in-depth interviews in groups of representatives of future users (i.e. IMGW, CBK, PAK, The Remote Sensing Laboratory or Pieniny National Park).

- *The role of Cyfronet in the project was to implement programming tasks related to the construction and maintenance of the Customer Service System, but also to provide the IT infrastructure necessary to collect and share data. Cyfronet is the operator of the Internet connection with the Institute of Meteorology and Water Management in Kraków, and as part of the PLGrid infrastructure, it provides computing resources necessary for the proper functioning of Sat4Envi. In turn, the Prometheus supercomputer installed in Cyfronet is used to process data and create information from raw satellite data* – emphasizes Prof. Kazimierz Wiatr, Director of ACC Cyfronet AGH-UST.

Sat4Envi supports in crisis situations

The analysis of satellite data with the smallest possible delays in relation to its acquisition may be of key importance for forecasting, monitoring and counteracting the effects of natural disasters. In particular, we are talking about monitoring floods, assessing the scale of fires, or monitoring damage to agricultural crops caused by drought or frost. Due to the importance of these activities and the need for rapid response, the Sat4Envi project includes support for special user groups such as the Crisis Information Center.

More information about the project can be found at: *https://sat4envi.imgw.pl/*.



Prometheus and its infrastructure are used to share data from Copernicus Programme. Image: Sat4Envi

EPOS – multidisciplinary platform supporting Earth sciences research

We need to use various tools and research methodologies to effectively monitor the processes ongoing below the Earth's surface and their effects perceptible on the surface. In this regard, EPOS – European Plate Observing System is to deliver complex solutions, which would cross the borders of different countries and research disciplines as well.

In October 2018, EPOS was granted the status of ERIC: European Research Infrastructure Consortium. This way legal and administrative framework was established, which allows EPOS to carry out international activities not only as a project implemented by many partners, but rather as a separate entity with its headquarters in Rome.

The idea standing behind EPOS is to recognise better Earth functioning as a complex system in which, on the one hand, natural episodes (such as volcanic eruptions, floods or earthquakes) have an impact on society and the economy. On the other hand, the environment is being changed by anthropogenic factors. For this reason, effective research requires a multifaceted approach. By integrating European research infrastructures into the transnational system, which delivers data, data products, software and services, EPOS works on enabling access to so far dispersed possibilities. The initiative focuses on ten main, linked with each other, Thematic Core Services (TCS): Seismology, Near-Fault Observatories, GNSS Data and Products, Volcano Observations, Satellite Data, Geomagnetic Observations, Anthropogenic Hazards, Multi-Scale Laboratories, Geo-Energy Test Beds for Low Carbon Energy, Tsunami.

Image: EPOS - European Plate Observing System

Computational Earth sciences

Providing the research infrastructure as an internet platform requires close cooperation of IT specialists with representatives of Earth sciences. Synergic activities are carried out on many levels – starting from measurements and experiments conducted by scientists generating a large amount of raw data, which is stored and processed in information systems – i.a., within Cyfronet's infrastructure. Cooperation is also significantly influential in building data visualisation tools, which are managed from the web browser level using Graphical User Interfaces (GUI). In the long run, EPOS aims to connect research communities within one multidisciplinary platform, provide effective and safe access tools, and develop new services based on previous achievements.

It's worth underlining that geophysical and geological data, advanced visualisation and analytic software will be available in open access and free of charge. Thanks to that, EPOS supports interdisciplinary research on the causes and effects of processes ongoing below and on the Earth's surface.

This is especially important in the light of monitoring environmental threats, such as earthquakes, floods, landslides or volcanic eruptions, and anthropogenic threats related to, among other things, the activities of mines. Through a comprehensive analysis of already observed phenomena, one can better prepare for events that have not yet taken place, limiting their adverse effects as much as possible.

EPOS has entered the operational phase of the infrastructure built within EPOS PP and EPOS-IP projects. The EPOS research platform is available at: *https://www.ics-c.epos-eu.org/,* and the portal gathering information about EPOS activities at: *https://www.epos-eu.org/*.

Cyfronet's work for EPOS

ACC Cyfronet AGH participates in subsequent EPOS projects, supporting the initiative with hardware resources and the specialists' knowledge and skills. The parallel activities include several areas, one of which is the development of the EPISODES digital platform for the Thematic Core Service – Anthropogenic Hazards (TCS AH). The platform, accessible at *https://tcs.ah-epos.eu,* is a tool for the analysis of anthropogenic seismicity and associated threats and for the assessment of the potential environmental impact of the exploitation of geo-resources. Thanks to data from seismic stations and shared industrial information, the platform allows more straightforward analysis of such processes as flooding artificial reservoirs, extraction of raw materials, shale gases or groundwater.

In addition to activities related to sharing and visualisation of data, Cyfronet's specialists take care of the maintenance of integration elements between the TCS AH and the superior EPOS ISC-C Portal, which gives access to the results of activities led in all of the EPOS Thematic Core Services.

Additionally, Cyfronet is sharing a virtual workspace for data management and for running software from various deliverers.

In turn, as part of the EPOS SP project (EPOS Sustainability Phase), Cyfronet is involved in actions aiming to gain new users for the EPOS infrastructure, including industry partners. Parallel work is in progress regarding the integration of subsequent software packages used in scientific research.

Additionally, by participating in national projects, Cyfronet works to develop the Polish Research Infrastructure EPOS-PL. The latest effort, within the EPOS PL+ project, regards developing software that supports modelling and conducting research using AI methods. The EPOS-AI platform will enable easy use of data integrated into the program, selection of machine learning methods, and launching resource-intensive stages of processing on computational resources of Polish High-Performance Computing Centres.

Sano: Centre for Computational Personalised Medicine - International Research Foundation

Owing to a unique initiative carried out in 2019-2026 by the Academic Computing Centre Cyfronet AGH along with five partner institutions in the framework of the EU Horizon 2020 *Teaming for Excellence* programme, the International Research Agendas programme implemented by the Foundation for Polish Science, and with financial support from the Ministry of Science and Higher Education, a new entity called **Sano –** Centre for Computational Personalised Medicine was established in Kraków.

The mission of Sano involves:

- development of new computational methods, algorithms, models and technologies for personalized medicine,
- introducing new diagnostic and therapeutic solutions based on computerized simulations into clinical practice,
- fostering creation and growth of enterprises which develop cutting-edge diagnostic and therapeutic technologies,
- contributing to novel training and education curricula which meet the needs of modern personalised medicine.

The **Sano Centre** (*https://sano.science/*) is situated in Kraków: a city well known for educating top-class medical practitioners and IT experts, whose teaching hospitals are well regarded among the academic community and whose life science technology sector is continually expanding.

The establishment of the **Sano Centre** directly contributes to regional scientific excellence by fostering new research collaborations and creating top-tier educational opportunities for postgraduate students. It will also improve knowledge and technology transfer by promoting creation of new commercial enterprises which deal with advanced technologies. The Centre's impact will transcend regional boundaries, contributing to advancements in medical research and thereby to the quality of medical care.

The Centre's objectives are based, among others, on the National Smart Specialisation Strategy. **Sano** aims to enhance collaboration between academic and commercial institutions on an international scale. Key performance indicators include the number of highly cited scientific publications and grants obtained by the Centre, the number of solutions based on computational models which have been introduced into clinical practice, and the number of innovative marketable products and services.

The Centre for Computational Personalised Medicine represents a joint international collaboration of the following institutions: ACC Cyfronet AGH, LifeScience Cluster Krakow – a Key National Cluster, University of Sheffield and Insigneo Institute, Forschungszentrum Jülich, Fraunhofer Institute for Systems and Innovation Research ISI, and National Center for Research and Development.

EOSC – transnational integration of scientific resources

The European Open Science Cloud (EOSC) is an initiative aiming to create a virtual environment which would complete the assumptions of the Open Science paradigm. EOSC aims to share (easily and transparently) research data and advanced tools and resources to store, share, process and manage this data.

Within EOSC, the connections between currently existing research e-infrastructures are made, and the integration takes place, i.a., by the unification of access and authorisation rules for researchers from different countries. Thanks to the achievements of EOSC-Hub, belonging to the family of EU-funded EOSC-building projects, a platform was created to do this. The EOSC-Portal is the interface between providers of scientific services and resources, and researchers who can benefit from those. ACC Cyfronet AGH team played a crucial role in this regard, becoming the Portal host and developing the Marketplace website, an extensive catalogue of services and documentation provided by EOSC partners.

Further development of EOSC

Development of the Portal and integration of services from the larger group of deliverers are the core assumptions of the planned enhancement of EOSC.

The EOSC Enhance project (*https://www.eosc-portal.eu/enhance*), which was realised from 2019 to 2021, aimed to improve the Portal in terms of convenience and speed of use. The works carried out by our specialists concerned, among others, advanced analysis of user behaviour in order to create and implement the best user experience practices. At the same time, new functionalities have been implemented.

The EOSC Synergy project (*https://www.eosc-synergy.eu/*), which lasted until the end of October 2022, has been underway to implement EOSC standards for another nine national e-infrastructures. Cyfronet, in addition to coordinating activities at the national level, supported the planning process by looking for new, effective solutions for integration in other countries.

In turn, the EOSC Future project is a continuation and an enhancement of the work that has been done so far. The project aims to upgrade the EOSC ecosystem so that it supports European research even better, and to convince researchers to use the resources which are offered. Cyfronet's responsibility as part of EOSC Future is the further development of one of the three main pillars of the Portal, which is the part aimed at the users, in particular related to the components supporting the Marketplace.

The latest in a series of EOSC-building initiatives in which Cyfronet participates is the FAIRCORE4EOSC project, which involves the expanding of EOSC infrastructure with components that support the FAIR paradigm. Cyfronet is responsible for expert support in data management and personalised search of EOSC resources (including scientific objects) using AI methods (mainly Machine Learning) and implementing solutions developed in the project in the EOSC User Portal.

EuroHPC – European Data-Processing Infrastructure

The European High-Performance Computing Joint Undertaking (EuroHPC JU) was established to radically develop the existing European HPC infrastructure so that it could provide European researchers with computing power comparable to the ones available in the USA, China and Japan. EuroHPC unites 33 countries, from 2018 including Poland, as well as private members. The joint aim is to buy and deploy two exascale supercomputers (EFlops = 10^{18} floating points per second) that will be on the TOP5 list of the world's fastest computers.

Reaching exascale is scheduled for 2023. However, other supercomputers of lower performance have been launched lately to build an infrastructure that will allow scaling the technology and software. By the time this publication is composed, EuroHPC has provided two pre-exascale supercomputers: LUMI (Finland) and Leonardo (Italy). The third supercomputer of this scale will start operating in Spain. Additionally, five petascale supercomputers were launched: Discoverer (Bulgaria), Karolina (Czech Republic), MeluXina (Luxemburg), Vega (Slovenia) and Deucalion (Portugal).

In June 2022, EuroHPC JU selected five more places where new supercomputers will be installed. They are Greece, Ireland, Germany, Poland and Hungary. At the Jülich Supercomputing Center, there will be JUPITER – the first European computer with a computing power above 1 Exaflop (10^{18} floating point operations per second), which, when fully installed, will become the fastest European machine. In Poland, at the end of 2023, a mid-range scale system will be launched, with power several times greater than the currently fastest supercomputer in Poland (Athena). The operator of the new supercomputer will be the Academic Computer Center CYFRONET AGH in Krakow, and access for Polish scientists will be provided as part of the PLGrid infrastructure.

More about EuroHPC JU at: *https://eurohpc-ju.europa.eu/*.

Karolina Supercomputer, Czech Republic. Image: EuroHPC JU

Vega Supercomputer, Slovenia. Image: Atos

EuroHPC PL – the Polish part of the EuroHPC JU

At the beginning of 2021, the National Supercomputing Infrastructure for EuroHPC JU – EuroHPC PL project was launched. It aims to deliver the modern infrastructure, particularly supercomputing resources, specialised accelerators and laboratories' services for academia, industry and society.

The project is run by the Consortium consisting of 7 members:

1. AGH University of Science and Technology in Krakow - Academic Computer Centre CYFRONET – Consortium Leader,
2. Institute of Bioorganic Chemistry of the Polish Academy of Sciences - Poznań Supercomputing and Networking Center,
3. Gdańsk University of Technology - IT Center of the Tricity Academic Computer Network,
4. Wrocław University of Science and Technology - Wrocław Center for Networking and Supercomputing,
5. National Center for Nuclear Research,
6. Institute of Theoretical and Applied Informatics of the Polish Academy of Sciences,
7. Center for Theoretical Physics of the Polish Academy of Sciences.

EuroHPC PL was established to facilitate and expand access to large-scale data processing for both scientists and entrepreneurs and thus increase the competitiveness of Polish entities compared to Europe and the world. The essential element of the infrastructure being built is the computing power provided by supercomputers. Their use is a crucial tool for conducting scientific research in many fields and significantly accelerates and reduces the costs of research processes. It is especially visible at the initial stage of the research process, in which the set of possible research directions is so large that their classic verification (e.g. in laboratories) would not be possible due to both time consumption and costs. Supercomputing resources will be supported by the architecture of quantum algorithms and neuromorphic accelerators. This will significantly increase the possibilities of data analysis and visualisation. Additionally, the project provides a wide range of user support and training.

Project tasks are performed within four laboratories. The main goal of the Laboratory of Modeling and Parallel Data Processing in a Pre-Exascale Environment is to design and implement advanced large-scale computing systems. The Laboratory for the Application of Hybrid Computing addresses the needs related to the application and implementation of practical use of quantum computers. In turn, the tasks of the Laboratory for the Application of Supercomputers in Medicine focus on the practical use of high computing power in medical applications. The Laboratory of Energy and Computing Efficiency of HPC Software works on the development of techniques to reduce machine operating costs, such as energy consumption and cooling demand, as well as works on optimising software running in a supercomputing environment.

More information about EuroHPC PL is at: *https://eurohpc.pl*.

The LUMI supercomputer

LUMI - No. 1 in Europe, No. 3 in the world

In June 2022, during the ISC High Performance fair in Hamburg, the TOP500 list of supercomputers with the highest computing power was announced.

Supercomputer LUMI, which was installed in the CSC Data Center in Finland, ranked the third place, and at the same time, it took the lead in Europe. This undoubted success in the HPC (High-Performance Computing) environment is the result of the cooperation of 10 countries, including Poland, which formed a consortium operating under the wings of a larger initiative, the EuroHPC Joint Undertaking. Importantly, thanks to the activities coordinated by ACC Cyfronet AGH and the Ministry of Education and Science, Poles have already gained access to LUMI resources.

LUMI among the world leaders in computing power and energy efficiency

Together with other petascale and pre-exascale systems in which EuroHPC JU is involved, LUMI is a part of a pan-European infrastructure, which will also include fully exascale computers in the coming years. Obtaining a place at the TOP3 of supercomputers is a great success for European countries, because it shows the possibility of competing with systems from the USA, China and Japan. It is worth noting that until the High-Performance Linpack (HPL) tests were carried out, in which LUMI showed the actual computing power of 151.9 PFlops, only a part of the GPU partition was installed. When fully installed, the capacity will increase to 375 PFlops, and the theoretical capacity will increase to 550 PFlops. Such performance has become possible thanks to the LUMI hardware architecture:

- the GPU partition is composed of 2560 nodes, each of which contains a 64-core AMD Trento processor and four AMD MI250X cards,
- a single MI250X card can deliver 42.2 TFlop/s of performance in High-Performance Linpack tests,
- each GPU node contains four 200 Gb/s network connection cards,
- the addition to the GPU partition is the CPU partition, which uses the 3rd generation AMD EPYC™ 64-core general purpose processors.

However, during the design and construction of the LUMI infrastructure, a strong emphasis was put not only on the computing power, but also on the efficient use of energy. For this reason, the LUMI supercomputer also took third place on the Green500 list of supercomputers with the highest energy efficiency, calculated as the ratio of computing power to electricity consumed. In addition, it is worth emphasizing that 100% of the LUMI's power supply comes from renewable sources (hydroelectric power plants), and the waste heat removed from the machine is used to heat the surrounding buildings. Thus, LUMI is now a flagship European example of implementing the idea of sustainable development in the construction of supercomputing systems.

More information about LUMI at: *https://lumi-supercomputer.eu*.

The representatives of ACC Cyfronet AGH during the inauguration of LUMI.
From the left: Mariusz Sterzel, Head of the Quantum Computing Laboratory and Marek Magryś, Deputy Director of ACC Cyfronet AGH for HPC

Access to LUMI for Polish scientists

ACC Cyfronet AGH coordinates work related to providing Polish scientists with the resources offered by LUMI. The aim is to ensure an efficient and user-friendly environment for working with a supercomputer, based on good practices developed within the PLGrid infrastructure.

From autumn 2021 to spring 2022, Cyfronet organised two competitions for pilot grants for computing on the LUMI supercomputer for Polish scientists. The pilot phase projects aimed to test the CPU partition (with general purpose processors) and the GPU phase (with graphics processors). Poland, as a member of the consortium, submitted five projects of potentially significant importance, requiring the use of many computing cores and enormous disk resources. Access is made via the PLGrid Portal (*https://portal.plgrid.pl*), and additional information is available at *https://cyfronet.pl/lumi*.

Cyfronet is the leader of projects included on the Polish Research Infrastructure Map

Among the strategic infrastructures included in January 2020 on the Polish Research Infrastructure Map there are two projects proposed by ACC Cyfronet AGH as the initiator and coordinator of the PLGrid consortium: *National Supercomputing Infrastructure for EuroHPC* and *National Cloud Infrastructure PLGrid for EOSC*.*

The aim of the **National Supercomputing Infrastructure for EuroHPC** program is to build a computing infrastructure for scientific research on solutions that meet the current and future needs of Polish society, the scientific community and the economy.

The infrastructure will be based on modern supercomputing systems enabling the implementation of both traditional simulation tasks and data analysis using artificial intelligence methods. The production computing systems built within the project will be among the world's leading supercomputers. In addition to the computing infrastructure, the project will also provide access to specialized training and expert technical support for users from science and economy, as well as the necessary procedures concerning allocation and accounting of used resources.

The project fits directly into the framework of the international EuroHPC – European High-Performance Computing initiative published in the European Commission Communication COM/2018/08 final – 2018/03 (NLE). EuroHPC is a project aimed at creating a European system of high-performance exascale computers, unique on a global scale, based on technologies developed in Europe. The project will be implemented by the PLGrid Consortium.

As a result of the implementation of the National Supercomputing Infrastructure for EuroHPC project, an infrastructure for conducting research for the needs of science, economy and society will be created, benefiting from the latest HPC technologies developed within the international EuroHPC cooperation. The infrastructure will offer services in the fields of massively parallel computer simulations, highly efficient processing of data sets, the use of artificial intelligence methods, software and high productivity tools, including data visualization, and user support and training. In addition to the main computing systems, the infrastructure will also include smaller test and research systems to verify new processor, accelerator, memory and network technologies in the context of using them to build production systems, as well as conducting research and development in the field of effective HPC infrastructures.

The services will be offered via the PLGrid infrastructure, which integrates most of the computing resources available in the country, what will facilitate the process of resource allocation and user support. Integration will also be implemented with European pre-exascale and petascale systems made available under the Euro-HPC program, in particular with a machine built by the LUMI consortium, of which Poland is a member.

National Cloud Infrastructure PLGrid for EOSC is a program for the use of cloud resources for scientific research that meets the current and future needs of the Polish society, the scientific community and the economy. The scope of this research includes data, infrastructures and data processing platforms, as well as effective algorithms and dedicated applications.

The program is based on the requirements of the society, economy and Polish researchers, in particular those cooperating within international research groups. These groups require advanced environments for the integration of distributed resources: software, infrastructures and dedicated services. These requirements can only be met by advanced IT technologies combined with computing, storage and data resources. Cloud technologies enable the interaction of all these elements within a flexible ecosystem.

National Cloud Infrastructure PLGrid for EOSC is part of the ecosystem of the European Open Science Cloud (EOSC, Declaration of 26.10.2017). Poland is currently developing two key components of this federated, globally available and multidisciplinary environment: Onedata – a system for unified data sharing and management, and the EOSC Portal. As part of the European ecosystem, the PLGrid National Cloud Infrastructure will offer trusted and open environments for users throughout the data lifecycle. This will allow scientists, the economy and society as a whole to publish, search, use and re-use the collected data, tools, software and other results.

The research planned within the National Cloud Infrastructure PLGrid for EOSC will allow for the development, validation and, consequently, the provision of services (general and dedicated), and thus the use of modern technologies and effective techniques for management, processing and reusing data by scientific communities, economic entities and society.

To this end, the Infrastructure will provide:

- solution technologies for distributed environments, including cloud environments, covering service management in a distributed environment, automation of complex processes, integration of research platforms and infrastructures,
- safe data sharing and management based on national technologies, in accordance with EOSC standards,
- research on the specific needs of users in the field of large-scale data processing in a distributed environment, including: „close data" processing, using the „data lakes" paradigm with new analysis models, scalable resources in a distributed environment,
- verification of solutions prepared for the needs of society, science and economy in advanced applications,
- a catalog of general cloud services for the economy, science and society as a result of research and development works in cooperation with EOSC.

The first stage of this work is currently carried out within the resources of the PLGrid infrastructure.

**Material from the "Polish Research Infrastructure Map" brochure of the Ministry of Science and Higher Education.*

The main goal of the project is to provide an open and fully operational FAIR EOSC ecosystem. The project aims to expand the EOSC infrastructure with components supporting the FAIR paradigm (Findability, Accessibility, Interoperability, Reusability).

The aim of the project is to define and develop a new and innovative European Master's degree program focusing on HPC. The program is designed to equip students with HPC competences and knowledge required by academia and industry.

The PROTEUS-RS project aims to implement models and a number of numerical simulations based on high-performance computer architectures for the design of the manufacturing process of metal long elements.

Through the development of the EOSC Portal, the project aims to extend the EOSC ecosystem, integrate existing scientific infrastructures and initiatives, and engage new domain and national research communities.

The main aim of the project is to deliver transparent and unified access to Earth Observation Data, primarily satellite data, in order to support food production and food and agriculture industry.

The goal of the project is to create a computational medicine centre in Krakow. The Centre will be the main driver of European progress in this fast-growing sector, developing advanced engineering methods for the prevention, diagnosis and treatment of diseases, and meeting the global need for radically improved healthcare systems.

The aim of the project is to expand the computing infrastructure enabling cooperation in the area of high-intensity problems connecting large volumes of data, which cross the boundaries of a single data center.

The purpose of the EPOS-SP project is to develop and implement assumptions to ensure the sustainability of the EPOS infrastructure produced in previous projects (EPOS-PP and EPOS-IP).

The EOSC-Synergy project introduced EOSC standards for national infrastructures in nine European Union countries. It was done by harmonizing policies and expanding access to research infrastructures, scientific data and domain services.

The goal of the PRACE-6IP project was to implement new solutions and maintain the operationality of the PRACE environment in the area of European HPC computing infrastructures.

The PRIMAGE project aims at creation of a Clinical Decision Support System (CDSS) for the treatment of cancer (neuroblastoma, glioma) in children. Patients' data was used in the multi-scale computational models of cancer designed to define disease biomarkers. The created CDSS system helps oncologists both in diagnosis and in predicting of disease progression and treatment effectiveness.

The main goal of the EuroCC project was to create a European technological support system based on closely related National Competence Centers in individual European countries, which allowed the academic environment, the enterprise sector (especially SMEs) and public administration to benefit from the available expertise, experience and resources of EuroHPC.

The EOSC Enhance project aimed to build an improved, more integrated version of the EOSC Portal, which enables improvement and extension of solutions that make it easier to find European scientific services and open science data sets.

The aim of the EOSC-hub project was to prepare the launch of a production infrastructure for open science in Europe and the practical application of solutions developed as part of the EOSC-Pilot project to a real large-scale environment scattered across most European countries.

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

The goal of the project is to develop and provide production services for storing, accessing, securing data and managing metadata, as well as integrating solutions for processing large and complex data volumes on the basis of a distributed e-infrastructure.

Within the project the construction of unique research laboratories based on the national PIONIER fiber optic network is planned. The main goal of the project is to build and make available platforms for research units, entrepreneurs and other entities interested in conducting scientific research and development works based on a new, nationwide research infrastructure.

The direct goal of the project is to create a specialized e-infrastructure for data processing, enabling the optimal use of specialized and new generation services to stimulate new areas of application in science, economy, education and social life.

As part of the project, the functionalities of the EPOS-PL research infrastructure will be increased. A new Research Infrastructure Center (Center for Research Infrastructure of Satellite Data – CIBDS) will be established, a new test site (Geophysical Safety System for mining protection pillars) will be created, and the WNiP will be established: „IT Platform for Research with Artificial Intelligence Methods (EPOS-AI)".

The goal of the project is to create computing infrastructure services and data storage services for the purposes of the PRACE project, within six dedicated laboratories: 1) L. of HPC and cloud processing, 2) L. of access to processing infrastructure, 3) L. of service management and monitoring, 4) L. of data management services, 5) L. of distributed data management and transparent access to data, 6) L. of infrastructure security.

The main objective of the project was to provide satellite data coming from the Sentinel satellites of the Copernicus network. The project created an infrastructure for automatically downloading data directly from satellites, its secure storage and sharing for the purposes of science, administration and training.

The project aimed at building the national research infrastructure for solid Earth Science and its integration with international databases and services implemented under the European Plate Observing System (EPOS).

The objective of the project was the development of the specialized technological competence centre in the field of distributed computing infrastructures, with particular emphasis on grid technologies, cloud computing and the infrastructures supporting calculations on large data sets. As a result, a great computing power and huge storage for digital data were offered to users. They also obtained access to a set of basic and end-user services.

The aim of the project was integration of selected services available in the PIONIER network, and the development of the new services, e.g., with the increased reliability and security.

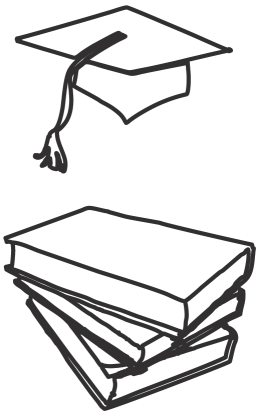
The aim of the project was the development of 21 environmental science data communication networks providing the scientific institutions across the country with access to a modern and secure network infrastructure, supporting the research and development of Polish groups of scientists.

The project envisaged the creation and launch of five services running on the basis of the PIONIER network. These services included: video conferencing services, eduroam services, campus services, universal archiving services and scientific interactive HDTV services.

ACC Cyfronet AGH at the ISC High Performance in Hamburg

Cyfronet acted as an exhibitor during one of the most important events for the environment of suppliers and users of high-scale computing solutions. ISC High Performance (May 29, 2022 – June 2, 2022) attracted 137 exhibitors and about 3,000 participants from all over the world. The exhibitors included companies – technology giants, such as Hewlett Packard Enterprise, Intel or NVIDIA, as well as leading initiatives and supercomputing centers, including i.e. EuroHPC JU, LUMI, Barcelona Supercomputing Center. The stand of ACK Cyfronet AGH was one of the few Polish accents during the fair. The main activities of Cyfronet included: maintaining relationships with existing partners, establishing new contacts with potential hardware and software suppliers, and assessing new technological solutions that can be implemented as part of the managed infrastructure. Cyfronet's participation in ISC was also related to promotional activities under the projects: National Competence Centres – EuroCC and National Supercomputing Infrastructure for EuroHPC – EuroHPC PL.

Work of young scientists in Cyfronet



The annual contest for the best PhD thesis conducted with the help of computing resources of ACC Cyfronet AGH is a tradition in our Centre. The scientific value of the submitted doctoral dissertation is assessed, as well as the possibility of its practical application and the scope of use of computing resources and disk storage in Cyfronet. In recent years, the Contest has become an important event promoting research conducted by young scientists. To subsequent editions of the Contest participants submitted many PhD theses focused on variety of scientific problems in chemistry, biophysics, physics and computer science. Also the utilisation of the resources varies, as the contesters use different tools running on a wide range of computing architectures offered by Cyfronet.

The laureates of the Contest are invited to give a talk during Cyfronet's Open Day. We are honoured to present here selected interviews with the Contest participants.

Join the next Contest edition!
http://www.cyfronet.pl/konkurs

The laureates of one of the previous Contest editions

Dawid Dułak

The interview with the author of the PhD thesis:
"The mechanism of amyloid transformation based on the change of hydrophobicity distribution"

How did your adventure with science begin? Can you share your story with us?

Conducting doctoral research resulted from a passion for Biophysics. The direction of my interests began to take shape in the subject of "Bioinformatics" during my graduate studies (Computer Science at the Jagiellonian University). I developed my medical knowledge through engineering studies in Medical Physics at the AGH University of Science and Technology. I wrote an engineering thesis related to protein folding. My initial understanding of the topic and my interest in research at this stage turned into regular cooperation with the Department of Bioinformatics and Telemedicine of the Jagiellonian University.

It had an impact on subsequent decisions ...

For ten years, I have been cooperating with the Department of Bioinformatics and Telemedicine of the Jagiellonian University, conducting research on the topic of protein folding in the context of the model created as part of the work of the Department mentioned above. What happened next was a natural consequence of the choices made earlier. I wanted to do research in an area that I had been working in for years. It was an open question which fragment or element of research to choose as the guiding axis of the doctoral dissertation.

Protein folding is a highly fascinating process in which many amino acids combine, always taking the same form (as long as the amino acids are arranged in the same order). Although the body performs this operation very quickly and very often, we cannot recreate it. I was researching this problem in the Department of Bioinformatics and Telemedicine of the Jagiellonian University. The leader of the group which created an IT model of protein folding was professor Irena Roterman-Konieczna. My goal was to improve this model by understanding the reason why some proteins fold incorrectly (the so-called amyloids, which contribute to the development of Alzheimer's disease, among others).

How do you remember working with the research team?

I was most involved in the work from the IT side. Contacts with the team allowed me to quickly and easily understand the nuances of the biological, chemical and physical world. I was most pleased with exchanging ideas, research results and reciprocal assistance in research work.

Was your computer hardware sufficient during the work?

For this type of calculations (folding simulations) scientists use computer modeling. They are computationally very demanding due to the number of elements (in this case, atoms) and the forces

The TAU protein, i.e. the amyloid forming fibrils (many chains connected together) in the intercellular space (Alzheimer's disease)

acting between them. The appropriate software had to be found first. I used the packages included in the GROMACS program and the program written by Mrs Małgorzata Gadzała. While the first of them is pretty popular and you can easily find information about it, let me briefly describe the second one. It enables the simulation of an aquatic environment as a field and its effect on protein folding. Modelling required alternating operation of both applications in different number of repetitions, so I performed about 500 simulations for one protein. During several years of research work, I simulated the folding of several dozen amino acid sequences (some of them a few times) that form natively specific proteins. I chose five of them for my doctoral dissertation due to their amyloid form. The main computers I used at the stage of computing were the Zeus supercomputer and later the Prometheus. An ordinary personal computer – even equipped with a top-class processor, graphics card and RAM – would perform a single simulation for many days, weeks or even months. So it is great that such machines are in Krakow, and we can use them for such necessary – in my opinion – scientific research.

The cooperation with the ACC Cyfronet AGH has been going on for years, and an essential application for research was created there. Only thanks to the supercomputers – which are crucial in molecular dynamics simulations – I could conduct the study described above. Cyfronet's team has always been available and helpful with all kinds of challenges, so I count on further, fruitful cooperation and the possibility of using the centre's resources.

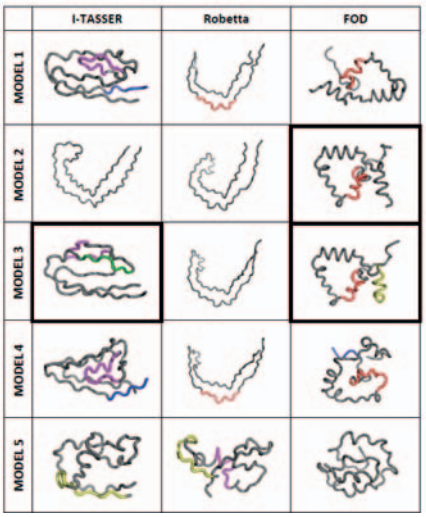
We are glad that we could help, and in the future, we invite you to contact Cyfronet! Based on your own experience, what advice can you give to young scientists?

It may sound like a fairy tale. A person starting a scientific career should get carried away by dreams and imagination, because they generate passion. And then, well... You have to persistently and patiently pursue new goals.

Do you plan to continue your research?

As I already mentioned, this topic is extremely interesting for me. Despite the rapidly developing area of artificial intelligence, which can essentially predict the structure of a protein based on its amino acid sequence, it cannot answer the question of what precisely this mechanism looks like. Therefore, there is still a long way to go to fully understand the natural process of protein folding. I will undoubtedly continue my research in this area, so who knows, maybe the time will also come for a higher level of 'initiation'.

Thank you for your time, and we wish you the fulfilment of your scientific plans.



Top 5 single chain structures (models) based on the TAU protein sequence generated using three different tools (I-TASSER, ROBETTA and DRIPPY). Those that potentially should not form fibrils are bolded. The structures (models) generated with the DRIPPY tool were obtained with the use of software available within the framework of the PLGrid infrastructure in the ACC Cyfronet AGH

Michał Koziarski

The interview with the author of the PhD thesis:
„Imbalanced data preprocessing techniques utilizing local data characteristics"

Can you synthetically characterize the subject of your dissertation?

The subject of my thesis was imbalanced data classification, and more specifically – development of new data preprocessing algorithms capable of reducing the negative impact of data imbalance on the quality of classification. We define data imbalance as a situation, in which one of the classes that we wish to classify is less represented than the other classes – in real datasets data imbalance is ubiquitous, and the question is usually how severe is the imbalance, and not if it is present at all.

Majority of the traditional classification algorithms are ill-equipped for handling imbalanced data and tend to prefer the more represented classes at the expense of the prediction quality on less represented classes – which is the opposite of the desired behavior. One of the approaches for reversing this trend is using data preprocessing techniques, that is algorithms manipulating the training data, either by reducing the number of observations from more represented classes, or by creating synthetic observations from less represented classes – development of such methods was the subject of my thesis.

Please tell me how did your interest in machine learning arise?

My interest in machine learning (and more broadly, artificial intelligence) started at the beginning of my university education, basically from the first contact with the subject – from my point of view it was the problem that, if solved, would help to solve all the other problems (meaning, being equipped with strong enough artificial intelligence algorithms we would be able to easily solve every other problem).

The development of artificial intelligence is perceived as an extremely important trend in science. Do you expect a spectacular breakthrough or shall we be dealing with evolution counted in decades?

The spectacular breakthrough was, essentially, already done – it wasn't just in the AGI (Artificial General Intelligence), which is, broadly speaking, the representation of artificial intelligence in movies and literature, but in practical applications of machine learning to specific problem domains. And success stories for that can be observed all around us, starting more prosaically from the contents of our smartphones, through different types of medical diagnosis and drug discovery, up to fundamental breakthroughs such as protein folding prediction – just to name a few examples.

Already at this stage, your algorithms support oncological diagnostics – I would like to ask for a few sentences about this cooperation.

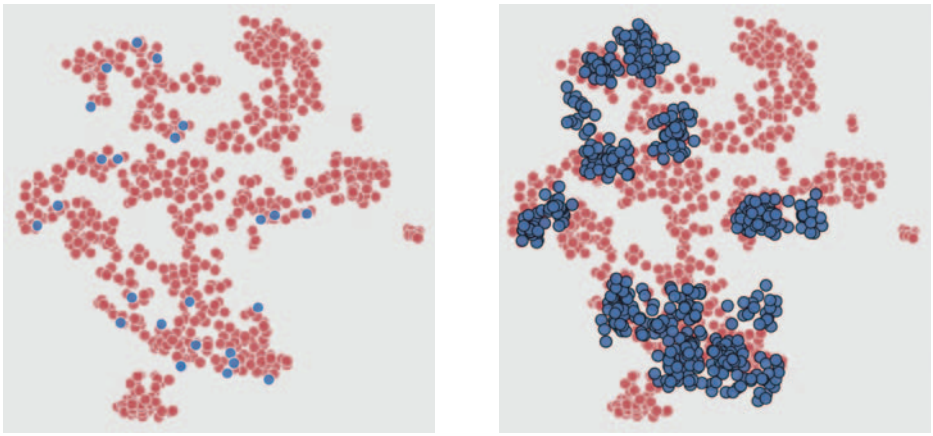
In the application section of my thesis I focused on automatic cancer detection on histopathological images, that is microscopic scans of tissue. The problem is particularly important due to the fact that we have a shortage of experienced histopathologists, and any improvement in the speed of diagnosis would have significant practical ramifications.

The problem boils down to classification of different tissue regions as either healthy or cancerous, and in the second case – prediction of the severity of the cancer. It's a highly imbalanced problem, because healthy tissue is most prevalent, and especially some of the types of cancer tissue are very rare. Additionally, from the point of view of machine learning, the problem is interesting, because both: whole datasets and individual scans are very large. In addition, we deal with significant levels of label noise, and

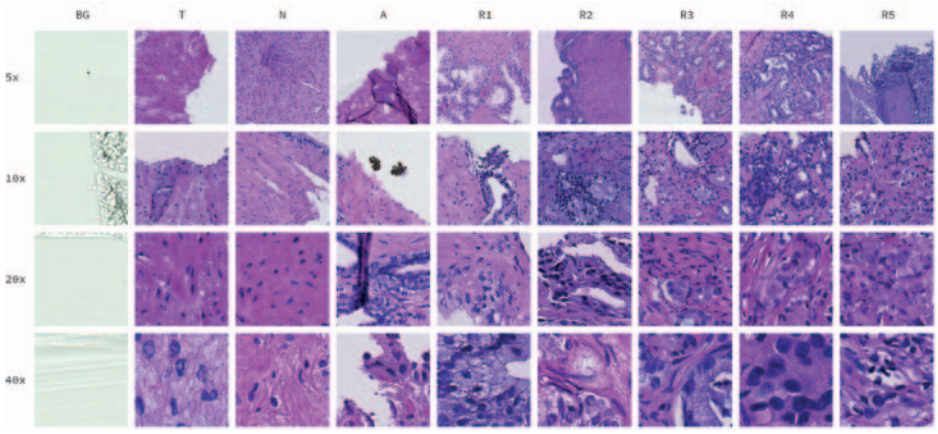
the entire classification process requires the usage of convolutional neural networks. I had the possibility of applying some of the algorithms designed in my thesis in the final pipeline that we developed.

Did Cyfronet meet your expectations in terms of availability and quality of computing resources?

Without the resources provided by Cyfronet my thesis would simply never come to light, at least not in its current form. Provided resources made it possible to design some of the more computationally expensive algorithms, and to thoroughly experimentally test the rest of them.



Left: An example of an unbalanced dataset, with the color of the observation class. Traditional classification algorithms cannot cope with highly unbalanced data. Right: a set after adding synthetic observations generated with one of the developed algorithms



An example of fragments of histopathological photos from different classes, the classification of which was the subject of the application part of the doctorate. This type of medical data is often highly unbalanced, with a significantly smaller number of photos containing neoplastic lesions

Mercedes Kukułka

The interview with the author of the PhD thesis:
„Theoretical investigations on absorption and luminescence properties of selected organic species for potential applications in optoelectronics"

How would you introduce your research to someone not involved in the field?

The main goal of my research was to identify and understand, based on quantum-chemical calculations, the relationships between the structure (geometrical and electronic) and photophysical properties of the selected organic systems that can find potential application in optoelectronics. For simplicity, my research allowed me to identify structural factors for several groups of chemical compounds that determine whether they absorb or emit light.

Is modern photophysics and photochemistry already applied in so-called everyday life, or are we at the stage of basic research?

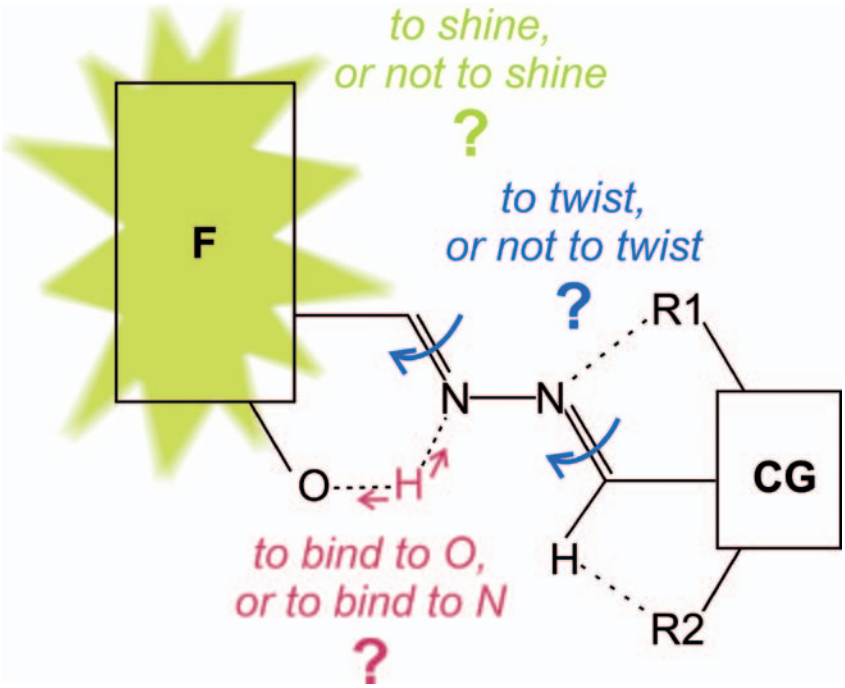
Of course, it is! It is clear that it takes a lot of time from discovery to real-life application, but in recent years, we have witnessed very rapid progress in the field of new technologies. Let us take a look at the development of the technologies used in display screen production. Almost a century has passed from a discovery of liquid crystals (in 1888) to a creation of the first liquid crystal display (LCD), and another three decades to mass-scale production of screens with a diagonal longer than 10". The story of screens based on Organic Light-Emitting Diode (OLED) technology is much shorter. The electroluminescence phenomenon of organic compounds was observed for the first time in the mid-20th century, the first efficient device was demonstrated in 1987, and OLED-based screens have been commercially produced, similarly to LCD screens, since the turn of the 20th and 21st century. Nowadays, in the market, we can also find Quantum dot Light-Emitting Diode (QLED) screens based on quantum dots. In this case, the idea appeared in the 1990s and the first QLED TV sets were already available in 2013.

The competitive, developing global industry of the 21st century is still in search of new materials that have better properties, or that are more effective. Research centers work more and more often in close cooperation with industry, which is partially supported by the government. Companies that carry out research and development, benefit from some tax credits; it is also easier for scientists to get financial support for research when their studies have an application character.

The appearance and dissemination of theoretical studies were also indispensable for rapid development in the research on new materials. The analysis of *in silico* research is fundamental to a meaningful interpretation of experimental data, and also allows us to gain an in-depth understanding of the relationship between the molecular structure and system properties, which is, in turn, crucial for the design of new materials with desired properties.

What further challenges can you see in the future and does the access to high performance computers change anything in this picture?

The effect of the interaction of light with chemical compounds is not always easy to predict, and quantum-chemical calculations for the electronic excited states are still challenging. Theoretical description of the excited-state properties often requires the usage of highly accurate multireference methods, which are computationally expensive, and thus such computations are limited to smaller systems. Moreover, machine learning techniques are being applied more and more often for discovering the relationships between the structure of chemical compounds and their optical activity and photochemical properties. In both cases, the access to high performance computers is indispensable.



Photophysical behaviour of the considered Schiff base ligands. F is fluorophore, CG – controlling group, R1 and R2 – substituents attached to the controlling group

Has Cyfronet provided sufficient support in terms of availability and quality of computing resources?

My research would not be possible without Cyfronet. Its infrastructure supports me with the access to scientific software, as well as the necessary computing resources. It is worth noting that the Helpdesk is really supportive – I always can count on Cyfronet staff for support, and their reaction time for every problem was immediate.

Do you have any advice or insights for students interested in pursuing doctoral studies?

Doctoral studies are a very time-consuming step in the young scientist carrier that also influences his/her personal life. It is extremely hard to finish it without passion and interest in the topic, if possible at all. The other important thing is to note that doctoral studies are more like a marathon, not a sprint. Usually, we do not get results easily, not every experiment finishes successfully; a lot of patience and persistence is needed. Finally, you do not only conduct research on the selected topic during doctoral studies, but you also develop your skills, familiarize yourself with the scientific environment, become more independent, and deal with challenges. This experience is of great value not only in the further work as a scientist, but, in general, in everyday life.

Ewa Machalska

The interview with the author of the PhD thesis:
„Research on amplification of Raman signal of optical activity of molecular systems"

What prompted you to carry out research based on the registration and analysis of ROA spectra to the greatest extent?

I became interested in Raman optical activity (ROA) and electronic circular dichroism (ECD) during my master's thesis, when I focused on an experimental and theoretical explanation of the aggregation process of chiral carotenoids. Then, the willingness to continue the research I have started and the possibility of the practical application of ROA spectroscopy in many fields of science, including biomedical areas, prompted me to undertake doctoral studies at the Faculty of Chemistry of the Jagiellonian University and to continue solving research problems of ROA method, which were not allowing to fall into a routine. Considering the potentially substantial impact of chiral change on the desired biological and healing properties of molecules, the aspect of practical use of chiroptical techniques, including in the pharmaceutical industry, is of particular importance to me.

Can Raman scattering spectroscopy be still considered an area with a relatively low level of exploration?

In my opinion, in recent years, Raman scattering spectroscopy, despite being based on a relatively weak effect, has strengthened its level of exploitation in many areas of science. The clear application nature of Raman spectroscopy is visible in biomedical research, e.g. in the analysis of proteins, sugars, lipids, and blood components, including erythrocytes and leukocytes. A group of scientists have learned to deal with some of the disadvantages that Raman spectroscopy undoubtedly brings, by developing new strategies to enhance the Raman signal obtained.

What was the biggest challenge in your research, and how did you deal with it?

Undoubtedly, the most time-consuming, but also intriguing challenge for me was the research on several vitamin B12 derivatives, on the example of which it was possible to discover and explain the ECD-Raman effect, i.e. the interference of the phenomenon of electronic circular dichroism and Raman scattering, which dominated the natural resonance cobalamin ROA spectrum. After many months of work in the laboratory, generally consisting in the registration of resonance-enhanced ROA spectra of cobalamin analogues, together with the supervisor of my doctoral dissertation, Prof. Małgorzata Barańska and Dr Grzegorz Zając, we managed to develop several experimental and theoretical strategies to eliminate the ECD-Raman effect on resonant ROA spectra. Moreover, meticulous studies of vitamin B12 derivatives have led us to the conclusion that obtaining a reliable ROA resonance spectrum of coloured substances is possible, but only if the ECD-Raman effect is taken into account. Otherwise, if we don't take into account the potential occurrence of the ECD-Raman phenomenon during the measurements of the resonance ROA spectra, it may lead to

misinterpretations of chiroptical spectra and wrong conclusions regarding the molecular structure of not only cobalamins, but also many other chiral compounds.

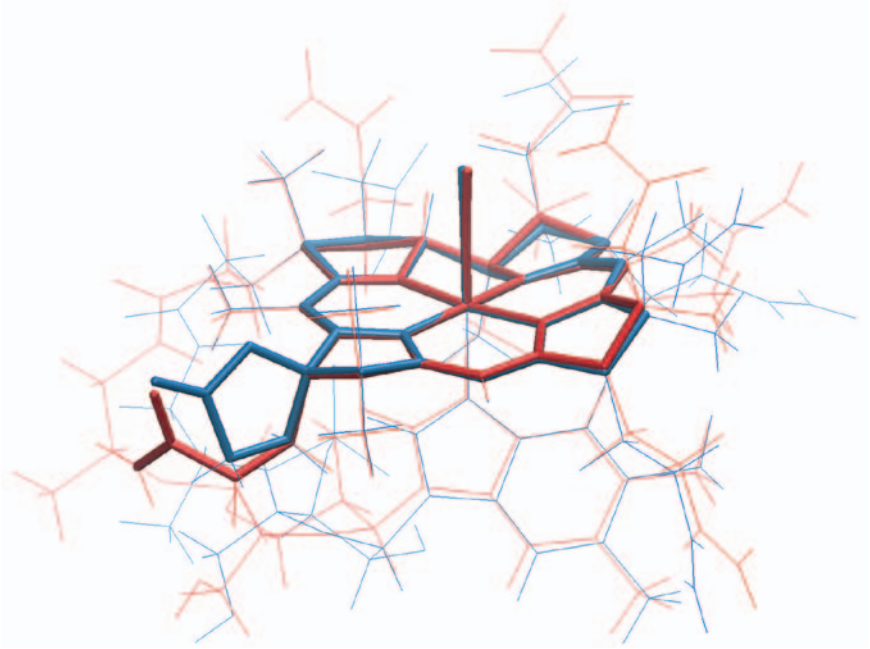
How did you use the resources provided by Cyfronet? What were they most helpful in?

In the research that formed the basis of my doctoral dissertation, through the PLGrid infrastructure, I used the computing resources of the Prometheus supercomputer. At this point, I would like to emphasize that effective quantum-chemical calculations would not be possible without the fruitful cooperation with Dr Grzegorz Zając from the Jagiellonian Center for Drug Development in Krakow.

Undoubtedly, the computational resources provided by Cyfronet were most helpful in explaining the resonance-enhanced ROA spectra of vitamin B12 and its derivatives. The performed calculations, on the one hand, allowed us to obtain ROA spectra, the profile of which was in excellent agreement with the experimental results. On the other hand, the spectral simulations were an undeniable confirmation of the double-sign character of the experimental ROA spectra.

What advice could young scientists who are just starting or planning to start working on their doctorate receive from you?

I think that the best advice I can with certainty give all candidates starting their doctoral studies is not to be afraid of the new challenges posed by science. The seemingly interdisciplinary nature of the research undertaken may, on the one hand, be an extremely difficult and terrifying task for every young scientist. Still, on the other, after some time, it will allow them to create a fresh, different approach to research problems. It is possible to achieve this mainly when PhD students take up topics that fascinate and enjoy them and at the same time, are a robust foundation for creating new ideas that these people can use in the future.



Comparison of the optimised structures of the most stable conformers of vitamin B12 ((CN) Cbl, blue) and its derivative ((CN) 13-epi-Cbl (e-lactone), red). The cortical ring and c side chains are bold

Bartłomiej Rzeszotarski

The interview with the author of the PhD thesis:
„Charge and spin transport in silicene-based nanosystems"

Due to its structure and properties, silicene is popularly compared to graphene. What made you focus your research on silicene-based nanosystems? In what fields can its use outweigh the use of graphene?

Silicene has built-in spin-orbit interactions that are strong enough to control the electron's spin or achieve a topological insulator state in the structure. This is due to symmetry breaking in the geometry of the arrangement of atoms, where two subnets are distinguished – upper and lower. Such an arrangement of silicon atoms in silicene is natural, i.e. in such a configuration, we obtain the minimum energy of the system due to bonds (the planar structure is unstable). However, this symmetry is not broken in simple single-layer graphene, and the spin-orbit interactions are negligible. To induce spin-orbit solid interactions in graphene, you have to break the symmetry by hand, e.g. by doping, rolling into nanotubes, or stacking several twisted graphene monolayers. The simplicity of the silicene system influenced its choice. Moreover, the silicon technology has already been very well known for years, which is why we saw the great potential of this material both in the context of the possibility of conducting experiments and possible applications in spintronics as elements of building quantum computers. Our attention was drawn to a working silicene-based transistor, the investigation of which was published in 2015 in the Nature journal – this sparked our willingness to further research on this material.

What do you think are the following challenges facing scientists researching silicene?

Stabilisation of silicene is very difficult – it is a material that oxidises quickly. In addition, the substrate on which the silicene is deposited should not affect its properties, i.e. the two-dimensional electron gas (for example, a metallic substrate makes silicene no longer a semiconductor). Currently, the work is underway to find a suitable substrate material that will not significantly modify the strand structure and, at the same time, stabilise the system so that it can work under the conditions imposed by the experiment. Although the current trend is dominated by the experimental work on twisted graphene double-layers, it does not exclude the emergence of new experimental results for silicene. However, there are still many challenges on the way to practical implementations.

What aspect of scientific work gives you the most excellent satisfaction?

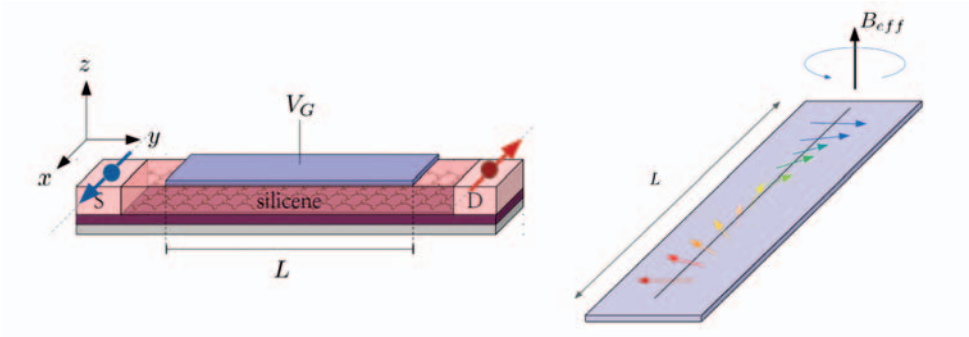
The greatest satisfaction gives me the understanding of the essence of the phenomenon under study to such a level that I can talk about it simply and understandably to everyone, even those not related to science. After many hours of work, a kind of intuition develops, which suggests some images for analogy and reasoning that I can pass on to others. Then, as a scientist, I feel fulfilled because I pass the knowledge on.

How did Cyfronet's resources contribute to implementing the part of your research related to calculations and their analysis?

All calculations were performed with atomic resolution, which generates a substantial numerical cost. Without the resources of Cyfronet, the solution of the systems included in the dissertation for the space of the examined parameters in the appropriate time (the duration of the doctorate) would be unattainable. Therefore, in the numerical part, all the credit lies in the infrastructure of the Zeus and Prometheus supercomputers located in ACC Cyfronet AGH, for which I would like to thank you very much.

What could you advise people who are just starting their doctoral studies? What should these people pay the most attention to?

In the scientific work, we use 90% of time for looking for solutions, we do not understand something, we check for errors or look for them. It is a period of "pain" that must be endured. In the end, we overcome this wall and get the results. And the remaining 10% of our time is spent on writing publications, analysing the results and defending them in front of the reviewers. Research work is not simple; it is often riddled with frustration and situations in which our knowledge and skills are challenged. If I were to advise people starting PhD studies, I would first recommend finding joy between the 90% and 10% above. And keep going, talking to others, looking for ideas. We are not alone on this challenging road.



The schematic view of the spin inverter based on silicene. Inversion of a spin is achieved by its precession around the B_{eff} effective magnetic field, which arises due to the intrinsic spin-orbit coupling in silicene. The precession rate can be controlled by the V_G gate voltage

Zuzanna Wojdyła

The interview with the author of the PhD thesis:
„Computational biochemistry studies on the mechanisms of reactions catalysed by 2-oxoglutarate dependent dioxygenases"

How did your adventure with science begin?

I discovered computational chemistry for the first time at the lectures given by prof. Ewa Brocławik at the Jagiellonian University. I liked the riddles that could be solved using quantum-chemical modelling. When I learned that computational chemistry could be employed to study enzymes, I knew this was the direction I wanted to follow. And this is how I graduated with my master's and then doctoral studies.

Which enzymes have you decided to deal with?

The enzymes I studied – 2-oxoglutarate-dependent dioxygenases, are promising yet unexploited targets for industrial application. They catalyse a chemically challenging cleavage and activation of C-H bonds. These reactions are highly selective (few by-products are formed). As a result, they could be used in the production of complex chemical compounds whenever the purity of the final product is essential, e.g. in the pharmaceutical industry.

Together with the research group I selected a number of representatives of this family of enzymes. We wanted to understand how they control the course of the reaction and what factors contribute to the wide range of reactions (hydroxylations, dehydrogenations or cyclizations) catalysed by the family of enzymes, which share the composition of the active site. This knowledge could be used, for example, in the engineering of enzymes to catalyze desired reactions – and that was our long-term goal. The research conducted by our team combines experimental and computational methods. Our cooperation is still developing. The calculations allow to complement and interpret the results of the experiments, and the experiments allow to confront the results obtained *in silico* with reality. I enjoy this way of doing research very much.

What aspect of computational chemistry is the most interesting for you?

My favourite aspect of computational chemistry is gaining the detailed insight into the catalyzed reaction from the enzyme-substrate complex to the end products. It is most compelling to investigate reaction channels that are not observed in experiments. It is valuable addition to experimental research, especially since the frequent "strategy" of enzymes is not only to lower the energy barrier for one particular reaction, but also to hinder alternative transformations. Calculations can show how these transformations are inhibited and what factors promote the "right" reaction.

What was the greatest challenge for you?

For each enzyme I investigated, the most interesting, but at the same time the most challenging step was to pinpoint the effect that sealed the selectivity of reaction. Enzymes can affect the reaction in a variety of ways. They can strictly determine the position of the substrate or engage in specific interactions. Depending on the analyzed enzyme-substrate complex, different effects dominate, and these, in turn, favour the formation of various products. Sometimes it took a lot of imagination and effort to identify this key factor.

What did your work look like?

The modeling of enzyme-catalysed reactions strongly relies on the experimental results – as a starting point I needed the structure of the enzyme-substrate complex. Later on, I frequently confronted the picture of the reaction obtained from the calculations with experimental results. The experimental data can be obtained from the literature, but a more interesting alternative is to compare the *in silico* results with the ones obtained in experiments carried out in parallel by my colleagues from the research group. To at least partially capture the complexity of enzymatic catalysis, many tools are needed, e.g. the Amber package to perform molecular dynamics simulations (to sample the flexibility of the enzyme-bound substrate), Gaussian (to model the reaction), NCIPLOT (to visualise the interactions in the system). I also used (mini)programs written within our team.

Large computing power is indispensable for computational studies of enzymatic catalysis. I used the possibilities offered by the Zeus and Prometheus supercomputers. Thanks to the cooperation with Cyfronet, I could focus on the (bio)chemical aspects of my work, and in case of technical problems, I could always obtain help from the experts via the Helpdesk. It allowed me to use my time efficiently, so it was definitely a good choice.

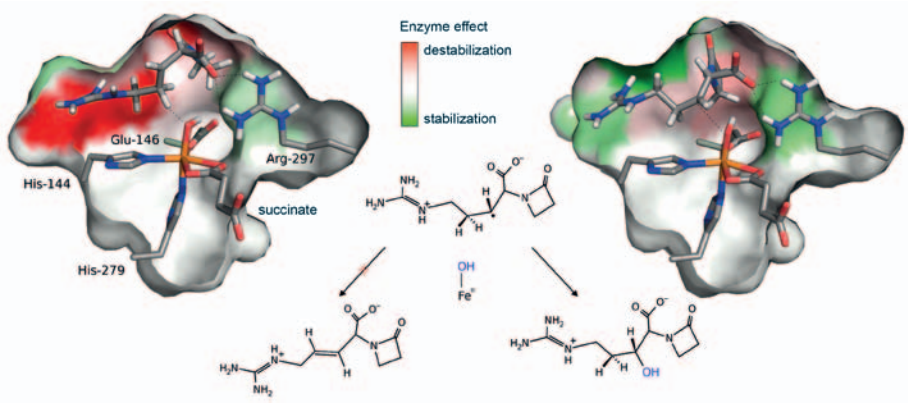
What can a young scientist do when he encounters obstacles?

I think a quote from Thomas Edison would be a good answer: *"I have not failed. I've just found ten thousand ways that won't work"*. This is one of the many quotes that Gaussian shares at the end of computations, and it reflects pretty well the approach that helped me with my research.

What are your plans for the future?

I am planning a new research project, during which I would like to approach the reactions catalyzed by metalloenzymes more systematically – to determine how changes in the environment (the binding pocket) and in the active site affect the course of the reaction. It will require a shift in perspective, so it will be a new interesting challenge.

Thank you for your time.



The binding cavity of the enzyme affects the selectivity of the reaction – the surroundings increase the energy barrier associated with double bond formation (left) and lower the barrier for hydroxylation (right)

Dominik Żurek

The interview with the author of the PhD thesis:
"Accelerating machine learning algorithms and selected population computational intelligence algorithms of arbitrary-precision arithmetics on the GPGPU"

How would you explain your research to someone outside the field?

The main goal of my research was to devise and analyse effective methods to accelerate calculations, which are suitable for parallel processing, from the area of machine learning and computational intelligence. It can be seen in my work that this takes the form of three different aspects, namely acceleration of calculations, artificial intelligence, and computational intelligence, all of which I am going to briefly present.

The concept of artificial intelligence is most often recognised as self-learning machines or IT systems that process information based on human reasoning rules. Artificial intelligence deals with issues which cannot be effectively resolved through the use of algorithms which are based on data modelling, so in order to resolve them, it is necessary to introduce algorithms which contain the earmarks of intelligence. The term "intelligence" should be recognised as the ability for the self-directed adoption to changing conditions – similar to human intelligence, which is a role model for AI. We can say that AI is an attempt to move the human brain's abilities to the programs running on machines (computers). The concept of the "learning process" refers to the ability to make an autonomous change in the system based on obtained experiences, which leads to the improvement of the efficiency of the algorithm. With this definition, we assume that there is a possibility to assess the quality of decisions being made – i.e. the ability to distinguish positive and negative changes. AI is able to acquire knowledge through the extraction of patterns from the raw data. By introducing elements of human intelligence to the algorithm, it is possible to train it to be able to recognise images, understand natural language or perform logical reasoning.

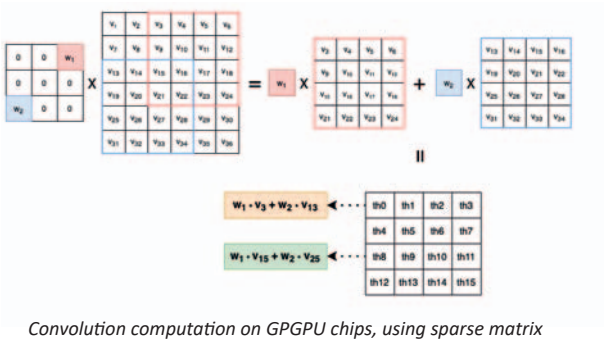
One of the subfields of artificial intelligence is computational Intelligence (CI), which mainly consists in the possibility of adjusting a system to a changing environment. CI places the emphasis on improving and developing applications in the real world. The main feature which distinguishes those groups of algorithms from other algorithms of AI is the lack of a defined model. Instead of this, the algorithm tries to build a model based on the use of training datasets. The important branch of CI's algorithms is formed by "evolutionary algorithms", which are inspired by the natural evolution process of the organisms. By transferring biological inspirations to evolutionary algorithms, the population of the specimens is equivalent to the number of solutions (a specimen = a solution). The evolution of the solution of the algorithm is the counterpart of the changes which occur in the population as a result of mutation (random modification) and recombination (exchange of genetic material), which finally consists of finding better solutions.

The aforementioned groups of algorithms need a huge amount of computing resources in order to be possible to work. This is why it is necessary to devise an effective method that enables faster processing of these calculations, namely the "Acceleration of the calculations".

What made you devote your doctoral dissertation to the issue of acceleration in GPGPU systems?

The GPGPU is dedicated to strong parallel calculations through the calculations conducted in accordance with the SIMD architecture (single instruction stream multiple data). As mentioned above, in my research I focused on the algorithms from

an artificial intelligence and evolutionary group, as a result of their broad spectrum of applications. Both groups of algorithms tend to be parallelising, thus the choice of graphics cards as hardware accelerators seemed to be the most natural.



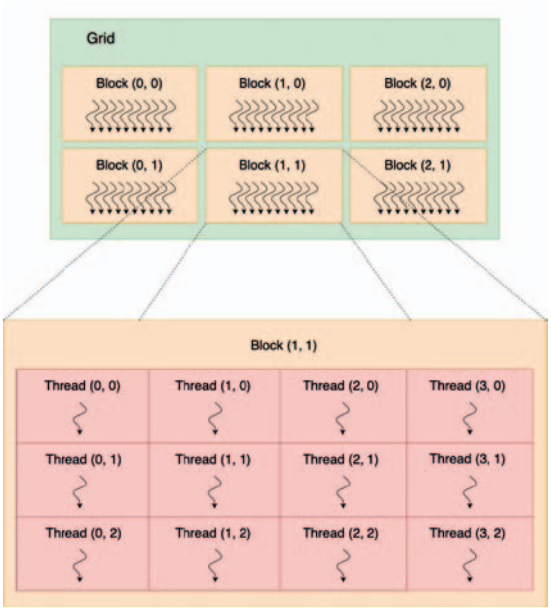
Convolution computation on GPGPU chips, using sparse matrix operations

Let's consider predicting the future. How can research translate into practical implementations? Is there an area that is particularly awaiting a breakthrough in this regard?

In my opinion, this kind of area is definitely neuroevolution, which I am currently working with thanks to knowledge and experience which I acquired during my PhD. Neuroevolution deals with developing neural-networks through the use of evolutionary algorithms (i.e. affects both of the groups of algorithms which were the subject of my research). This neural-network development could be realised on two levels, namely on the level of searching the optimal topology of neural networks or during the process of neural-network training (weight mutation). In both cases, a tremendous computing power is required. Thereupon, the methods which will be able to process more efficiently artificial intelligence and evolutionary algorithms, will certainly translate into practical implementation in this regard.

How do you evaluate the role of Cyfronet in the context of research in the field of acceleration and data processing?

In my research, I used a very effective and modern graphic card, namely Nvidia Tesla V100-SXM2-32GB. Thanks to this GPGPU, for some problems which were undertaken in my PhD, it was possible to achieve an exceptional level of acceleration compared to the older generation of graphic cards and multi-core processors. I had access to all of these accelerators thanks to Cyfronet, so I can certainly claim that without Cyfronet, it would not have been possible to obtain such impressive results.



Graphics card threads clustering scheme

Would you have any advice for colleagues considering taking up doctoral studies?

At the beginning of my science adventure I joined the "Calculation acceleration" team in Cyfronet, where I met outstanding specialists who turned out to be wonderful people. Thanks to this, a passion and a deep desire awakened in me to learn more about the subject that I took up in my doctoral dissertation and which I have been dealing with up until today. Therefore, I strongly recommend young scientists to establish cooperation with a research team which is dealing with interesting subjects. In such a team, they will contribute to complex research projects and they will thus be able to gain valuable experience and enrich their knowledge faster.

Awards of the National Science Centre for users of ACC Cyfronet AGH computing resources

NATIONAL SCIENCE CENTRE POLAND

The award of the National Science Centre (NSC) is a distinction for young scientists for their significant scientific achievements, performed as part of basic research carried out in a Polish scientific unit, documented by publications affiliated to such a unit. The award was established by the Council of the National Science Centre in February 2013. The award is granted in three research areas: 1) humanities, social and art sciences, 2) life sciences, 3) technical sciences. Winners of these awards often emphasize the role of Cyfronet computing resources in achieving research results.

Award of the National Science Centre for Sebastian Glatt, PhD DSc from the Małopolska Centre of Biotechnology of the Jagiellonian University in Kraków*

On October 6, 2021, Sebastian Glatt won the Prize in the field of life sciences. The prize was awarded for the following achievements: establishing the spatial structure and mechanism of action of the elongator protein complex; describing the mechanism of tRNA modification by the elongator complex; structural and functional characteristics of eukaryotic tRNA modifying complexes; determining the structure and function of the components of the elongator complex and linking their mutations with the formation of neurodevelopmental and neurodegenerative disorders.

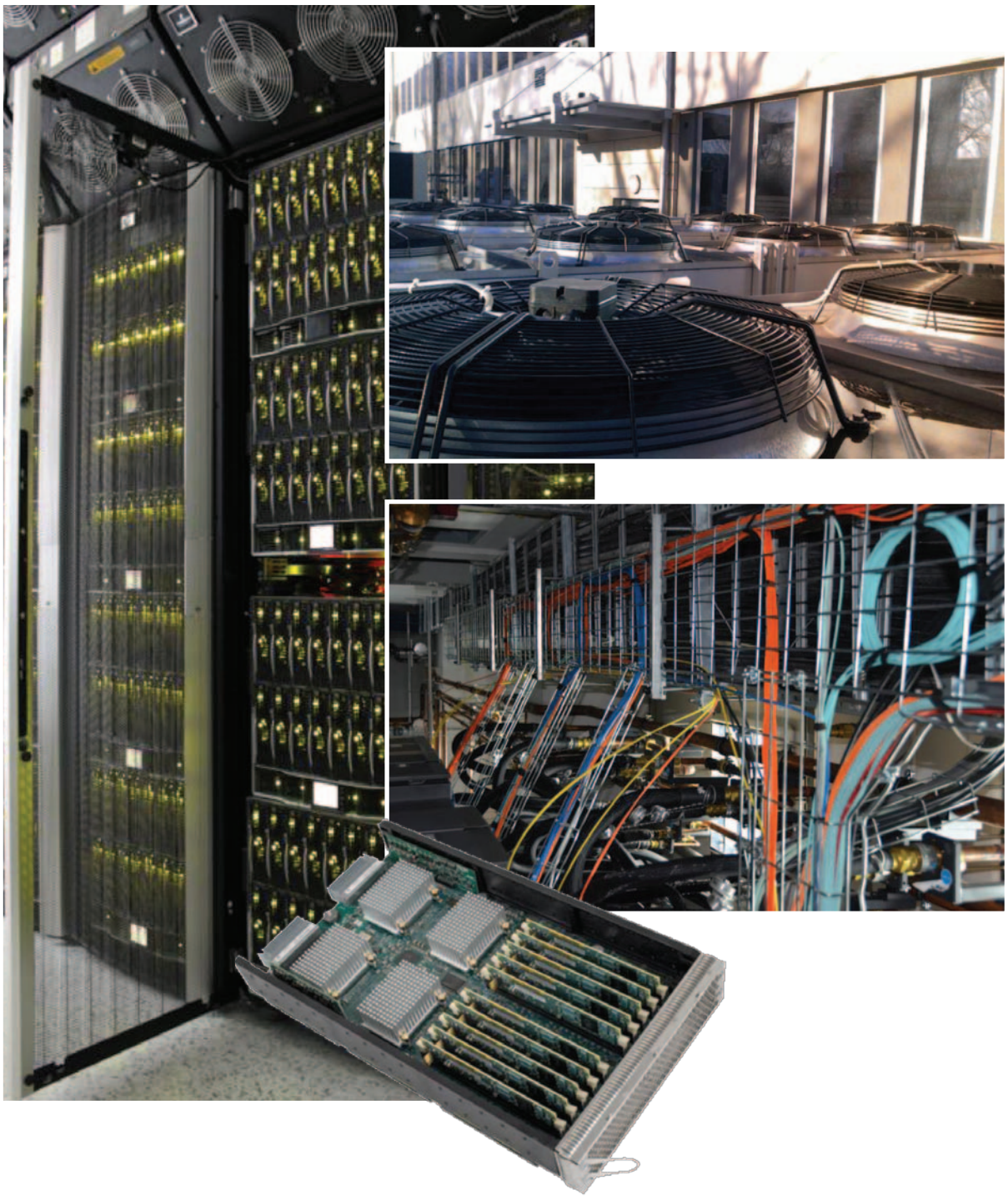
Sebastian Glatt graduated and received his PhD from the University of Vienna. He conducted research at the European Molecular Biology Laboratory in Heidelberg. He has been working at the MCB of the Jagiellonian University since 2015, he also manages the Cryo-EM laboratory team at the National Center for Electron Cryomicroscopy. He initiated very fruitful cooperation with scientists from other centers in Poland, Germany, France, Belgium, Switzerland and Australia.

- In my research group, we specialize in studying the three-dimensional shape of proteins and nucleic acids using cryoelectron microscopy and crystallography. Awards, such as the National Science Center Award, are received individually, but the whole team works for them – says Dr. Sebastian Glatt.

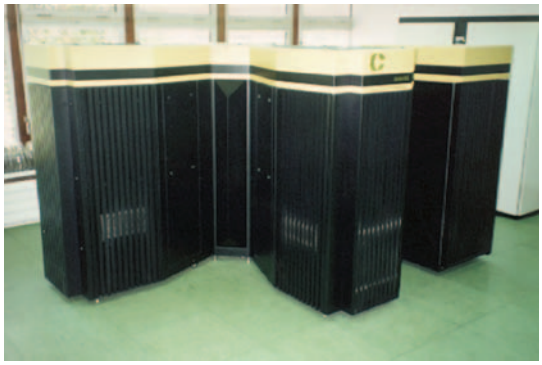
Sebastian Glatt supported his research with calculations on the Prometheus supercomputer at Cyfronet, within the *plgribosomes2021* computational grant: "Eukaryotic Ribosomes".

Congratulations to the winner and we wish you further success!

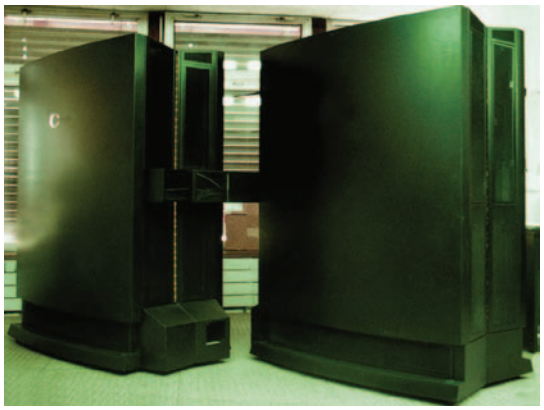
**Press material: https://www.ncn.gov.pl/*



CDC CYBER 72



Convex C3840



Exemplar SPP1600/XA

1973 CYFRONET is established

1975 A CDC CYBER 72 computer is deployed at the Centre

1990 The first KraKow node of the EARN / BITNET network is deployed at CYFRONET (on an IBM 4381 computer)

1991 CYFRONET installs a Convex 120 machine – the first vector computer in Central and Eastern Europe.

The first Polish national Internet link is established between Krakow and Warsaw.

Construction begins in the Krakow MAN

1994 A 2 Mbps link is deployed between Krakow and Warsaw

1996 An Exemplar SPP1600/XA computer deployed at CYFRONET took a position on the TOP500 list.

The first automatic tape library (ATL 2640) is installed at the Centre

1997 The ATM communications subnet is deployed within the Krakow MAN.

CYFRONET joins the POL-34 national backbone

1998 An SGI Origin2000 computer is deployed at the Centre

2000 Increasing the Centre network connection bandwidth to 155 Mbps

2002 A RackSaver PC computer is deployed at CYFRONET as part of the CrossGrid project

2003 An HP Integrity SuperDome computer is deployed at CYFRONET (the first such computer in Poland)

2005 An HP Storage Works XP12000 disk array is deployed at CYFRONET.

Increasing the Centre network connection bandwidth to 622 Mbps

2006 An HP Storage Works EVA 8000 disk array and an SGI ALTIX 3700 supercomputer (Baribal), with 0.8 TFlops of theoretical peak performance, is deployed at CYFRONET

2007 An agreement concerning the creation of the Polish Grid (PLGrid) Consortium was signed.

An SGI ALTIX 4700 supercomputer with the SGI RASC acceleration module is deployed at CYFRONET.

IBM BladeCenter HS21 servers are deployed at CYFRONET (6.2 TFlops).

An HP Storage Works EVA 8100 disk array is deployed at CYFRONET

2008 The configuration of SGI ALTIX 3700 supercomputer is extended to 1.5 TFlops.

MAN 10 Gbps started.

The Metropolitan Area Network is directly connected to Warsaw and Bielsko-Biała through the PIONIER network links, each of 2x10 Gbps capacity.

Zeus supercomputer (HP Cluster Platform 3000 BL) with 2 048 cores is deployed at CYFRONET

2009 Start of the PL-Grid project – Polish Infrastructure of Supporting Computational Science in the European Research Space

2010 The configuration of Zeus supercomputer is extended to 9,544 Intel Xeon cores, Zeus has been placed on 161st position on the TOP500 list

2011 Deployment of Hitachi Data Systems High Performance NAS Platform for computing infrastructure.

Total amount of installed disk space exceeds 2 PB.

The configuration of Zeus supercomputer is extended to 12,032 Intel Xeon cores.

Zeus has been placed on 80th position on the TOP500 list

2012 Start of the PLGrid Plus project – domain-oriented services and resources in the PL-Grid.

In April, ScaleMP, a leading provider of virtualization solutions for high-end computing, announced that Zeus-vSMP system at CYFRONET is the largest virtual SMP system in Europe.

Zeus among 100 fastest supercomputers on the TOP500 list.

The Metropolitan Area Network is directly connected to Rzeszow through the PIONIER network link of 2x10 Gbps capacity

SGI Origin2000

SGI ALTIX 3700

HP Cluster Platform 3000 BL

Anniversary Medal

New Machine Hall

Prometheus supercomputer

2013 After upgrading of Zeus supercomputer configuration to 25,468 cores, its theoretical peak performance reached 374 TFlops.

Anniversary Medal has been minted

2014 The new Machine Hall is completed.

Start of two new projects – PLGrid NG and PLGrid Core.

The Metropolitan Area Network is directly connected to Katowice through the PIONIER network link of 2x10 Gbps capacity

2015 The Prometheus supercomputer (41,472 cores) is deployed at CYFRONET, and ranks high, 49th place on the TOP500 list (the July edition), and next (after upgrading to 53,568 cores) 38th place (the November edition).

For the first time in history two supercomputers from Cyfronet (Prometheus and Zeus) are ranked on the TOP500 list, in one edition.

The new backup Data Center is completed.

CYFRONET starts active participation in INDIGO-DataCloud, EGI-Engage, EPOS-IP and PRACE-4IP projects.

High Performance Computing centres in Poland (Gdańsk, Kraków, Poznań, Warsaw and Wrocław) are integrated with links of 2x100 Gbps capacity

2016 Prometheus ranks 48th (the June edition) and 59th place (the November edition) on the TOP500 list

2017 Prometheus ranks 71st (the June edition) and 77th place (the November edition) on the TOP500 list.

Further dynamic development of the Centre, including establishment of 6 new laboratories.

Sat4Envi, Gliomed, EPOS-PL and eXtreme DataCloud projects launched

2018 Prometheus (53,604 cores, 2.4 PFlops) ranks 103rd place (the June edition) and 131st (the November edition) on the TOP500 list.

EOSC-Hub and PRIMAGE projects have been launched

2019 Cyfronet represents Poland in the LUMI consortium, composed of eight countries that will jointly build one of the fastest European supercomputers.

Prometheus ranks 174th place (the June edition) and 241st place (the November edition) on the TOP500 list.

Cyfronet exhibition stand at the ISC'19 conference.

Cyfronet provides a new computational system for research using AI methods, with computational power over 4 PFlops for tensor operations and 256 TFlops for standard calculations.

PRACE-LAB, PRACE-6IP, SANO, EOSC-Synergy and EOSC Enhance projects launched

2020 Among the strategic infrastructures included in January 2020 on the Polish Research Infrastructure Map there are two projects proposed by ACC Cyfronet AGH as the initiator and coordinator of the PLGrid consortium: *National Supercomputing Infrastructure for EuroHPC* and *National Cloud Infrastructure PLGrid for EOSC*.

Prometheus (53,748 cores, 2.7 PFlops) ranks 288th place (the June edition) and 324th (the November edition) on the TOP500 list.

The Prometheus supercomputer supports scientists in the fight against coronavirus.

EPOS PL +, PRACE-LAB2, EPOS SP, PROTEUS-RS and EUROCC projects launched

2021 Ares supercomputer with theoretical peak performance over 4.0 PFlops is deployed in Cyfronet.

Prometheus ranks 373th and 440th place, and Ares ranks 216th and 267th place on the TOP500 list.

A new version of the ACC Cyfronet AGH website has been launched.

EuroHPC PL, PIONIER-LAB, KMD3, AGH – PANDA3, EGI ACE, EOSC Future and FINDR projects launched

2022 Athena supercomputer with theoretical peak performance 7.7 PFlops is deployed in Cyfronet.

For the first time in the history of Polish computer science, three supercomputers from one Polish computing centre (Athena, Ares and Prometheus) are ranked on the TOP500 list.

Prometheus (53,748 cores, 2.7 PFlops) ranks 475th place, Ares (37,824 cores, 4.0 PFlops) ranks 290th place and Athena (6,144 cores, 7.7 PFlops) ranks 105th place on the TOP500 list.

Inauguration of the LUMI supercomputer.

EUMaster4HPC, FAIRCORE4EOSC, EuroScienceGateway, InterTwin, DT-GEO and Geo-INQUIRE projects launched.

Inauguration of the 50th anniversary of ACC Cyfronet AGH

AGH UST CAMPUS

LEGEND

1. Rector's Office
2. Faculty of Mining and Geoengineering
3. Faculty of Metals Engineering and Industrial Computer Science
4. Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering
5. Faculty of Computer Science, Electronics and Telecommunications
6. Faculty of Mechanical Engineering and Robotics
7. Faculty of Geology, Geophysics and Environmental Protection
8. Faculty of Mining Surveying and Environmental Engineering
9. Faculty of Materials Science and Ceramics
10. Faculty of Foundry Engineering
11. Faculty of Non-Ferrous Metals
12. Faculty of Drilling, Oil and Gas
13. Faculty of Management
14. Faculty of Energy and Fuels
15. Faculty of Physics and Applied Computer Science
16. Faculty of Applied Mathematics
17. Faculty of Humanities
18. AGH UST Academic Centre for Materials and Nanotechnology
19. AGH UST Centre of Energetics
20. Main Library
21. Walery Goetel School of Environmental Protection and Engineering
22. Department of Foreign Languages
23. Department of Sport and Physical Education
24. AGH UST Swimming Pool
25. Centre of e-Learning
26. **AGH UST Academic Computer Centre CYFRONET AGH**
27. University Computer Centre
28. Department of Education
29. Centre for International Students
30. Education Centre
31. University Admissions Board for Prospective Students
32. AGH UST Student Campus
33. University Board of Student Government
34. Career Centre
35. Centre for Transfer of Technologies
36. Administration and Business Cooperation Department
37. Krakow Centre for Innovative Technology INNOAGH
38. Centre for Project Management
39. Department of International Collaboration
40. Disability Support Office
41. AGH UST Museum
42. Geological Museum of the Faculty of Geology, Geophysics and Environmental Protection
43. AGH UST Press
44. Academic Cultural Centre, Club STUDIO
45. Student Club Gwarek
46. Student Club Zaścianek
47. Student Club Karlik
48. Student Club Filutek

ISBN 978-83-61433-42-2

Academic Computer Centre
CYFRONET AGH
Nawojki 11

30-950 Krakow, P.O.Box 386
phone: +48 12 632 33 55
fax: +48 12 633 80 54

cyfronet@cyfronet.pl
www.cyfronet.pl