Academic Computer Centre

CYFRONET AGH

2020

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!

We present our next information folder. In the coming 2020, at ACC Cyfronet AGH, we will constantly provide our ICT services supporting scientists in research work. In June 2019, the Prometheus supercomputer with 2.4 PetaFlops of computing power, this time as the only supercomputer from Poland, once again took the high 174th place in the TOP500 list of the world's fastest supercomputers. In 2018, Prometheus and Zeus with computing power reaching a peak performance of 374 TeraFlops altogether performed over 8 million computing tasks for scientists representing many universities and research institutes.

Scientific research carried out with the help of the Centre's resources concern many fields, and the number of scientists using our supercomputers is constantly growing. Access to the computing power, IT tools and memory systems offered by Cyfronet, allows researchers to obtain scientific calculation results in a much shorter time when compared with a personal computer. We are glad to support researchers in their scientific work and in making important discoveries. Each success of Cyfronet infrastructure's Users is extremely important to us, that's why we present here selected doctoral dissertations describing interesting scientific issues, innovative applications of conducted research and National Science Centre awards for significant scientific achievements – it is worth noting that in all presented cases research was carried out with the help of Cyfronet supercomputers.

The important aspect of our activity is the construction and development of a nationwide, distributed, grid and cloud computing infrastructure under the PLGrid Consortium led by Cyfronet. Thanks to the resources available through the PLGrid platform, many Polish scientists conduct their research and perform analyses at a world-class level and can compete with specialists from foreign research centres.

ACC Cyfronet AGH carries out the mission of supporting science in many areas and works for both small research groups and individual researchers, as well as large, international projects. Cyfronet participates in a number of Polish and international scientific and research projects that provide us with extremely important experiences in the field of new ICT technologies. Cyfronet's activity particularly concerns many projects of the Horizon 2020 program, the Smart Growth Operational Programme, the Digital Poland Operational Programme and others, under which the Centre provides the HPC infrastructure along with extensive resources of memory systems and engages its competences.

I would like to thank all our Friends and Users for cooperation and valuable advices regarding the further development of the Centre. I would like to invite you to personal contacts, also during our annual meetings: HPC Users' Conference and Cyfronet Open Day.

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 12 632 33 55
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 Storage Department – Marek Magryś, MSc (manager)
Phone: +48 12 632 33 55, e-mail: Marek.Magrys@cyfronet.pl

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

High Performance Computing Users Department – Mariusz Sterzel, PhD (manager)
Phone: +48 12 632 33 55, e-mail: Mariusz.Sterzel@cyfronet.pl

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

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

Technical Department Pychowice – 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

PLGrid Program and European Collaboration Team – 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. (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 – Tomasz Szepieniec, MSc (manager)
Phone: +48 12 632 33 55, e-mail: Tomasz.Szepieniec@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 Visual Techniques – Jacek Przybylski, MSc (manager)
Phone: +48 12 632 33 55, e-mail: Jacek.Przybylski@cyfronet.pl

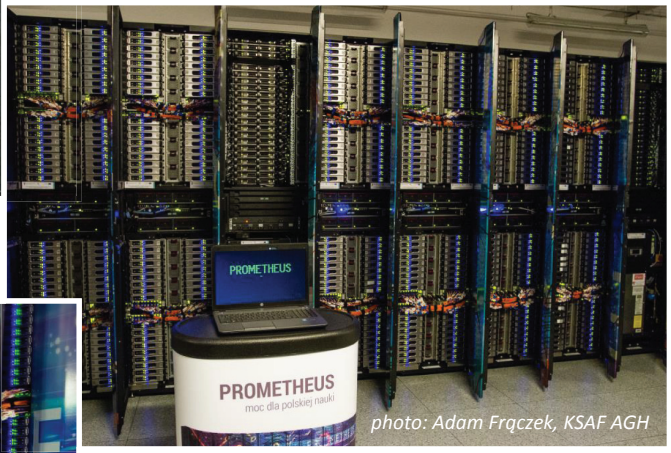


photo: Adam Frączek, KSAF AGH

photo: Krzysztof Mastalski, KSAF AGH

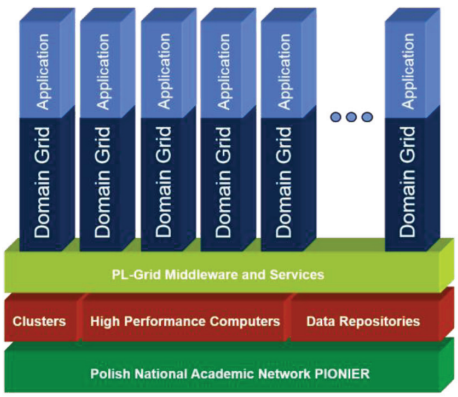
Marek Jarnicki, Prof. *The Chairman*	Jagiellonian University
Marek Jeżabek, Prof. *Deputy Chairman*	Institute of Nuclear Physics Polish Academy of Sciences
Bogdan Bochenek, MSc	Institute of Meteorology and Water Management
Andrzej Dietrich, MSc	Institute of Petroleum and Natural Gas
Jan Dudek, MSc Eng.	Cracow University of Economics
Florian Gambuś, Prof.	Agricultural University in Cracow
Jerzy R. Jaworowski, PhD Eng.	Cracow University of Technology
Paweł Kruszelnicki, MSc Eng.	Academy of Physical Education
Robert Gryboś, PhD	Jerzy Haber Institute of Catalysis and Surface Chemistry Polish Academy of Sciences
Wojciech Majka, MSc Eng.	Jagiellonian University Medical College
Paweł Mikłaszewski, MSc	Academy of Music
Roman Mazur SDB, PhD	Pontifical University of John Paul II in Cracow
Andrzej R. Pach, Prof.	AGH University of Science and Technology
Wojciech Palacz, PhD	Jagiellonian University
Bartłomiej Pokrzywka, PhD	Pedagogical University in Cracow
Marek Policht, MSc Eng.	Institute for Advanced Manufacturing Technologies
Tomasz Szmuc, Prof.	AGH University of Science and Technology
Jacek Barski, Deputy Director	Marshal Office of the Małopolska District
Grzegorz Żych, Director	Cracow City Hall

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 consisting Interdisciplinary Centre for Mathematical and Computational Modelling in Warsaw, Poznan Supercomputing and Networking Center, Wroclaw Centre for Networking and Supercomputing, Academic Computer Centre in Gdansk and **Academic Computer Centre Cyfronet AGH as the initiator and coordinator 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 81st 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.

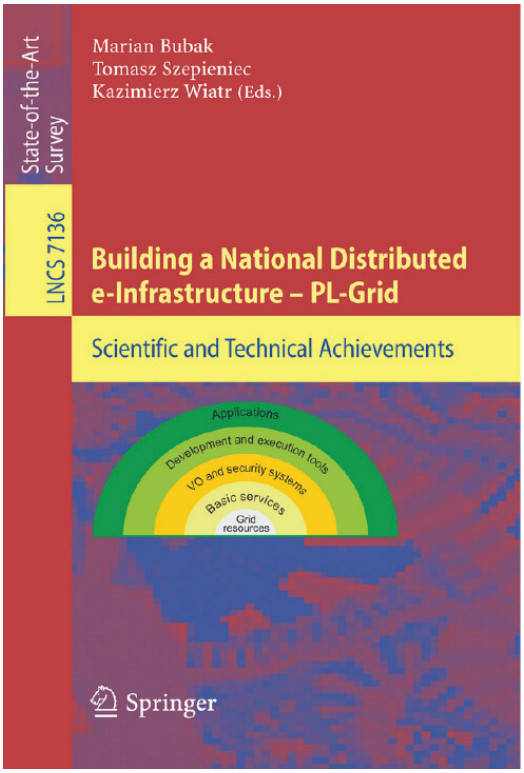
In the PLGrid NG project, the domain-specific grids were developed for several other groups of scientists, representing fourteen research fields (**in total, in the two projects, IT support tools were built for 27 scientific disciplines**).

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!**

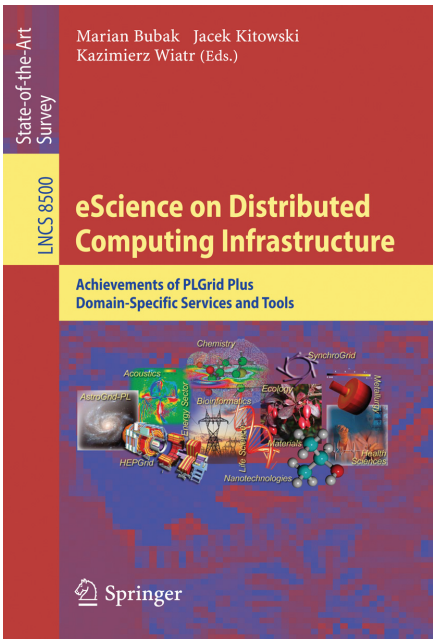
At the beginning of 2019, more than 5 PFlops of computing power and more than 60 PB of disk storage were available within the infrastructure. In addition, many tools supporting organization of computational experiments, designing and running applications, computationally supporting research and results' 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.

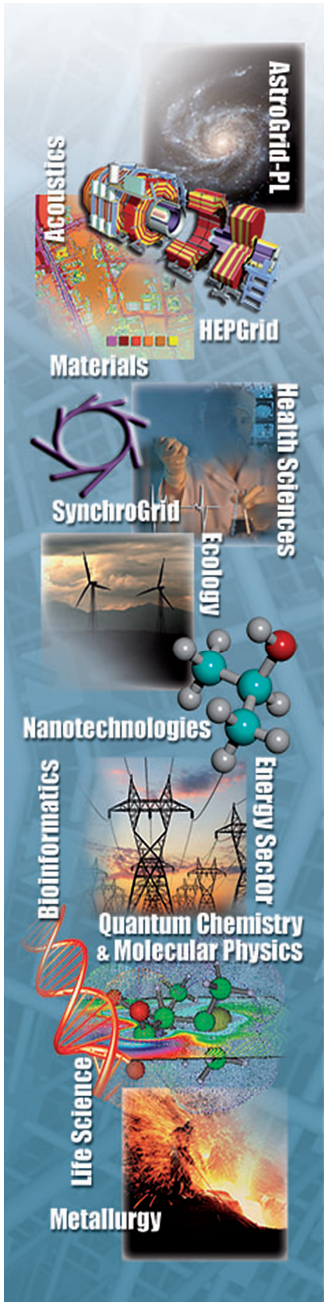
The second part (chapters 9 to 13) presents new environments and IT services that can be used by all of the previously mentioned groups of scientists. The third part (chapters 14 to 36) describes how specific environments, tools and services, prepared within the PLGrid Plus project, are used in advanced 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.

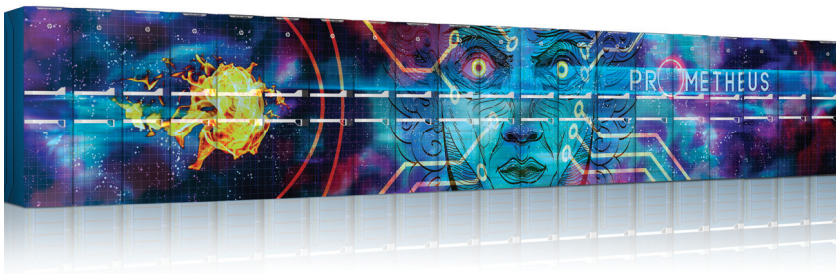
Preparation and implementation of a set of domain-specific services fit very well with the need of development of an advanced IT infrastructure designed for the implementation of modern scientific research. The well-tailored PLGrid e-infrastructure does not only fulfil researchers' needs for suitable 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.



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 2,200 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 53,604 cores. These are accompanied by 282 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 GPGPUs. The theoretical performance of Prometheus is 2.4 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. To achieve the same computing power in case of Zeus (Prometheus' predecessor), it would have to take about 160 racks. Baribal, the predecessor of Zeus with computing power of 1.5 TFlops, was placed on 8 racks. To achieve the computing power of Prometheus it would take as many as 12,000 Baribal's racks.

Prometheus in numbers	
Number of computing cores	53 604
RAM	282 TB
Number of GPGPUs	144
Computing power	2.4 PFlops
TOP500 – the list of the world's fastest computers (June 2019 edition)	174th 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.

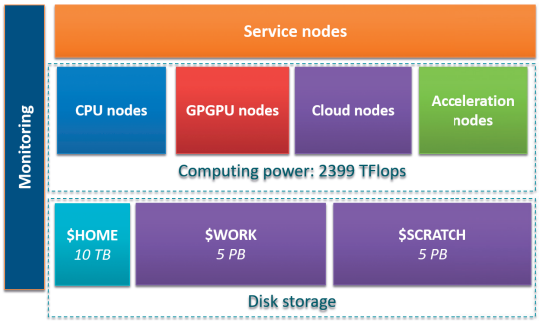
Prometheus once again has been listed on the TOP500 list of the world's fastest computers (June 2019 edition) and took the **174th position**, this time as the only supercomputer from Poland.

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,
- set of servers designed for Cloud computing,
- 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).

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

Prometheus architecture



Year	No. of Jobs	CPU time in years
2015	1 099 822	5 811
2016	3 080 543	21 239
2017	5 032 438	36 600
2018	5 430 811	39 946

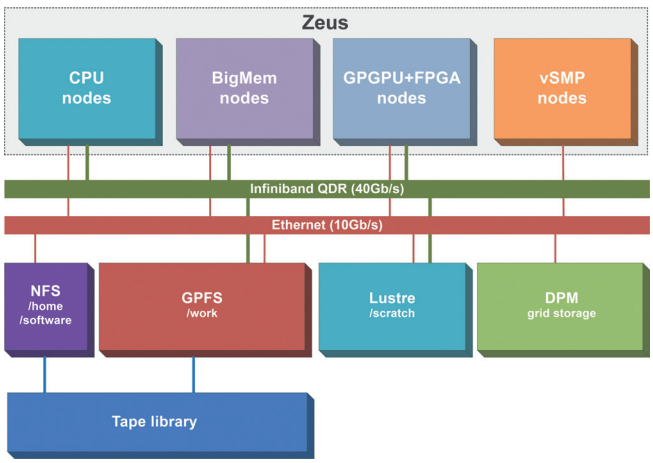
ZEUS – over 57 000 CPU-years

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 **81 – the highest noted spot**. Zeus was 10 times the fastest in Poland.

The architecture

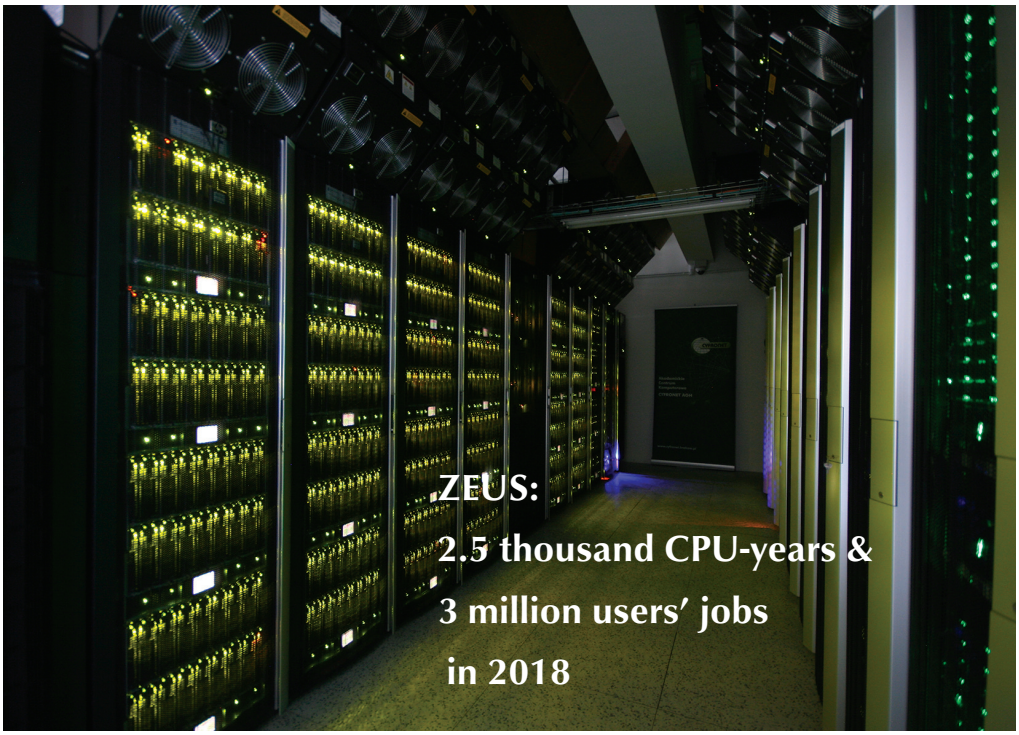
Zeus is a heterogeneous computing cluster. It constitutes of four classes of nodes, varying in terms of architecture of computing resources, specifically tailored to the requirements of the scientific communities. The Zeus architecture is a composite of four partitions:



- **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 2018 alone Zeus performed almost **3 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**!

Year	No. of Jobs	CPU time in years
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
2015	6 405 941	10 141
2016	4 668 134	3 414
2017	4 034 454	2 632
2018	2 911 875	2 490

Supercomputers usage

Prometheus and Zeus are part of the European cloud and grid infrastructure under the European Grid Infrastructure (EGI). At the same time, Prometheus and Zeus 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 Prometheus and Zeus 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 Zeus and, recently, Prometheus is quite reach. It includes:

- modeling of the electron structure of composites containing graphene oxide,
- research on the physicochemical properties of biodegradable polymers,
- phase transitions in laser welded steels,
- fatigue tests of the drill string and drilling fluids' flows,
- the use of gravitational lenses as space telescopes,
- modeling the activity of selected anticancer drugs through carbon nanoparticles,
- research on the activity of storm clouds on a local and global level,
- using quantum methods for image analysis.

A wide range of research topics is evidence of constantly increasing number of scientists, who are aware of advantages of supercomputers like Zeus or Prometheus. With the help of these powerful supercomputers 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 projects, including experiments like CTA, LOFAR, EPOS, Large Hadron Collider in CERN and the recently discovered gravitational waves in LIGO and VIRGO detectors.

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

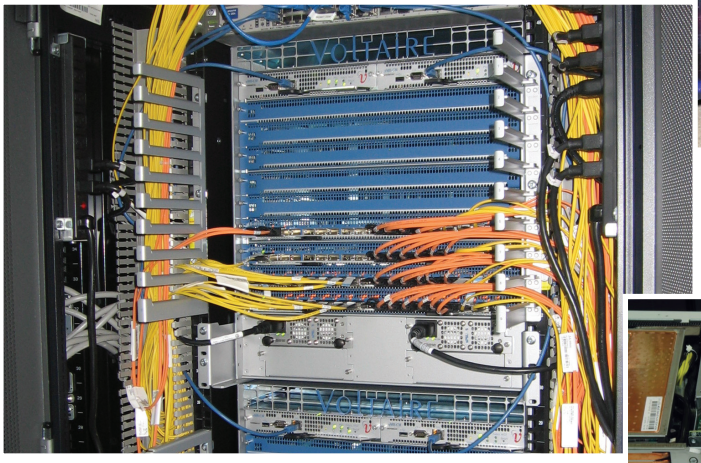
the Polish scientific community. The usefulness of supercomputers provided by ACC Cyfronet AGH as a tool for conducting research is best evidenced by statistical data on their use.

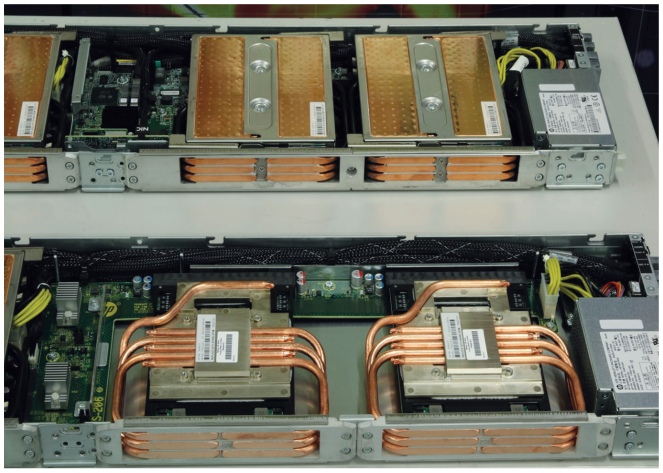
The table presents the aggregated key data on the number of computational tasks and their duration, performed by Cyfronet for other units.

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.

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
Prometheus supercomputer		
2015	1 099 822	5 811
2016	3 080 543	21 239
2017	5 032 438	36 600
2018	5 430 811	39 946
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


PROMETHEUS


VOLTAIRE
VOLTAIRE



Data storage

Only the proper teaming of computing infrastructure with the right selection of storage solutions can assure the best quality of services provided to users. The scale of problems in this area increases with the complexity and the efficiency of high performance computers. At present, disk storage systems attached to Cyfronet's supercomputers store over 500 000 000 data files (with the file sizes up to several terabytes). A wide variety of research conducted on the Centre's resources requires not only diverse configuration of key Cyfronet's computers, but also an efficient, dedicated storage system.

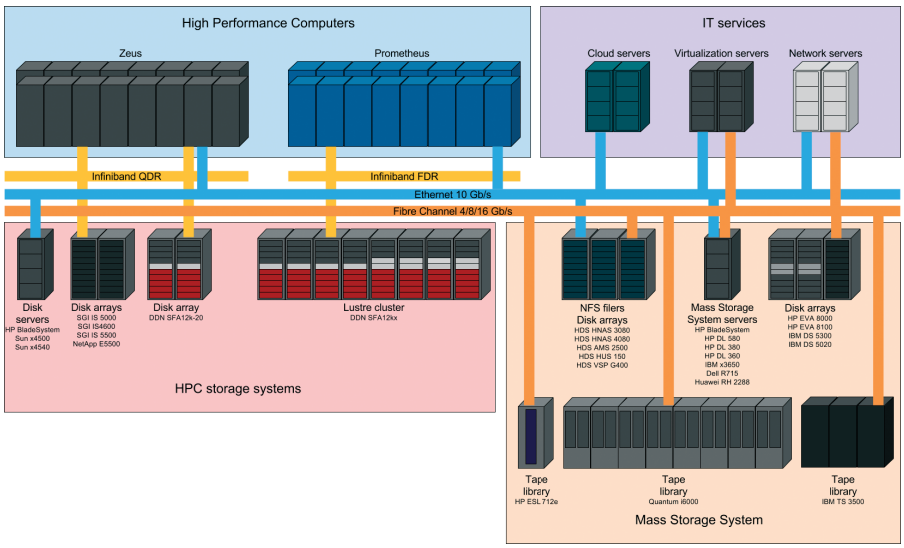
The most fundamental is the one used for keeping users' home directories. In this case all the crucial elements provide a very high level of availability and data security, which are supported by mechanisms such as snapshots and backups to external tape libraries. Zeus and Prometheus (the two main supercomputers of the Centre) offer such functionality through using specialized HNAS file servers (so called filers), produced by Hitachi Data Systems. These servers support hardware implementations of the NFS protocol and provide very high performance and high availability of the file systems. HNAS filers are coupled with Hitachi Data Systems AMS 2500 and HUS 150 disk arrays, used as repositories of physical disk space. These devices also provide extremely high levels of security and performance, fitted to the specific characteristics of the data stored in home directories.

Another type of storage space used in supercomputers is the scratch space, in which the crucial factor is speed. To address this requirement, Cyfronet uses the Lustre distributed file system, which is capable to scale both space and performance by aggregating storage capacity of many servers. Moreover, throughput and/or capacity can be easily increased by adding more servers dynamically, without interrupting user computations. Nowadays, all Cyfronet's supercomputers can use scratch spaces based on Lustre. In Zeus case, it is the file system with almost 600 TB capacity and 12 GB/s read/write bandwidth. Prometheus' scratch has enormous capacity of 5 PB and 120 GB/s read/write bandwidth. For even more demanding disk access requirements it is possible to use a super-fast RAM-disk provided by the vSMP partition of the Zeus supercomputer.

However, the major part of Cyfronet's storage resources is dedicated to the needs of users of domain-specific services developed in the PLGrid program. The PLGrid infrastructure provides a dedicated workspace for groups in domain grid environments – the functionality essential for enabling cooperation of scientists from geographically distributed locations. Zeus provides almost 200 TB of such disk space with the use of HNAS filers and the NFS protocol. Prometheus offers similar functionality with higher performance, using the Lustre file system. The maximum capacity of the */archive* resource in this supercomputer reaches 5 PB and the total rate of read/write operations attains 60 GB/s.

A special case of mass storage are the resources for large projects and international collaborations, in which Cyfronet takes part, such as WLCG (Worldwide LHC Computing Grid), which stores and analyzes the data coming out of the LHC detector in CERN, or CTA (Cherenkov Telescope Array). Such projects demand high volumes of disk space available by a set of specialized protocols, such as SRM, xroot or GridFTP. Cyfronet provides such space with the use of the DPM (Disk Pool Manager) instances and dedicated networks, such as LHCone. Total amount of disk space provided by these services exceeds 1 PB. The overall data storage space exceeds **49 PB**.

Comprehensive infrastructure of efficient and safe storage of digital data

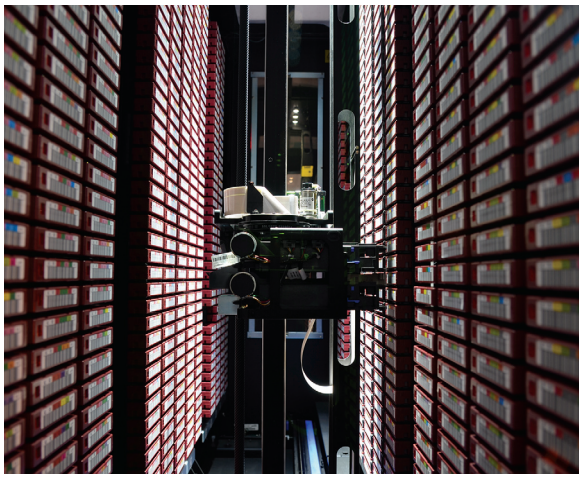


The currently observed phenomenon of the rapidly growing amount of digital information also applies to the scientific community. Access to very efficient supercomputers enables performing analyses of large-scale research problems, what results in generation of huge data sets. They require a completely 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 resources is **24 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 three tape libraries having in total 6 thousand slots for LTO magnetic tape drives and 36 drives of the III, IV, V and VI generation. A single LTO-6 magnetic medium has a physical capacity of 2.5 TB and allows recording at the speed up to 160 MB/s, which theoretically allows the storage of almost 15 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.

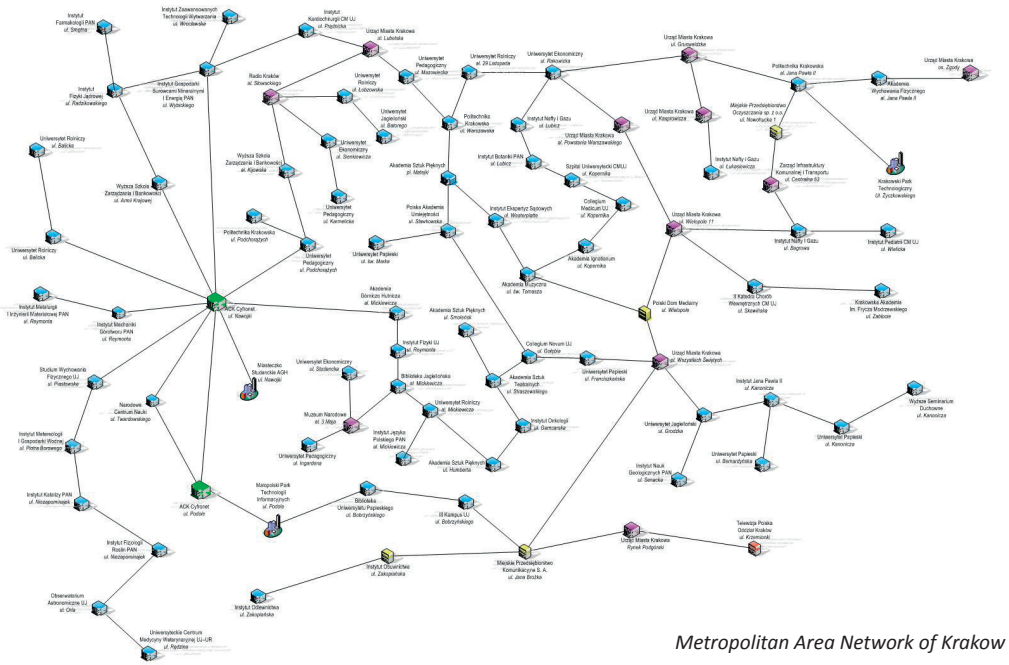
Backup is performed on the active data – that might be currently in use – through a replication process from the source location to a separate, isolated destination. The ideal backup procedure ensures consistency of the source and backup data, both at the level of a single object (a file located on a 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**.

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.



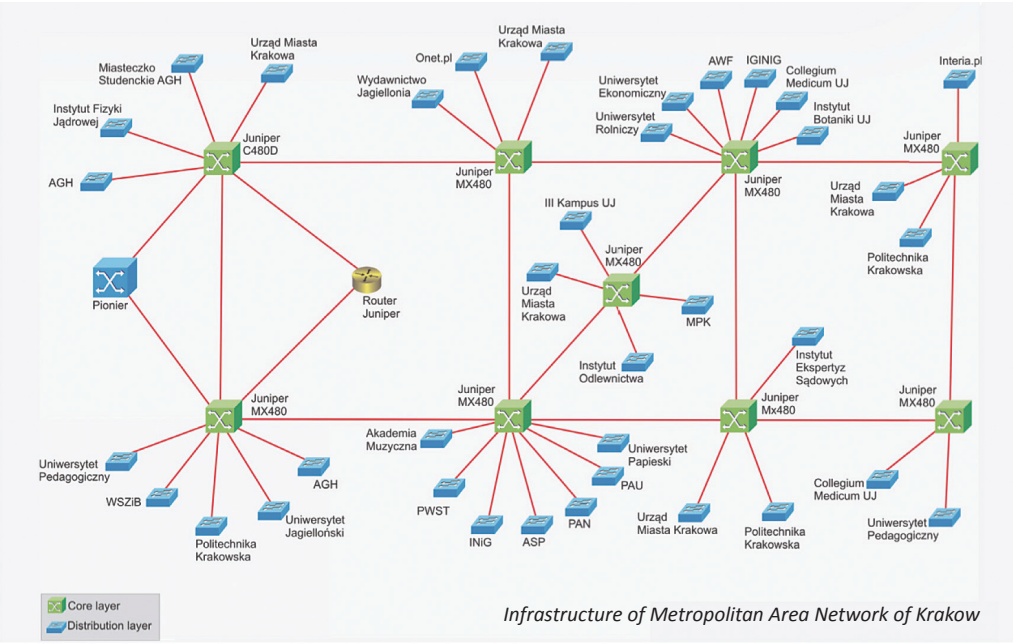
Metropolitan Area Network of Krakow

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 almost 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 3 Gbps capacity is established to the International Carrier TeliaSonera Poland network.

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. The news server operated by ACC Cyfronet AGH registers over 20 000 new messages each day,
- **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 2018	
Number of e-mails	> 18 000 000
Number of e-mail server sesions	> 45 000 000
Number of news groups	> 5 800
Daily number of news messages	> 20 000
Annual number of news messages	> 7 000 000

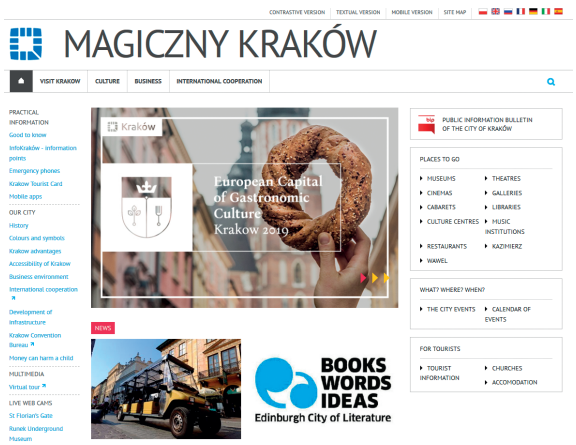
Portals and mobile applications

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.

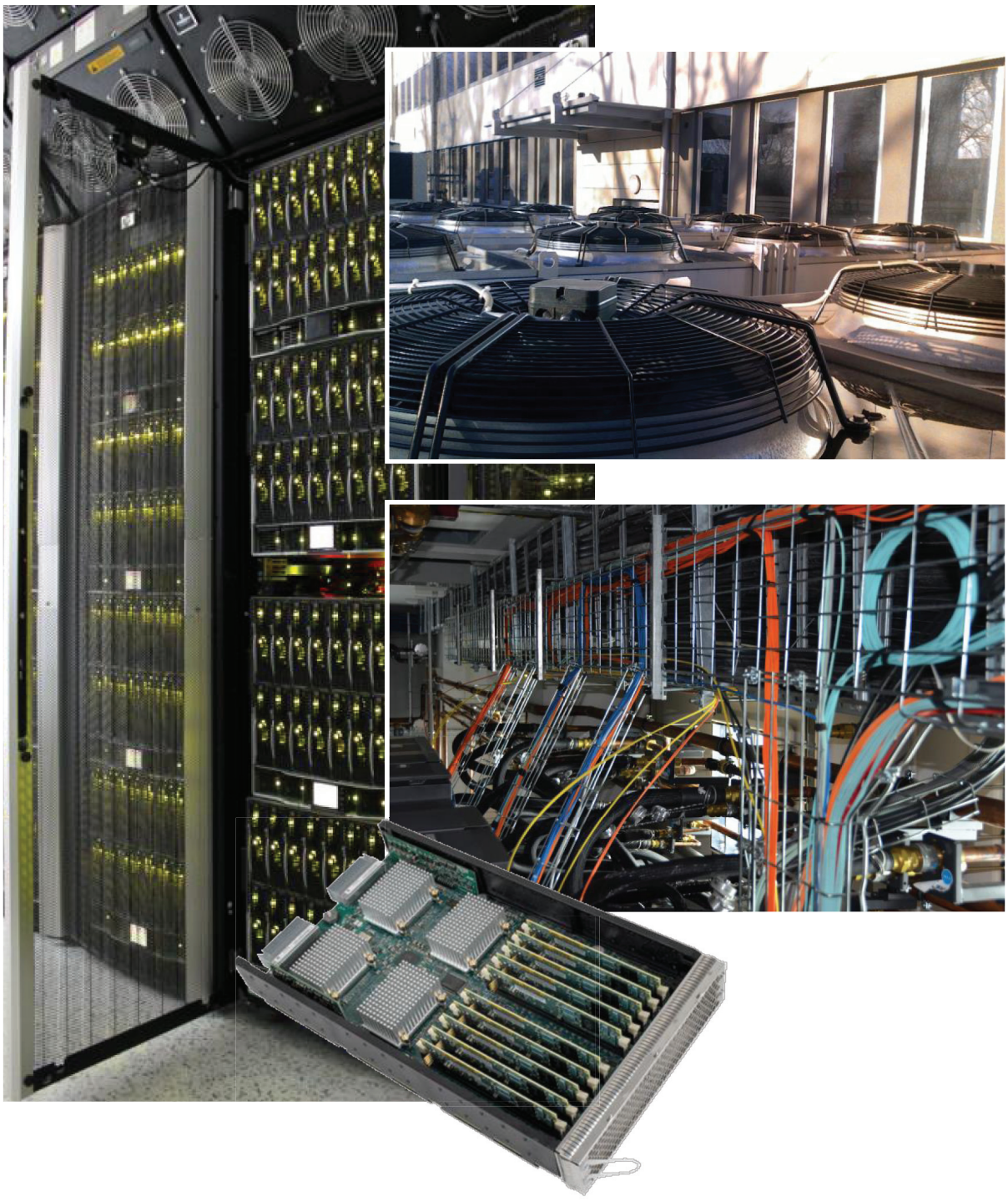
Cooperation with Kraków authorities is of particular importance for the Centre. The agreement between the Municipality of Kraków and CYFRONET, regarding the promotion of the City has resulted in the creation of an up-to-date portal. Aside scientific information the portal introduces its readers to the 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 Polish, English and Spanish.



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.

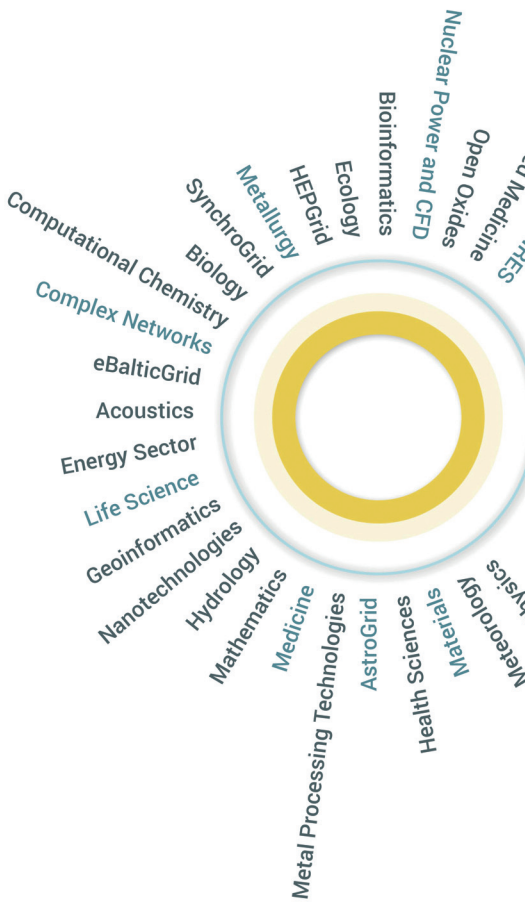
– *The Cloud Computing in PLGrid and Cyfronet is an innovative solution on a European scale. We have a strong belief it will bring a new quality level for research conducted by our users* – concludes Prof. Wiatr.

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

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.

Computational resources

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

Computing resources

Prometheus and Zeus supercomputers provide:
2,75+ PetaFLOPS
78 000+ cores
300+ GPGPUs
340+ TB RAM

Storage

24 PB of disc 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.

Tools for scientific collaboration

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

Computational cloud

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

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 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

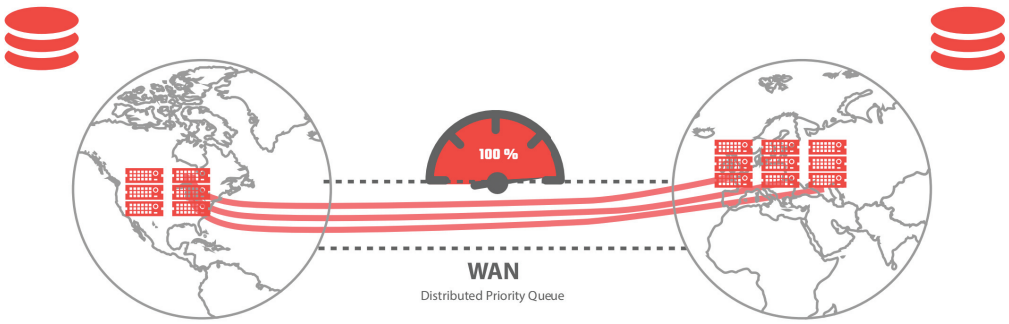
- Simple deployment based on Docker containers using a friendly command line client.
- Easy storage support for user requests based on secure tokens.
- Complex monitoring information available on all aspects of the system, accessible through REST API or directly visualized in the administration panel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



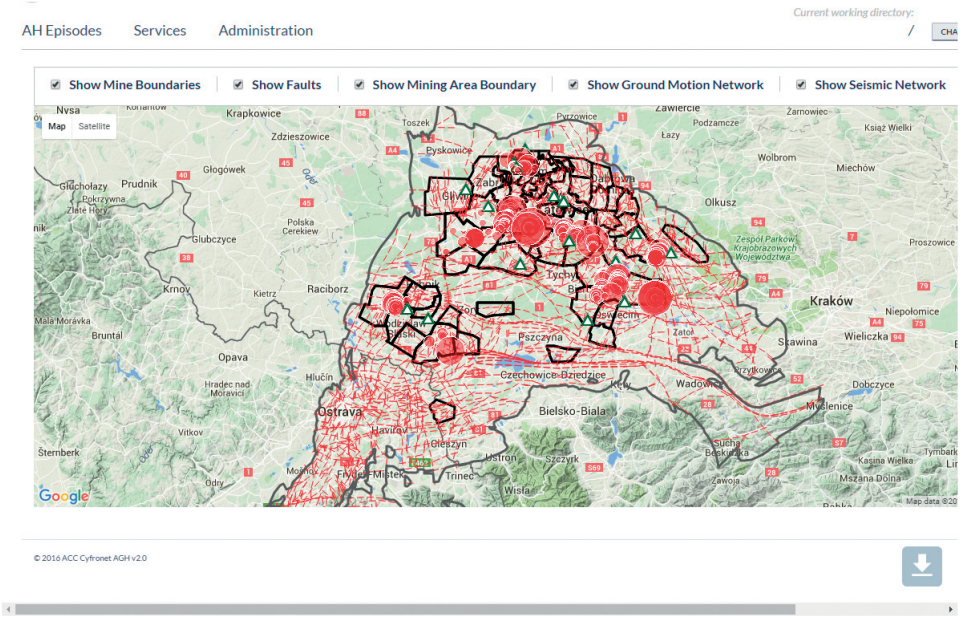
More information: *https://onedata.org*

InSilicoLab is a framework for building application portals, also called Science Gateways. The goal of the framework development is to create gateways that, on the one hand, expose the power of large distributed computing infrastructures to scientists, and, on the other, allow the users to conduct *in silico* experiments in a way that resembles their usual work.

The scientists using such an application portal can treat it as a workspace that organizes their data and allows for complex computations in a manner specific to their domain of science.

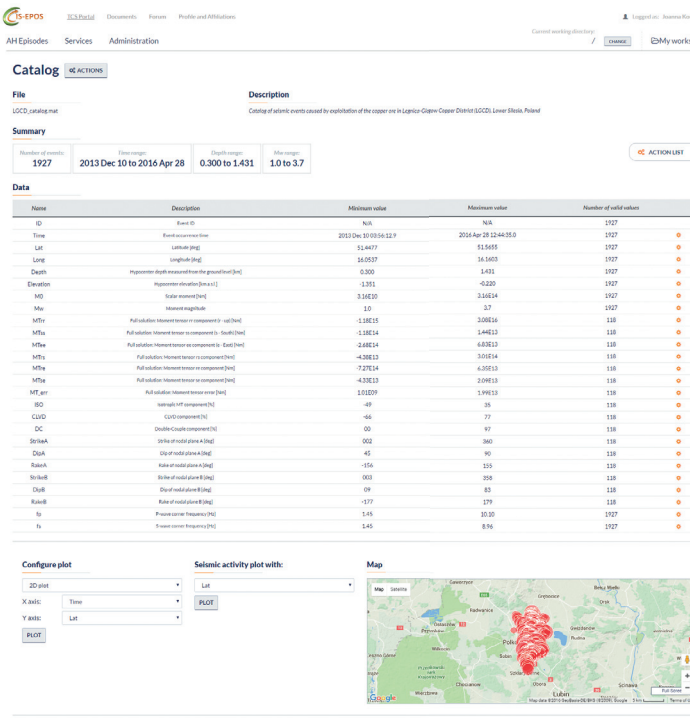
An InSilicoLab-based portal is designed as a workspace that gathers all that a researcher needs for his/her *in silico* experiments. This means:

- capability of organizing data that is a subject or a product of an experiment – this should include:
  - facilitating the process of preparation of input data for computations,
  - possibility of describing and categorizing the input and output data with meaningful metadata,
  - searching and browsing through all the data based on the metadata,
- seamless execution of large-scale, long-lasting data- and computation-intensive experiments.



The approach

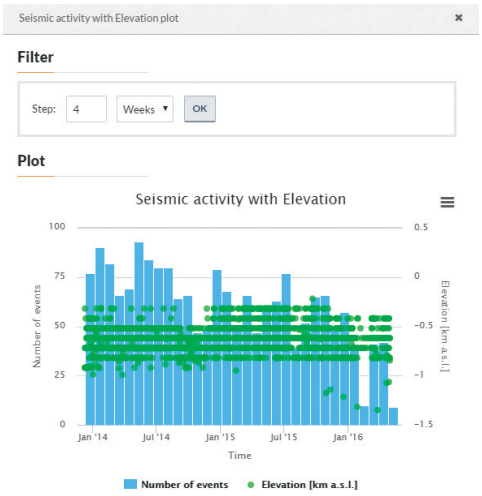
InSilicoLab is not meant to be a "*Yet another engine for...*", therefore, its developers has put maximum stress on the utility of the tool. This means user-friendliness, but, even more importantly, serving real scientific problems. This requires focusing on solving specific problems, rather than building a platform to solve any scientific problem, as the latter cannot be done in a universal and comprehensive way. Therefore, building a framework, which obviously is a generic solution, has to be performed in a bottom-up approach – starting from the particular problems, and building the generic tool from the common parts of the specific solutions.



Domain applications

Every gateway based on the InSilicoLab framework is tailored to a specific domain of science, or even to a class of problems in that domain. The core of the framework provides mechanisms for managing the users' data – categorizing it, describing with metadata and tracking its origin – as well as for running computations on distributed computing infrastructures. Every InSilicoLab gateway instance is built based on the core components, but is provided with data models, analysis scenarios and an interface specific to the actual domain it is created for.

http://insilicolab.cyfronet.pl

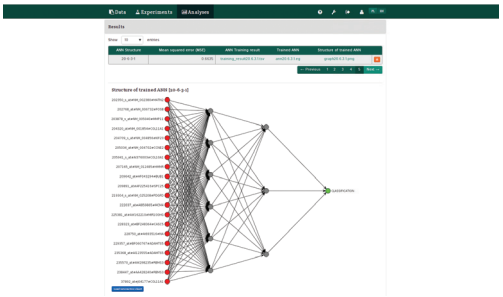


Life science and personalized medicine services

Out of disk space at your institute? Stuck waiting for alignment results? Can't easily share your results with collaborators? Or, perchance, are you looking for bioinformatics experts to assist you with managing large datasets? We can help!

We have developed a suite of tools which enables researchers representing various domains of science to perform complex large-scale experiments and manage their results. The efficiency of analysis tasks and security of scientific data are looked after by experts from the Academic Computing Centre Cyfronet AGH. Our tools reside in the PLGrid infrastructure which allows Polish scientists and their foreign collaborators to make use of vast computational services – free of charge and in a convenient manner.

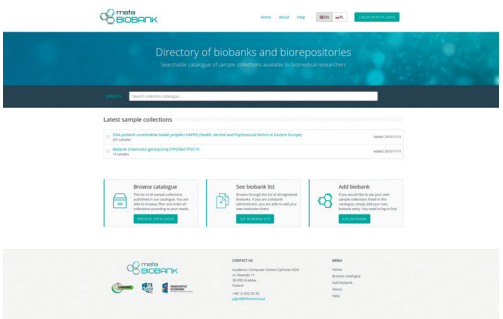
The *life science* toolkit includes:



DNA microarray integromics analysis platform

Designed for researchers who perform biological assays with the use of DNA microarrays. This service helps analyse gene expression data and correlate it with other clinical data sources characterising the target organisms.

https://integromics.plgrid.pl



MetaBiobank

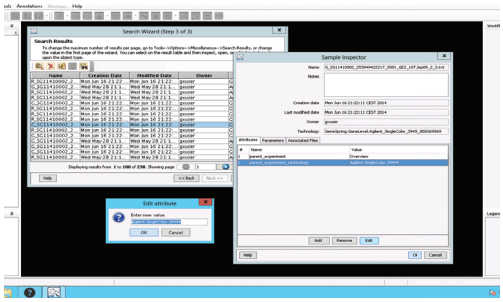
A convenient directory of biobanks and biorepositories, with searchable catalogue of sample collections available to biomedical researchers. It makes it possible to publish your own biological samples. Also supports searching for interesting samples in available biobanks and helps to establish new collaborations with other researchers. Supports ICD10 dictionary.

https://biobank.plgrid.pl

GeneSpring GX

An Agilent software package facilitating execution of statistical analyses and visualization of microarray data.

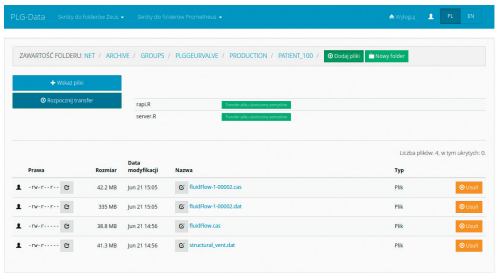
Accessible through a remote desktop interface.



PLGData

A tool for management of data stored in the PLGrid infrastructure. Comes with a user-friendly web interface. Currently supports both Prometheus and Zeus supercomputers at Cyfronet. Allows for easy but controlled computation output sharing with your team or research project consortium, using group folders. Existing API interface enables direct integration of your software with the service through a handy set of HTTPS operations.

https://data.plgrid.pl



Interested in collaboration?

We are on the lookout for collaborators interested in jointly developing new services for the life science domain. We also offer direct support in planning and carrying out scientific research. Our team takes part in collaborative, large-scale research projects as an international consortia member.

Write us at *plgrid@lifescience.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. Main advantages of the presented service include openess to using any technology and programming language and integration with the user authorization system of the PLGrid infrastructure.

Readiness for various applications

Applying REST principles in the implementation of the rimrock service allows to use its functionalities independently of any programming language chosen to build applications on top of the computing infrastructure. 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.

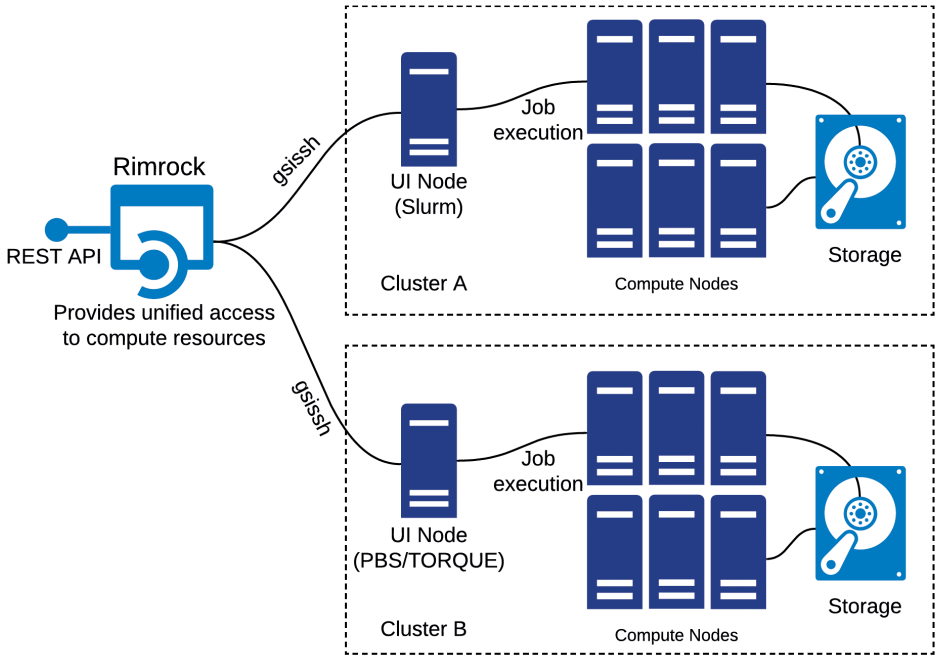
Another advantage of using REST interfaces is good support in any programming language by specialized libraries. It considerably facilitates the integration process with the computing infrastructure for both computation and data management. Each operation provided by the service is available through a single endpoint and the communication protocol uses a well-known JSON format, which also has good support in any programming platform.

The rimrock service has been successfully used during development of a web application in the domain of energy sector, allowing for harnessing the computing power of the PLGrid infrastructure for analysis of different scenarios of building a national power grid and their influence on the environment and human health. Through the integration with the rimrock service the time required for interfacing with the computing infrastructure has been minimized and the results can be obtained directly from the secure storage resources of the computing centres, without any additional data replicas.

Support for several job management systems

The rimrock service uses several job management systems, what ensures support for their unique features made available via the REST interface (e.g. using appropriate directives in existing computing scripts). It allows for easy integration of legacy applications in newly developed systems. With rimrock it is possible to submit jobs directly through multiple queuing systems such as the PBS system (*TORQUE*) or Slurm. The user may also submit interactive jobs by running instructions sequentially, what makes it possible to control the execution based on the returned data. 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. Through delegation of the mentioned certificate it is possible to constantly communicate with rimrock with just a single user login action. Furthermore, communication among several services within the PLGrid infrastructure is feasible, which allows to build complex systems on top of the existing services.

https://submit.plgrid.pl

Plgapp is a service designed for creators of web applications built on top of the PLGrid computing infrastructure. Developing applications with plgapp considerably shortens the implementation time by providing reusable elements such as handling of user login process, test and production environments for the developed application (setting up servers is not necessary any longer) and a set of libraries enabling management of computing tasks as well as access to computation results on behalf of end users.

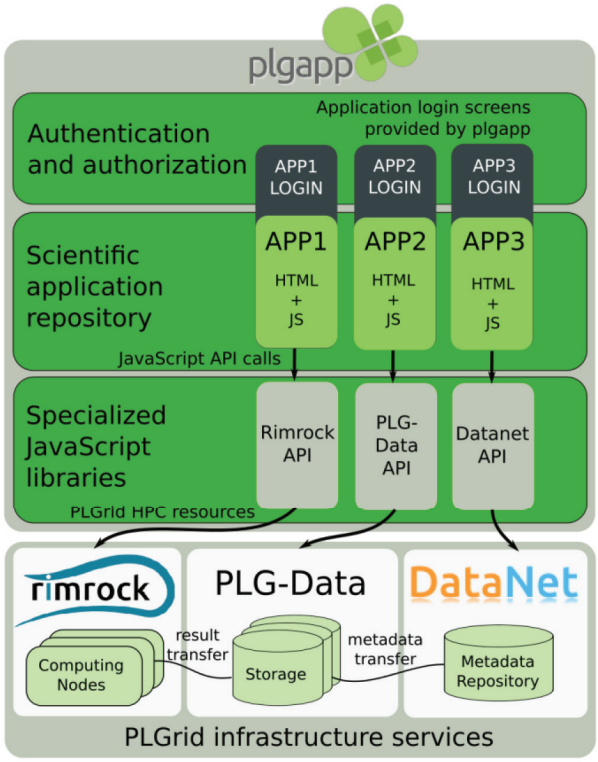
Developing web applications with plgapp

To start developing a web application with plgapp you need a text editor on your local computer and access to the service main page where you can register new web applications with a simple web form. The application right after registration is equipped with a login screen and its own address. Application code written with the text editor should consist of HTML, CSS or JavaScript files and can use any of the available *Web 2.0* programming components and libraries (e.g. JQuery, Angular, HighCharts, etc.). The libraries, among other things, allow to reuse graphical components such as interactive charts, animations or even 3D visualizations on the pages of the developed application. Submitting modified application code to the plgapp server for execution can be accomplished with the help of the already mentioned web form available on the service main page. If a given application is implemented by several people, it is possible to create a group within which the application code can be shared. A given application at any point in time can be published to the end users with just a single mouse click on the application panel on the plgapp web page. The key aspect, which is communication with the computing infrastructure, is addressed by providing a set of specialized JavaScript libraries that can be easily added and used in the created applications.

Specialized libraries available in plgapp

Integration with the computing infrastructure is handled with a set of specialized JavaScript libraries, which allow to retrieve basic information about the currently logged in end user, submit and monitor computing jobs, send and download data files as well as manage metadata. Each library uses one of the production services of the infrastructure and therefore ensures stable and coherent programming interfaces. The presented approach enables straightforward extension of the set of specialized libraries and makes the solution open for continuously expanded service offer of the PLGrid computing infrastructure. Undoubtedly another advantage is a single location of the documentation pages for each of the specialized libraries, which greatly facilitates their use. In order to improve the learning curve of using plgapp, the documentation contains code samples for each of the libraries and web page templates to be copied and pasted.

In the picture below a structure of the applications created with plgapp is given. The parts provided by the developers of the applications are depicted with blocks titled from APP1 to APP3. As can be seen, plgapp takes care of a considerable number of application elements, which would have to be provided by the developers themselves without using the presented service.

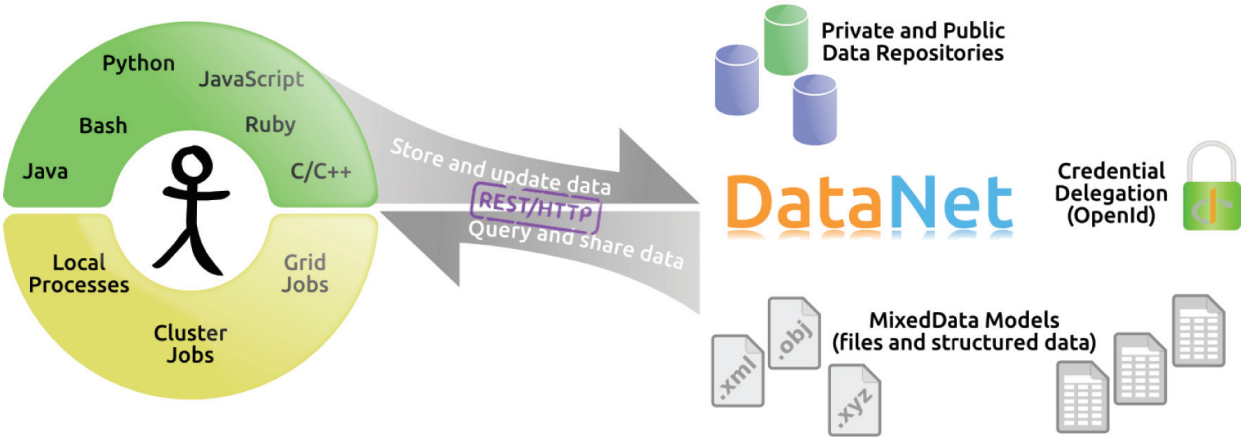


Plgapp service applications

Using the plgapp service is not limited to a particular science domain or a given scientific application. The service may be used wherever access to large computing resources is required via a modern graphical interface. As an example of using plgapp, building a portal managing simulations of reactors powered with atomic fuel can be mentioned (*https://mcb.app.plgrid.pl*). In this case, plgapp allowed to shorten the implementation time of the final solution and improved cooperation among developers and domain experts.

https://app.plgrid.pl

DataNet is a service built on top of the PLGrid high-performance computing infrastructure to enable lightweight metadata and data management. It allows creating data-models consisting of files and structured data to be deployed as actionable repositories within seconds. All this is accomplished with a convenient web interface where also access-restrictions to data stored in the repositories can be configured. The use of the service is not limited only to single users, but is open to others, like providers, who can access the repositories from their servers by delegating a user proxy certificate. It is even possible to gain access to the data directly from the web browser with the use of DataNet's CORS (Cross-Origin Resource Sharing) extension.



Interoperability

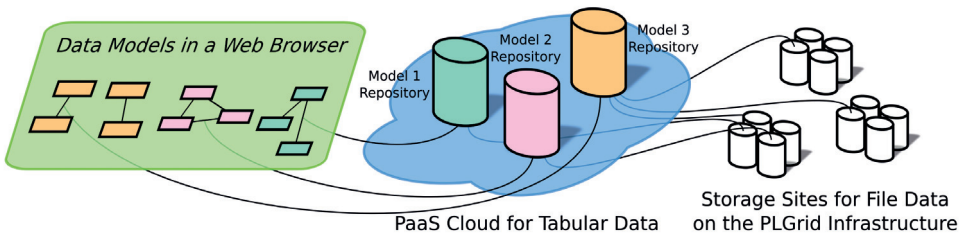
One of the main goals of DataNet is to make it usable from the largest set of languages and platforms possible. That is why we used the HTTP protocol as a basis for transferring data between computing nodes and the service, and – to make it even more convenient – we applied the REST methodology to structure the messages sent to and from the repositories. This way of transferring data has additional support in every major programming language, which makes the integration process straightforward. Our web interface has code templates for popular programming languages, which can be copied and pasted to get things going faster. DataNet is also independent of the computing location with the only requirement for it to be online, so you can use it from your local machine, a cluster job or a Grid job executed on any computing site.

Security

DataNet is fully integrated with the PLGrid authentication and authorization system, so existing users can quickly gain access to the service with a fully automated registration process. While using the service, all access rights to files and repositories are ensured and no additional privileges than those established within the PLGrid infrastructure are granted. Access to owned repositories can be configured on the level of individual users of the infrastructure. A temporary proxy certificate is used as the carrier of user credentials and it is possible to delegate it to other services for implementing complex, inter-service scenarios.

Resources

For storing metadata and data, DataNet uses resources available in the PLGrid infrastructure, thus the limits for storing capacities are specified only by individual user and group's quotas. In cases where big data volumes are processed and default user grants are not enough, a grant negotiation procedure should take place and resulting additional resources will be used by DataNet. In order to ensure user data separation, each repository is deployed on dedicated PaaS (Platform-as-a-Service), which ensures scaling and database service provisioning for structured data. For high-throughput scenarios, it is possible to configure the system to expose several instances of a given repository to increase request processing rate.



https://datanet.plgrid.pl

Scalarm is a platform dedicated for parameter studies and data farming experiments, which involve multiple executions of the same application with different input parameter values to explore its behavior in different conditions.
The main goal of Scalarm is to support such experiments in their essential stages (Fig. 1):

1. input parameter space specification,
2. application execution management on heterogeneous infrastructures,
3. results collection, analysis and visualization.

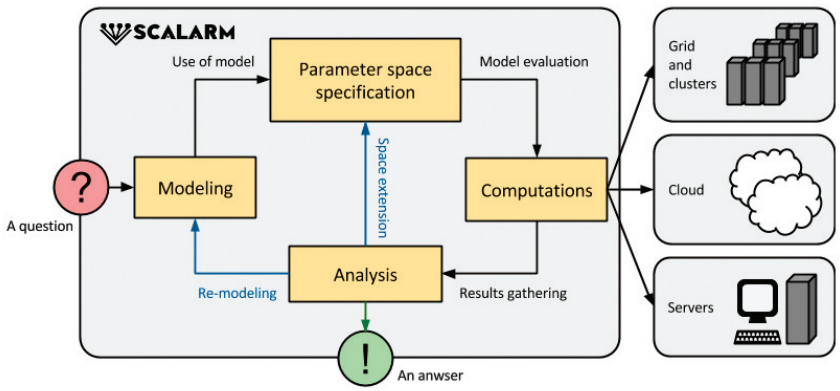


Fig 1. A diagram of experiments supported by the Scalarm platform

Input parameter space can be specified with: simple methods (single values, ranges, random values), design of experiment algorithms (Orthogonal Latin Hypercubes, 2^k), and imported CSV files. Each data point in the input parameter space constitutes a single application run in an experiment.

Scalarm can execute any scientific application including: sequential scripts, OpenMP and MPI parallel applications, and even scientific workflows by integrating with the Pegasus workflow management system. By using the adapter pattern, the user specifies how to provide input parameters, execute, monitor and collect output from the application. The results are automatically collected by Scalarm for analysis with provided graphical user interface.

Scalarm unifies access to and hides complexity of different computing infrastructures. The user decides where the application will be executed from an extensive list of supported e-infrastructures including:

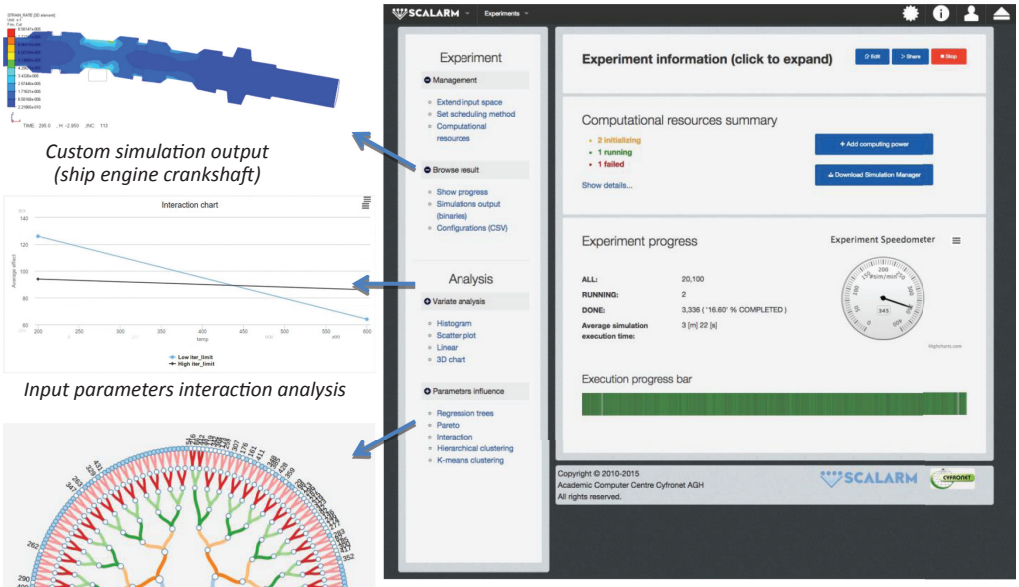


Fig 2. Scalarm web-based graphical user interface

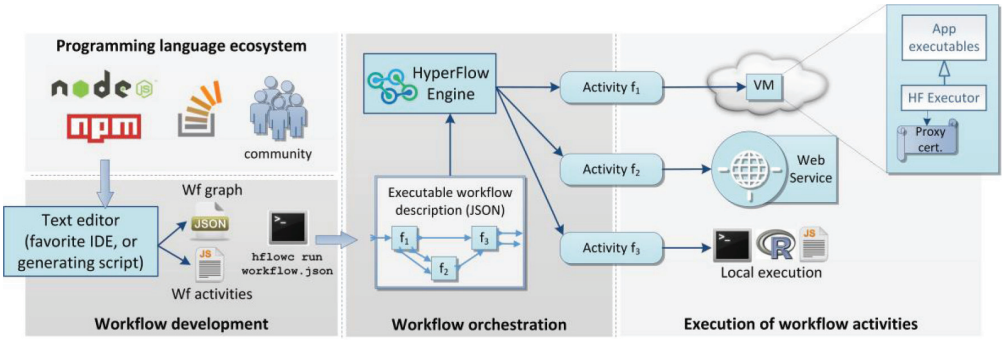
- the PLGrid infrastructure accessed via the QosCosGrid middleware,
- supercomputers with SLURM or PBS queuing systems such as Prometheus and Zeus at ACC Cyfronet AGH,
- public compute clouds – Amazon EC2, Google Compute Engine, or PaaSage environment,
- any server accessible via the SSH protocol.

Scalarm exposes a web-based graphical user interface (Fig. 2), and an HTTP-based application programming interface, which enable users to: register applications, create experiments, manage computational resources, monitor progress of the executing applications and experiments, analyze results with built-in methods. Any custom data produced by user's application can be downloaded from GUI.

Advanced features of the Scalarm platform include: a) automated input parameter space exploration with global optimization methods and sensitivity analysis techniques, b) computing resources scaling with different objective functions.

https://scalarm.plgrid.pl *http://www.scalarm.com*

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.

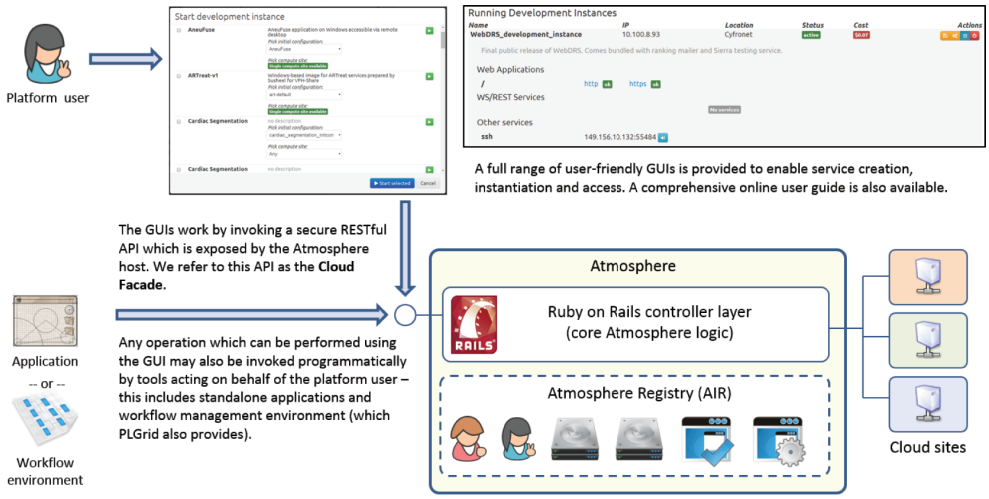
Contact

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*.

Atmosphere

The Atmosphere Cloud Platform is a one-stop site for management and interaction with the computational cloud resources operated by the PLGrid e-infrastructure. Atmosphere is a user-friendly environment where hybrid cloud resources contributed by various participating institutions and sites (including public clouds) are seamlessly integrated into a coherent, unified resource space, made available to PLGrid users. Accessing the Atmosphere platform enables you to exploit the cloud computational resources which are part of PLGrid.



Cloud service abstraction

The principal goal of Atmosphere is to make interaction with cloud sites easy for beginners and experienced users alike. Atmosphere can deploy virtual machines into the cloud, create snapshots and support sharing of computational services by PLGrid groups without the need to use any low-level cloud service libraries. A convenient GUI encapsulates all features offered by Atmosphere to each class of users: system administrators, application developers and end users. In addition, a set of APIs is provided to enable integration of the Atmosphere platform with external applications, tools and workflow management systems.

Security

The Atmosphere cloud platform is integrated with PLGrid authentication and authorization mechanisms. All users of the PLGrid infrastructure can request access to cloud resources simply by joining the **plgg-cloud2** user group. Thereafter, each user can create, use and share cloud-based virtual machines in the context of their own research team, or for individual research purposes. The visibility of each virtual machine (and the corresponding VM images) is restricted to the PLGrid team, in which the machine was created. Atmosphere performs automatic billing and resource consumption auditing for all VMs.

Resources

Atmosphere can interact with many different types of cloud resources contributed by individual cloud platforms. For the purpose of PLGrid, a dedicated cloud site has been set up at ACC Cyfronet AGH, comprising of computational nodes managed by the OpenStack cloud middleware. Atmosphere can also interact with public cloud providers, such as Amazon, RackSpace, Google Compute and many others. All this is done without forcing the user to learn any technology-specific libraries or APIs.

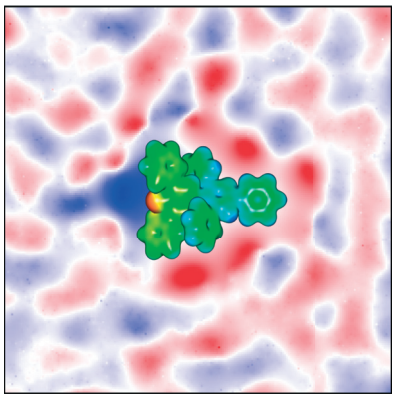
Applications

In addition to raw OS templates for service developers, the PLGrid cloud site supports a variety of ready-to-use applications. Any web or REST service can be deployed into the cloud using Atmosphere. It is worth noting that an earlier version of Atmosphere, developed in the framework of the VPH-Share project was successfully exploited in the Virtual Physiological Human community by approximately 25 research teams affiliated with the VPH-Share and VPH-DARE consortia, as well as external partners who collaborate with ACC Cyfronet AGH. Additionally, Atmosphere was used to provision computational services for the EurValve project in which ACC Cyfronet AGH participated as a member.

https://cloud.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 those being current challenges of nanotechnology or biological sciences) tractable and improve the accuracy of obtained results. Fortunately, constant progress in computer technology together with specialised software offered by Cyfronet meet this demand and enable various kinds of chemical computations.



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

Zeus and Prometheus clusters' nodes provide up to 1.5 TB of RAM and 64 cores per physical node, which enables quantum chemical computations that require large amount of memory or high number of cores with shared memory. Moreover, 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 to storage systems. The parallel-distributed Lustre scratch file system and possibility to use RAMDisk on selected nodes enable that.

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

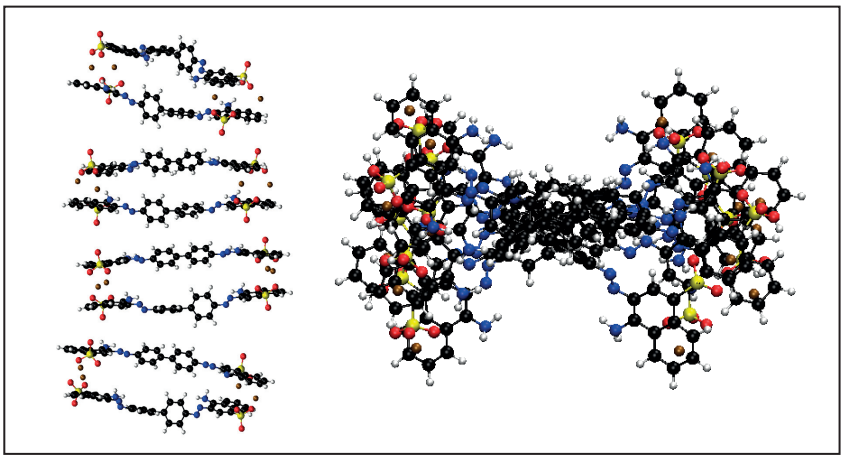
- Versatile and widespread used quantum chemistry codes such as **Gaussian**, **GAMESS UK**, **NWChem**, **Schrödinger**, **Q-Chem** and **TURBOMOLE**, which are capable of calculating electronic structure and various properties of diverse molecular systems using both *ab initio*, density functional 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 **(ADF, DFTB, MOPAC, COSMO-RS)** which 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 (**ADFInput**, **ADFViev**, etc.) ADF package is used by many of our users.

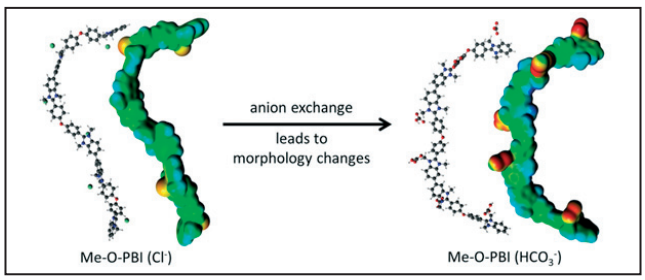
- Several packages, which could be used for **solid-state systems**. Among them **BAND**, **Quantum ESPRESSO** and **SIESTA** are worth mentioning.

- **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. Among software prepared to run on graphical processors our administrators' team prepared quantum chemical packages such as **GAMESS**, **Terachem**, **NAMD**, and **Quantum ESPRESSO**. 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.

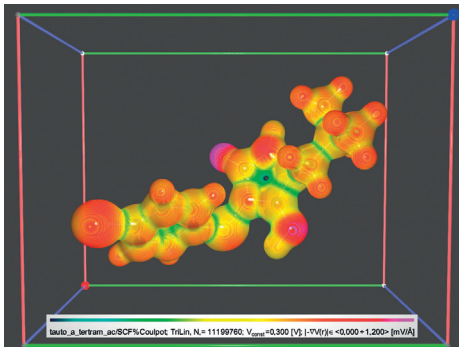
Scikit-learn is a software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN. It has been designed to interoperate with the Python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

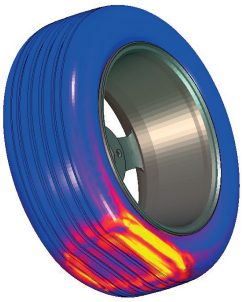
The most important modules are:

- 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 nowadays essential tools in the process of developing and building almost everything – from car parts to buildings. Through computer simulations, engineers can check 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 simulations package with 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 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.

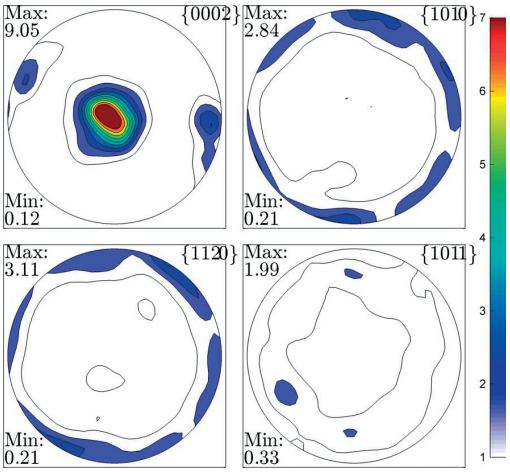
FLUENT software offers the broad physical modeling capabilities needed to model flow, turbulence, heat transfer and reactions for industrial applications ranging from air or liquid flow to semiconductor manufacturing. FLUENT can be used in numerous science domains, including chemistry, metallurgy, biomedicine,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

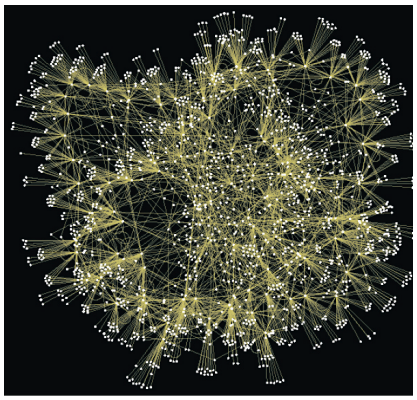


Bartosz Sułkowski: Results of texture simulations by visco-plastic self-consistent model of Zn after hydrostatic extrusion at 250 °C

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)

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.

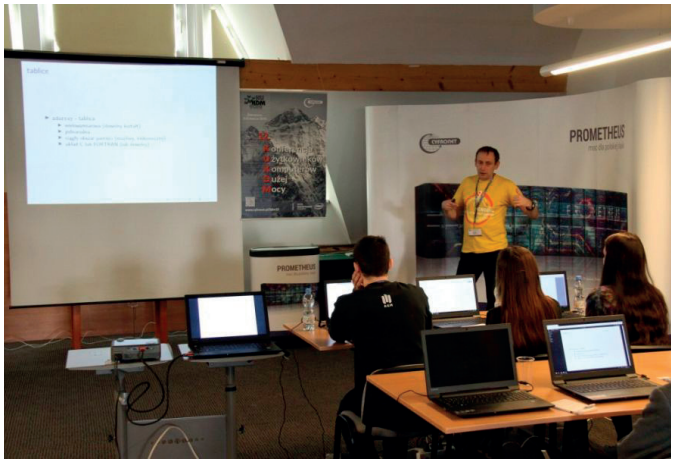
The first edition of the conference attracted much attention and increased the interest of users in Cyfronet resources. It proved that this type of event was much awaited and needed.

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.

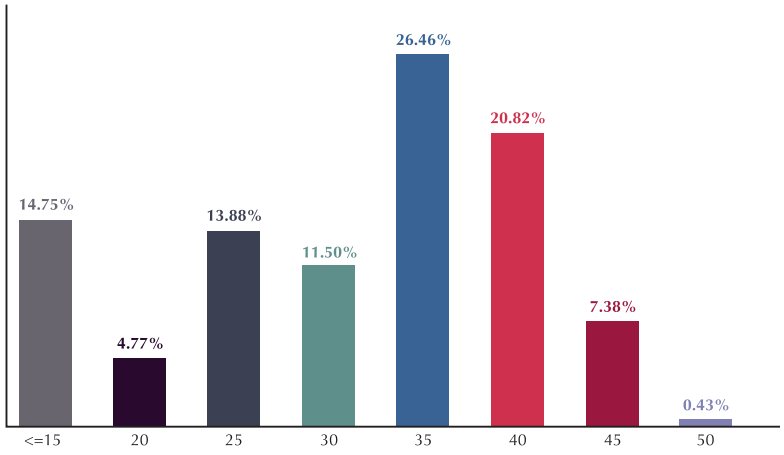
The crucial parts of the conference are meetings with suppliers of Cyfronet hardware and software, as well as the panel discussion on efficient use of these resources. The latter is always attended by users – researchers, who use the chance to get familiar with news regarding the computing infrastructure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



The conference is accompanied by poster and training sessions – in 2019 the training was focused on the programming of scientific calculations in the Python language.

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*).

http://www.cyfronet.pl/kdm19/



Ministry of Science and Higher Education marks of articles published in 2018 by Cyfronet Users in scientific journals

CGW Workshop

The CGW Workshop was organized for the first time in 2001 by ACC Cyfronet AGH, in cooperation with the Institute of Nuclear Physics Polish Academy of Sciences and the School of Banking and Management in Krakow. This event was related to the Cyfronet's participation in a large, international project CrossGrid (2002-2005), funded by the EU. Among the guests invited to the conference were eminent scientists, experts in the field of grid technologies: Fabrizio Gagliardi, Wolfgang Gentzsch, Jarosław Nabrzyski and Piotr Bała.

Since then the conference has been organized annually and has become a very important event in Europe. ACC Cyfronet AGH and Department of Computer Science AGH have been the organizers of the conference.

CGW Workshop was an opportunity for the presentation of research and development activities, supported by large-scale simulations, using grid and cloud technologies, and other computing techniques. The conference also provided the overview of research carried out in the EU and national projects, addressing distributed infrastructures.

In particular, scientific and technical achievements within the PLGrid Programme were presented, together with those related to the use of the PLGrid infrastructure.

The conference program included invited lectures and oral presentations of participants. In parallel, the poster session was organized, as well as the exhibition stand with the materials promoting the EU grid projects, and the ones realized under the PLGrid Programme. The conference was also accompanied by training aimed at familiarizing the participants with the latest tools and platforms,

which facilitated the use of grid and cloud infrastructures.

Contributed papers accepted for presentation at the conference provided a very good overview of the research activity in the area of e-Science and distributed computing infrastructures.

In addition to publication in the CGW Workshop proceedngs, the extended versions of selected papers were printed in the Computer Science Journal (CSCI), published by the AGH University of Science and Technology in Krakow. The journal is prepared in cooperation with many renowned computer science researchers from all over the world and its main purpose is to create a forum for exchanging research experience for scientists specialized in different fields of IT.

The Steering Committee of the CGW Workshop made every effort to address during the conference such topics, which were in line with the latest trends in IT. The result of these endeavours is the fact that each year the conference had a growing interest of researchers involved in the creation and development of information technologies, as well as the users of these technologies. The last edition of the conference took place in 2018.

EuroHPC – towards exascale computing

The European High-Performance Computing Joint Undertaking (EuroHPC JU) was established to radically develop existing European HPC infrastructure, so that it could provide European researchers with computing power comparable to the ones available in USA, China and Japan. EuroHPC unites 29 countries as well as private members with the aim to buy and deploy two exascale supercomputers that will be on the TOP5 list of the world's fastest computers.

The LUMI consortium

To achieve the EuroHPC goal, firstly pre-exascale machines will be operated. Poland, represented by ACC Cyfronet AGH and Ministry of Science and Higher Education, has joined the Finnish-led LUMI consortium, which works on the machine that will be ten times more powerful than the most powerful supercomputer currently in Europe (Piz Daint). The other consortium members are: Belgium, Czech Republic, Denmark, Estonia, Norway, Sweden and Switzerland.

LUMI, which stands for The Large Unified Modern Infrastructure, also means "the snow" in Finnish. And like the snow, it can completely change the landscape – in this case the one of data-driven research. The sum of the unique expertise and experience of building and operating HPC systems, managing Big Data, as well as developing and using the advanced software, shall result in creating an easily-accessible, efficient, safe and powerful HPC environment.

Pre-exascale supercomputer

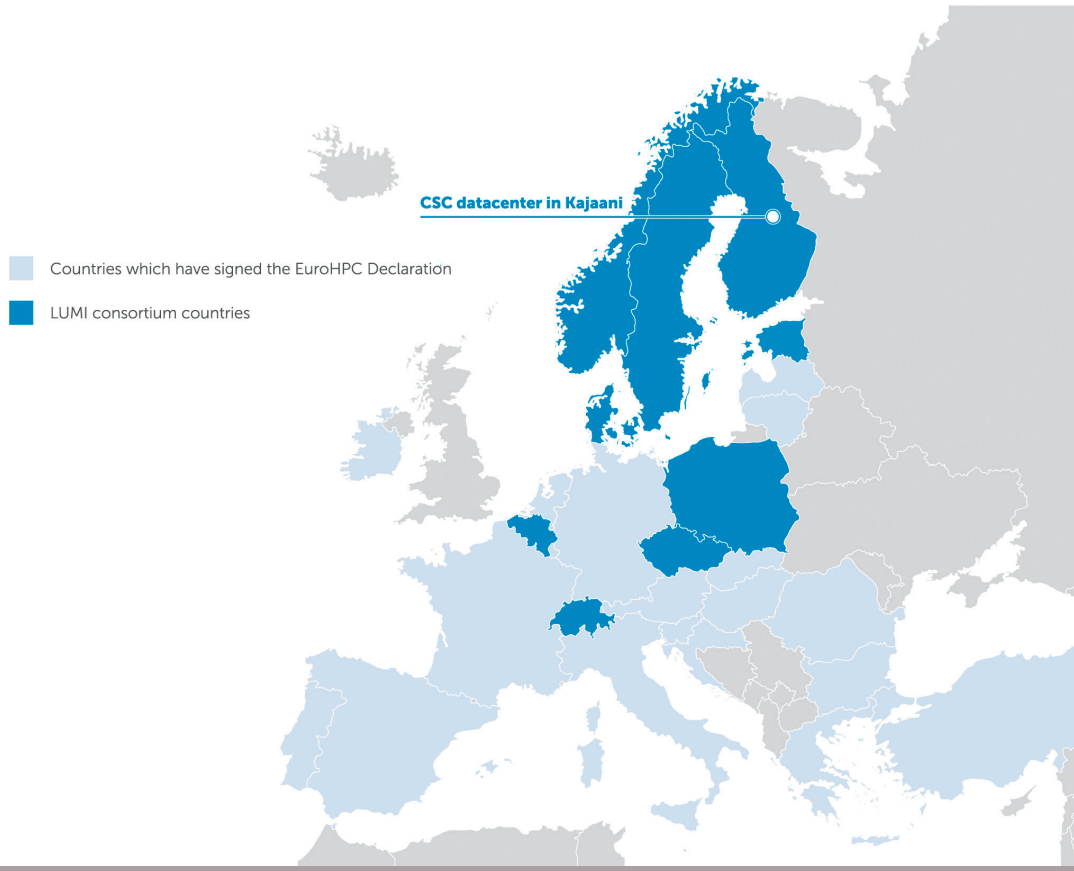
To meet the growing need for computing resources of academia and industry, the LUMI supercomputer is designed to have computing power exceeding 200 PFlops (0,2 EFlops), which will be achieved by a combination of General Purpose Processors (GPP) and GPGPU partitions. Jointly with over 60 PB storage and cloud services, it will give researchers the possibility to perform significantly more complicated calculations in shorter time, and faster process bigger sets of data. The expected result is a rapid development of many research directions, including chemistry, biology, nanotechnology, material engineering and other domains crucial for modern industry. The new supercomputer will also contribute to many achievements in areas such as e.g. astrophysics, weather prediction, seismic activity, personalised medicine.

"We will be able to use the supercomputer to solve the problems that are unsolvable, taking into account the currently existing machines – says Marek Magryś, Cyfronet's Deputy Director for HPC. – *The flagship example are brain simulations."*

At the time of installation, the LUMI supercomputer will be one of the world's fastest computer systems. It will be operated at the CNC data centre in Kajaani (Finland), but its computing power, storage and services will be available for Polish scientists. One of Cyfronet's goals is to provide them with an efficient and user-friendly working environment with this supercomputer, based on good practices developed and national infrastructure adapted to new conditions.

Carefully selected site

The choice of Kajaani data centre was preceded by a thorough analysis of possible sites, as they should fulfil the strict EuroHPC criteria. CNC data centre is situated in Finland and due to the climate it benefits from natural cooling provided all-year-round. Kajaani is also equipped with three local hydro power plants (30 MW) and a biomass power plant (88 MW) which guarantees green energy delivery. And the waste heat from the supercomputer will be utilized to warm data centre and surrounding buildings, what will reduce the negative impact on the climate.



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. This international research foundation is one of three Polish beneficiaries of the prestigious *Teaming for Excellence Phase 2* call, as well as the only participant representing the Małopolska region.

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** 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 know-ledge 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.

An important aspect of the activities of **Sano** is its collaboration with the University Hospital in Kraków and its personnel.

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-Hub – a step closer to easily-accessible, transparent and data-driven science

The Open Science is expected to become one of the most important paradigms shaping the future research. This idea stands behind the European Open Science Cloud (EOSC), which is created with the aim of giving the researchers an easy access not only to data repositories, but also to efficient storage and data-management services (including e.g. searching, analyzing, re-using and sharing data).

EOSC-Hub is one of the key projects supporting that initiative directly. It started in January 2018 and since then has been playing the role of integrator of the resources of existing European e-infrastructures and research infrastructures. So far more than 240 services from participating providers, as: EGI, EUDAT, INDIGO-DataCloud and research communities, have been shared, out of which almost 80 are accessible via Marketplace, meant to become a single platform giving researchers opportunity to find means supporting their research, no matter the scientific domain they represent.

EOSC Marketplace, Portal nad AAI

The Marketplace, as the name implies, was designed for finding, ordering and accessing various types of products – which in this case are represented by eight sections: Compute, Data Management, Networking, Processing and Analysis, Security and Operations, Sharing and Discovery, Storage, Training and Support. With the use of browser with useful filters, the users are able to quickly find a service fitting they needs. Every service has its subpage with short description, links to documentation (including also terms of use) and Helpdesk. This structure facilitates the process of discovering and accessing resources, and thanks to that it supports data-driven research in multiple disciplines. The Marketplace is a substantial part of EOSC Portal (also operated by EOSC-Hub) and is to become the heart of future EOSC.

The Portal itself plays the role of a hub connecting supply- and demand-side. It allows the providers – which are European e-infrastructures and research infrastructures of different size (from local to multinational) – advertise their services, in that way broadening the audience of potential users. On the other hand, researchers obtain not only access to diverse resources, but also to instruction and technical assistance directly from the providers. One of the distinguishing features of the Portal is its aim to lower barriers of for both sides.

To ensure safety and reliability of the Infrastructure use, EOSC-Hub builds the Authentification and Authorisation Infrastructure (AAI). It's been developed upon the best practices from EGI Federation, EUDAT CDI, and INDIGO-DataCloud. Since some promising solutions have been implemented so far, EOSC AAI is to become a unified, secure, pan-European access mechanism for researchers.

Cyfronet's main responsibility is to develop the EOSC Marketplace platform and to host the EOSC Portal. Other contributions are related to Dirac services, Indigo PaaS and federating of storage.

EOSC-Hub will run until 2020 with funding from the European Union's Horizon 2020 research and innovation programme, as well as from contribution of participating units.

PRIMAGE
Medical imaging Artificial intelligence Childhood cancer research

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

EOSC-hub

The aim of the EOSC-hub project is 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.

PROCESS

The solutions developed within the project will be a breakthrough step in the creation of innovative, exascale data processing services, maximizing the benefits of modern data processing systems.

eXtreme DataCloud

The goal of the project is to build specialized solutions for managing and processing large-scale data in a hybrid cloud, thus introducing access and data migration in distributed cloud environments.

SERA
Seismology and Earthquake Engineering Research Infrastructure Alliance for Europe

The overall objective of SERA is to give a significant contribution to improve the access to data, services and research infrastructures, and deliver solutions based on innovative R&D in seismology and earthquake engineering, aiming at reduction of the exposure of our society to the risk posed by natural and anthropogenic earthquakes.

AARC

The goal of AARC2 was to design an AAI (authentication and authorisation infrastructure) framework to develop interoperable AAI, to enable researchers to access the whole research and infrastructure service portfolio with one login.

PRACE

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

EPOS
EUROPEAN PLATE OBSERVING SYSTEM

The EPOS Implementation Phase project (EPOS IP) built on the achievements of the successful EPOS preparatory phase project (EPOS PP). The EPOS project was integrating the diverse, but advanced European Research Infrastructures for solid Earth Science, and built on new e-science opportunities to monitor and understand the dynamic and complex solid-Earth System.

EurValve

The aim of the project was to develop a comprehensive, clinically-compliant decision-support system to meet the challenge of treatment optimisation in case of the Valvular Heart Disease, by quantifying individualised disease severity and patient impairment, predicting disease progression, ranking the effectiveness of alternative candidate procedures, and optimising the patient-specific intervention plan.

EGI

The aim of the EGI-Engage project was to accelerate the implementation of the Open Science Commons by expanding the capabilities of a European backbone of federated services for computing, storage, data, communication, knowledge and expertise, complementing community-specific capabilities.

VirtROLL

The aim of the project was the development of a model-based predictor system, supporting the flexible design of strip rolling, joining functionality of numerical simulations, material modelling, sensitivity analysis and optimization.

The main objective of the VPH-Share project was creation and implementation of a platform for the sharing of scientific applications implemented by the members of the European Virtual Physiological Human (VPH) consortium, as well as to adapt existing applications and data repositories for sharing within the framework of the aforementioned platform.

CROSSGRID

The project aim was to develop and implement new tools and services used to run interactive applications, which required high computing power and large data collections in the grid environment.

The main objective of the project is digitization, preparation of descriptions of digitized data and the final publication of digitalisers on the PAUart platform. 13 140 new records will be published on this platform, including graphics from the modern period, valuable in artistic and scientific terms, localised in the Print Room of the Polish Academy of Arts and Sciences (PAAS) and 19th and 20th century photographs from the Lanckoronski Phototheca of the PAAS and from the Archives of Science of the Polish Academy of Sciences (PAS) and PAAS in Krakow.

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

Within the project it is planned to develop a comprehensive diagnostic test based on analysis of the material obtained from the tumor via surgery or biopsy. This will lead to a liquid biopsy-type method based on tumor cfDNA in blood, which could be a quick and minimally invasive alternative to biopsy.

The project aims 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, allowing for easier integration of their solutions, specific to the selected fields of science, with the PLGrid infrastructure.

The project aimed at integration of new groups of researchers from the next 14 disciplines with the PLGrid computing infrastructure. Thanks to deployment of new domain grids, research teams from these areas are able to faster obtain results of their calculations and to better integrate with the national computing infrastructure for science.

The project supported, by means of IT, Polish research teams in conducting research and also enabled extensive collaboration among these teams, as well as international cooperation in the area of e-Science.

The preparatory phase of Polish input for Cherenkow Telescope Array Project. This project has been undertaken to help design high level tools easing CTA data analysis. The main area of work covered the development of InSilicoLab Science Gateway for CTA with focus on job results analysis and visualisation, in particular.

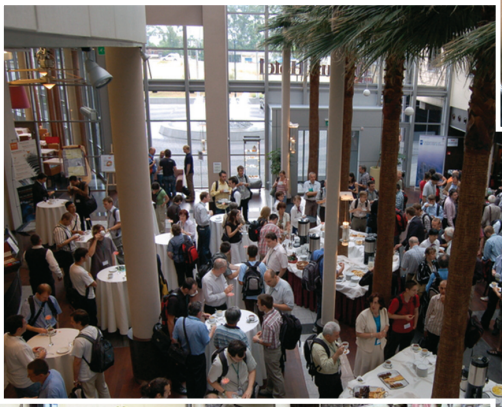
The aim of the project was integration of selected services available in the PIONIER network, and the development of the new services, e.g., with the increased reliability and security.

The 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.

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.

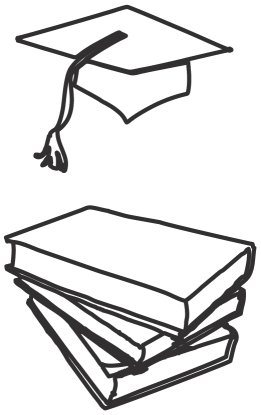
Within the project the Polish Grid Infrastructure (NGI) has been built to provide the Polish scientific community with an IT platform based on computer clusters, enabling research in various domains of e-Science.


AKADEMICKIE CENTRUM KOMPUTEROWE
CYFRONET AGH




www.CYFRONET.krak

Work of young scientists in Cyfronet



The contest for the best PhD thesis conducted with the help of computing resources of ACC Cyfronet AGH is a tradition in our Centre. Organized annually, it is becoming an important event promoting research performed by young scientists. For the 2019 edition of the Contest, participants submitted many PhD theses focused on variety of scientific problems in physics, chemistry, computer science and others. Also the utilisation of the resources varies, as the contesters used different parts of the Cyfronet's computing infrastructure: both supercomputers as well as clusters and the GPGPU platform.

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 2020 Contest edition!

http://www.cyfronet.pl/konkurs

The laureates of one of the previous Contest editions

Jakub Chęciński

The interview with the author of the PhD thesis: "Modeling of magnetization dynamics in spintronic oscillators"

What influenced you the most to choose spintronics as research direction?

I have always tended to be interested in the computational side of physics: all models are inaccurate to a certain extent, but what makes some models more useful than the others? Can we recreate the complex phenomena we observe in the real world if we start only from a simple set of building blocks and basic rules? What if we could tilt those rules slightly? At the same time, as an undergraduate student I had no realistic idea about which area of study could be a good fit for such an approach. Spintronics emerged as an answer to this question, since as a field it requires a highly realistic computational approach, but at the same time it gives you an opportunity to work in a close feedback loop with the experiment. I was lucky to find a group of highly competent and motivated researchers working in spintronics here at the AGH-UST, which has been a major advantage for me in terms of my scientific career progress.

How would you explain the subject of study to a non-specialist?

Nanotechnology and various types of nano-devices have already transformed the world we live in and are likely to continue doing so as one of the future industry pillars. However, experimental investigations of nanosystems, while indispensable, tend to be tied with significant costs in terms of both money as well as the researchers' time and effort. It is therefore very important to have tools that allow for reliable modeling of the nano-world, serving as a guidance and a source of scientific inspiration for the experimental teams. Due to the complexity of the physics involved, this approach almost always has to rely on numerical methods and often requires a great deal of computational resources. It is also the main motivation for my own subject of study: micromagnetic modeling, which focuses on understanding and predicting magnetic phenomena occurring in nano-sized systems, especially in spintronic devices which typically depend on their magnetic properties as a way to receive, store and process information.

What results regarding spintronic oscillators can we expect in the nearest future?

People enjoy speculating upon the next technological breakthrough, but in reality those tend to be notoriously hard to predict – after all, we are dealing with scientific research, which by its very definition involves exploration of the unknown. This being said, spintronic oscillators do show a lot of promising signs for the nearest future. It is especially true for antiferromagnetic oscillators. They constitute an entirely new subfield of spintronics, which could lead to major advances in areas such as extremely fast data transmission and processing. I am particularly excited about the so-called single-layer antiferromagnetic oscillators that could work based on a modified type of architecture, which was described theoretically just a few months ago. Another very promising direction right now

includes neuromorphic computations, which utilize oscillators as artificial intelligence building blocks, designing biology-inspired algorithms that can solve computational problems significantly faster than their "classical" counterparts. Spintronic oscillators fit this trend very well, because they combine GHz (or even higher)-range speed with very low energy usage and room temperature tolerance in a way few other hardware implementations could achieve.

Aside from creating theoretical models you have also programmed a tool aimed to facilitate micromagnetic simulations. How would you describe MAGE?

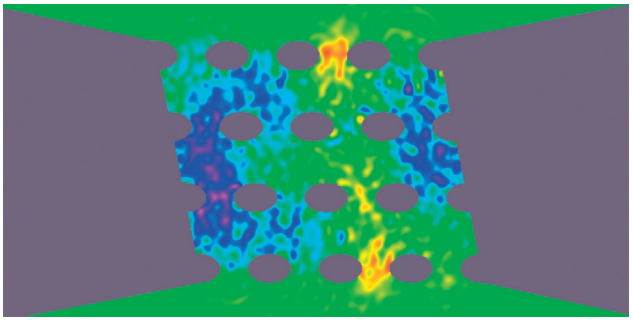
MAGE, or Mif/M-file Automatic GEnerator, is a software tool we have written to make designing micromagnetic simulations easier, especially for users that are less experienced or less inclined towards programming. It was my first independent project as a PhD student and served as a sort of methodological preparation for the later physics research I aimed to do in the area of spintronics. Together with my team we have published the code in a peer-reviewed journal specializing in scientific software tools and made the program publicly available on an online platform. Since then, it has been used regularly by researchers from both academia and industry, and we hope it can remain a useful tool for designing micromagnetic simulations.

Which of Cyfronet's resources have you used and how could they help?

Since my research has mostly numerical character, I have been relying on Cyfronet's resources quite heavily, starting from Zeus supercomputer back when I was a graduate student to Prometheus nowadays. It is really impressive how this infrastructure keeps growing over time, providing its academic users with very powerful computational capabilities while keeping bureaucratic procedures at practically zero. My foreign colleagues from the same area of study tend to be genuinely surprised by the accessibility and the sheer numbers of what we can do here, as their own research often has to be based on either small local clusters which offer very limited possibilities, or large and complicated computing systems with byzantine rules and long waiting queues. I think this can be a real source of competitive advantages for Polish scientists. In my own work, I particularly appreciate the infrastructure for computations based on graphical processing units (GPUs), since they gradually become the recognized standard in the field, with more and more software tools being rewritten or optimized with this architecture in mind.

From your experience, what are the most important things the PhD students should focus on?

I believe it is helpful to keep a clear vision of what kind of research projects would be both realistic and interesting to do. An advice I would give to everyone starting is: read the literature regularly to see what other people are working on, and be ready to adjust your own plans accordingly. For people dealing with modeling in particular, try to begin with very simple examples and then gradually increase the complexity, producing iterative results on your way. A lot of people overlook the availability of different kinds of resources that could make your efforts significantly more effective. Cyfronet and PLGrid infrastructure are actually good examples of that – spending one afternoon on learning how to use them could literally save you months of work in the future. Finally, if you have an opportunity to work abroad for a while (for example using the National Science Centre Etiuda program), it can be very helpful for your career, especially if you choose a group with a good research track in your field.



An example array of spintronic oscillators performing a neuromorphic computation. Different colors represent different relative phases of the magnetization

Karol Grzegorczyk

The interview with the author of the PhD thesis: "Vector representations of text data in deep learning"

What inspired you to get interested and then engage in the research on Machine Learning and AI?

Selection of the research area was strongly influenced by my supervisors. Their passion for conducting research on AI and Machine Learning inclined me to investigate the topic in my dissertation.

Why in your opinion is it worth to teach computers natural language processing? How would you evaluate the role neural networks may play in that regard?

Natural language processing is a very complex problem. It's very difficult to explicitly encode all the rules that govern the language. The fact, that natural language is a living creature, adds another complication. Some new words appear in the dictionary and at the same time some other fall into disuse. That's why self-learning systems, like neural networks, can prove themselves the best.

What are the advantages of vector representation in large text data-sets analysis?

The merit of hidden vector representations is their ability to model similarities. In original space every piece of text (i.e. single word, sentence or document) is represented by numeric ID. The algorithm doesn't know that similar words are similar, because each of them is represented by an ordinal number from the dictionary. When we use hidden vector representations, the algorithm is able to get to know that two words are similar to each other and two other are not. That property becomes crucial in many natural language processing tasks.

What kind of achievements do you expect to happen in the analysed area in the next few years?

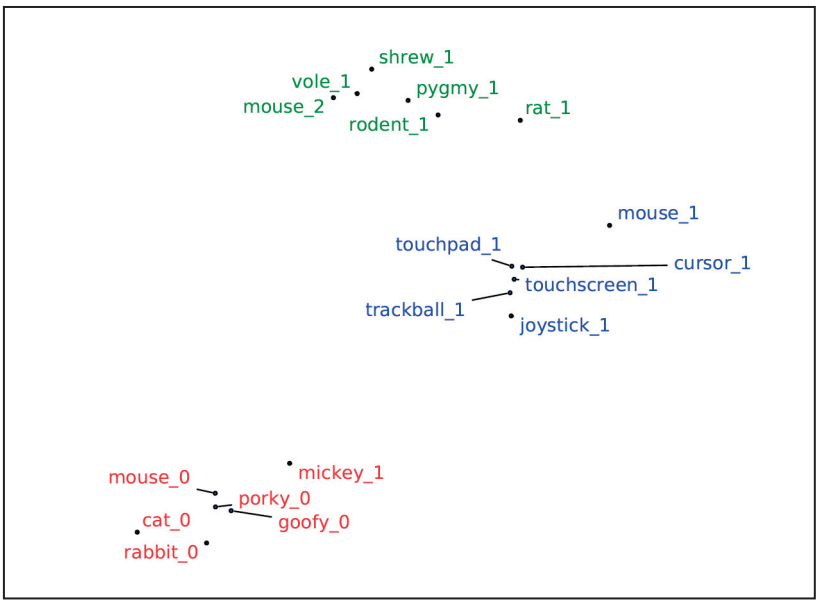
I suspect that within the next few years there will appear chatbots enabling long complex conversations. At the moment most chatbots are rather simple systems that answer the questions. The possibility to use information given by the user during a long conversation seems a key to creating intelligent assistants that imitate people.

Which of Cyfronet's resources have you used?

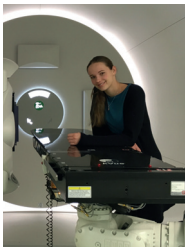
I have used Prometheus supercomputer, which is equipped in numerous strong compute nodes with as well traditional compute cores as well as GPGPU's.

What, in your opinion, should PhD students focus on? What advice could you give to them?

I think that at the point of choosing research area it is a good idea to be driven by an authentic curiosity and the will to know the world. At the beginning of the scientific career it is worth to devote as much time as possible to learning, self-development and broadening one's horizons. It is not worth to choose one's thesis theme too early. Moreover, I think that at the beginning of the career one ought to try to work on difficult issues and challenges in a given area. It gives you more freedom in the future career, as it is easier to move from more difficult issues to the simpler ones than in the opposite direction.



2-D PCA (Principal Component Analysis) visualization of nearest neighbours of three meanings of the word „mouse"



Magdalena Kłodowska

The interview with the author of the PhD thesis: "Application of Monte Carlo methods in transport modelling of the therapeutic proton beam"

How did your interest in medical physics develop?

I was just curious about radiotherapy. It all started from a bachelor thesis on radiation leakage through a multileaf collimator applied in conventional radiotherapy with photons. Then, during my Erasmus stay at the Stockholm University in conjunction with Karolinska Institutet, I took a multitopic semester course on radiotherapy. Next, I did my master thesis in Particle Therapy Centre (currently MIT) in Marburg. So an interest in proton therapy in Bronowice Cyclotron Centre (CCB) was just a natural consequence.

In which aspects can we think of your research as of a contribution to cancer treatment? How do you feel about that?

A computer beam model, which I developed within my work, is obligatory for a modern proton therapy centre. It allows to improve the precision of calculation of the dose delivered during the treatment. In my PhD thesis one can find description of the parameters needed for a reliable beam model to be built. Frankly speaking, I am somewhat proud that I could have developed such a model for CCB. It can be applied to any simulation concerning the usage of the CCB proton beam, not only in cancer treatment but also for experiments.

Could you shortly describe the advantages of Monte Carlo methods in dosimetry? On the other hand – what limitations of them can be perceived?

Monte Carlo (MC) methods can definitely improve the precision of the dosimetric measurements. It leads to better agreement between the dose calculated prior to the treatment and the dose delivered during the irradiation which, in general, results in an improvement of the treatment precision.
Moreover, in some particle therapy centres worldwide MC simulations replace the dosimetry measurements preceding the treatment, saving man-power and beam time. MC methods enable also the reconstruction of the treatment dose based on e.g. beta-emitters registered during PET (*positron emission tomography*) scan.
In each aspect, a reliable computer beam model is required, which must be developed for each therapy centre individually. Additionally, access to the efficient computational resources is mandatory, because high precision calculations require a lot of computational time.

Cyclotron Centre Bronowice (CCB) is one of the small number of such installations in Europe. Regarding that, would you place Kraków among leaders of innovative eye-cancer treatment?

Yes, eye treatment rooms with proton beam from state-of-the-art cyclotrons such as IBA Proteus C-235 are unique. In general, I find proton irradiation as an innovative method for treatment of some types of eye cancers and of course it is a great opportunity to do it here, in Kraków, at CCB. Nevertheless, for gantry installations, with two available at CCB, it becomes a new standard for cancer patients in almost every country in Europe.

What further challenges stand in the way of scientists of your discipline? What achievements do you expect in the nearest future?

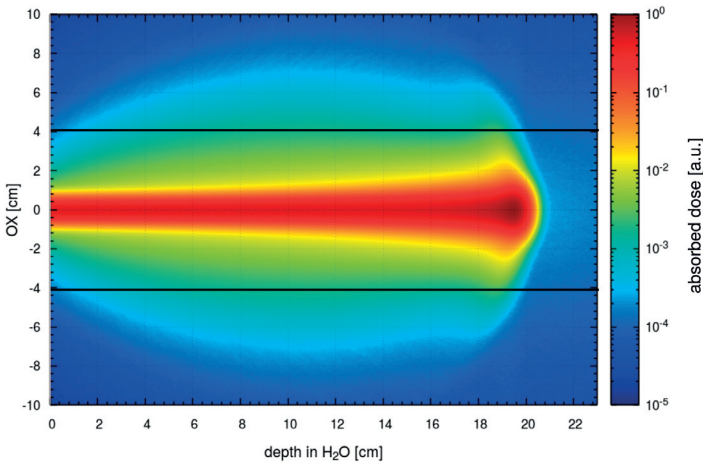
In particle therapy precision plays a very important role – energy deposited by e.g. protons highly depends on the density of the traversed tissue. Hence, research concerning the irradiation of moving targets such as lung cancers with beam triggered by patient breath for example, is the aim for the nearest future. The other researchers develop techniques of the *in-vivo* dose monitoring, which will enable better knowledge of the exact dose delivered to the patient. The most advanced studies, in my opinion, include individual therapy, based on radiobiological studies of the tissue response.

What did you use the ACC Cyfronet AGH infrastructure for?

It was used for Monte Carlo simulations to develop CCB computer beam model. Moreover, all the dosimetry calculations for CCB proton beam were done with ACC Cyfronet AGH resources.

By defending your PhD thesis, you have closed some stage. What advice could you give to people, who have just started it?

The longer it takes, the worse it gets – so try to finish your thesis as soon as possible. For me it was a great lesson of discovering my true interests, overcoming own weaknesses and getting more and more humble... For sure it was worth doing – good luck!



Dose deposited in water by 180 MeV protons, marked in colour. Simulations based on Monte Carlo methods enable calculation of dose deposited beyond the routinely used detectors (dimension marked with black lines)

Karolina Kula

The interview with the author of the PhD thesis: "A synthesis of new nitro-substituted analogues of dihydropyrazole in a reaction of 1,3-dipolar cycloaddition"

In which circumtances have you decided to become a scientist? What influenced you to choose chemistry as a discipline?

I've been always keen on chemistry. As early as in the primary school I associated my future with it, that's why I chose Chemistry Technical College in Krakow. This choice even stronger confirmed me in belief, that chemistry is what I want to do in my professional life. I decided to continue study at the Department of Chemical Engineering and Technology, Cracow University of Technology. After first year of study I started to work at Physical Organic Chemistry and Organocatalysis Research Group under the supervision of prof. Radomir Jasiński. Thanks to that, at the very beginning of my professional career I was able to see what the scientific work is all about. My research resulted in participation in multiple conferences, as well national as international, and in BEng and MSc theses defence. At the science club I have synthesized some connections, that hadn't been received before. After I graduated, all these factors induced me to continue research on PhD studies and to enrich my experimental research with quantum-chemical calculations.

During the research you have synthesized 16 new organic compounds. What properties could they have?

The connections synthesized by me mostly belong to the group of pyrazolines and pyrazoles. The majority of these class of organic compounds have a biological activity and are succesfully used as painkillers, anti-inflammatory and antipyretic medicines. What's more, these compounds are used as pesticides or dyes. Moreover, it's worth noticing, that in the structure of the pyrazolines and the pyrazoles, which I synthesized, two functional groups are present: nitro and trichloromethyl. First of them, due to its high reactivity, is a valuable molecular building block. Furthemore, a nitro may be converted to various other functional groups. It may be a precursor in preparation e.g. nitronates, amines, hydroxylamines, oximes, hydrazines, azo compounds, nitriles and their oxides, and carbonyl compounds. It creates the wide scope for further functionalization of such designed target structures. In turn, trichloromethyl group stimulates the biological activity of many heterocyclic structures. The compounds containing this group are used e.g. as building blocks in peptide preparation and as insecticides or rodenticides. In addition, I have obtained nitrilimine bromides, which are the precursors of valuable TAC's components, used in the synthesis of heterocyclic connections in a reaction [3+2] cycloaddition.

What were the biggest obstacles you have encountered and how did you overcome them?

Undoubtedly, the most difficult phase of my PhD research was the synthetic part. The tasks included among others designing receiving pathways for individual compounds, their isolation, purification,

and then identification of the obtained connections based on elemental analysis data and spectral characteristics. All of above mentioned tasks require knowledge of appropriate skills and tools, and additionally, they are time-consuming and expensive. Regarding that, I have decided to perform quantum-chemical calculations to limit workload related to the experiment.

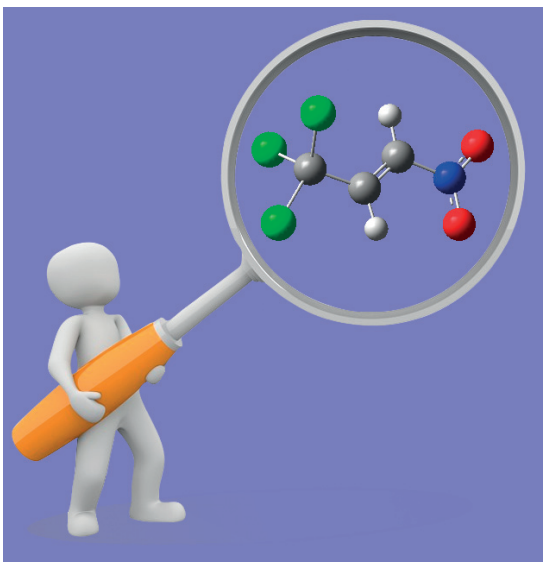
On the basis of quantum-chemical calculations you have created the models that then served for performing experiments. In this respect, how did you benefit from Cyfronet's resources and what are the advantages of such sequence of actions?

Thanks to Cyfronet's resources I was able to prepare well for performing the experimental part of my PhD research. Mainly it were quantum-chemical calculations that let me choose the proper conditions to analysed reactions. This approach allowed me to shorten time of simulations by experimental way, as well as save expensive, sometimes harmful, reagents. The use of quantum-chemical calculations for organic chemistry research is especially important because of the fact that in majority of reactions we are dealing with formation of more than only one product. Connections of individual atoms in forming compounds and their steric arrangement are of great importance, and theoretical considerations allowed me to perform simulations of hypothetical pathways for considered reactions, and also helped me to define, which of the possible structures is the most privileged to form in a given reaction.

On what actions should, in your opinion, focus the people who have just started their science career? What could you advise them?

As a person doing in my PhD dissertation as well experiment-, as theoretical-based research, I think that good theoretical foundations and manual skills are extremely important. These two factors should harmonize. That's why I encourage all young people to make use the most of their studies. As a lecturer working with students I would primarily suggest to seek for science problems, try to solve them independently, but don't be afraid of asking questions.

However, man does not live by science alone. Chemistry is of course not only my job, but also a passion, nevertheless it's important to have other interests. At the time I was writing my doctoral dissertation, long-distance running helped me a lot. At such a hard for me period of intensed mental effort, sport allowed me to take a rest. For all of us, but especially for the young, support from the closest people is of great importance. Regarding that, I would like to extend my heartfelt gratitude to my Supervisors, Parents, Fiancé and all kind people for continual support and cheering me in the most demanding moments of working on my dissertation.



The structure of (E)-3,3,3-trichloro-1-nitroprop-1-ene used as dipolarophile in reactions of [3+2] cycloaddition for obtaining pyrazoles and pyrazolines

"Science walks forward on two feet, namely theory and experiment"
Robert A. Millikan, American physicist, Nobel Prize laureate

Klaudia Maj

The interview with the author of the PhD thesis: "Measurement of the azimuthal anisotropy in Pb+Pb collisions at $\sqrt{s_{\mathrm{NN}}}$ = 5.02 TeV with the ATLAS detector at the LHC"

What had the strongest impact on your decision to study heavy ion collisions and quark-gluon plasma?

By studying heavy ion collision products we can find answers to many questions about the initial conditions of the universe. I came across the concept of quark-gluon plasma while writing an engineering thesis. Plasma is extremely hot and very dense nuclear matter, which behaves more like a liquid with superfluid properties (i.e. a liquid with very low viscosity) than like a plasma gas. It is expected that the universe was filled with such matter in the first fractions of seconds of its existence. I chose high energy physics and heavy ion collision as the subject of my master's thesis as there is quite a large group of physicists in Krakow involved in this field. During my studies, I joined the ATLAS experiment and conducted my first analysis of heavy ion collisions based on data collected at the LHC. I was studying the anisotropic flow of particles produced in lead-lead collisions. The subject attracted me so much that I decided to continue my research as part of my doctoral dissertation.

You conducted research using the ATLAS detector in the Large Hadron Collider in Cern. How could you describe work in such a large international experiment?

Working in such a large collaboration is a rewarding but often challenging experience. Each member of the experiment is assigned his tasks, which later form a beautiful whole in the form of an efficient detector and high-quality published content. The awareness that you have some influence on the operation of entire cooperation is very encouraging. At the same time, in such a large experiment, you can't just focus on your research. Cooperation with others is very important. The data analysis itself is a multi-stage process, a team of several people is working on one issue, and progress is regularly presented to a larger working group. The ATLAS experiment has also an advanced privacy policy, according to which prior to publication, results of research undergo a multi-stage internal review process to eliminate any possible factual or editorial errors. This is a necessity, because the author is everyone, not just people directly involved in the given analysis. In this way, all Cooperation takes responsibility for published content.

What was the biggest challenge in your work?

For me, the most difficult part of work is public speaking. And while working in the ATLAS experiment there are a lot of such: from presentation of work results in the working groups' forum to speeches at international conferences. However, the main obstacles in obtaining measurement results are hardware limitations. Colliding nuclei of heavier atoms, such as lead, is associated with a huge number of collisions occurring at short intervals. It is physically impossible to record exactly every

case observed in the detector. Therefore, for each analysis, experimental data filtering systems are created that select and save only interesting cases on disks. Despite this, the amount of data to be processed in one analysis is of the order of terabytes. In addition, there is also the production of Monte Carlo simulations, which are an important element of any analysis, especially when comparing whether the theory correctly predicts what we can see in the experiment.

How can heavy ion collision studies affect the whole of society?

Studies of products resulting from collisions of heavy ions do not generate direct, visible to the naked eye, impact on our daily lives. However, like any basic research, it has long-term effects. The obvious advantage of performing these studies is learning about the world around us and the rules that govern it. All science or all the things we use every day, used to have their origin in theory. In addition, studies of proton collisions or heavier ion collisions require specialized equipment. Therefore, new technologies are developing – e.g. electronics used in detectors must be more and more accurate, more efficient and occupy less space. Patents and ideas often arise during such experiments, and are later transformed into industry and everyday use. I think that this type of research has a positive effect on all fields of science.

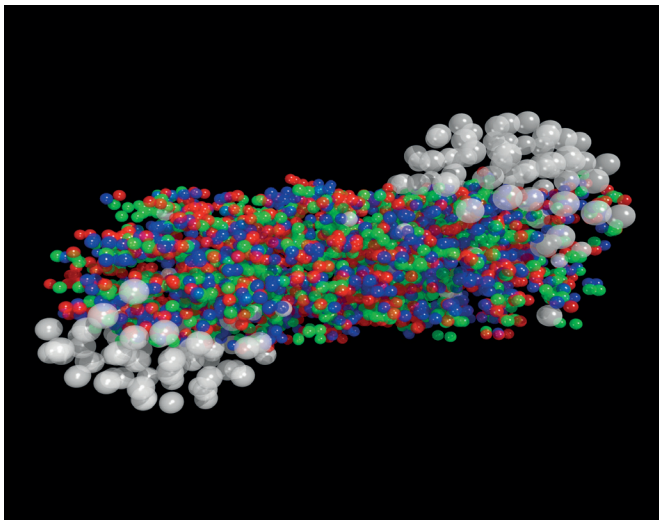
How did Cyfronet resources support your work?

The ATLAS experiment provides such a large amount of experimental data that it can be analyzed only with the help of distributed computing power, in this case in the Worldwide LHC Computing Grid (WLCG), to which ACC Cyfronet AGH also belongs. In addition, the Krakow group of the experiment carries out physical analyses locally on the Zeus supercomputer, which significantly speeds up research. The analysis presented in my doctoral dissertation required processing of terabytes of experimental data. To obtain satisfactory results, it was necessary to "review" millions of lead-lead collisions recorded on the disks of the ATLAS experiment. By storing data in Cyfronet, it was possible to carry out quick tests and eliminate errors at an early stage of the analysis.

Can you share advice with people who have just started or are planning to start doctoral studies?

In my opinion, the most important thing is to choose the right supervisor, and if there is no agreement between the doctoral student and the supervisor, do not be afraid to choose another one. One can say that the supervisor initially controls the career of his student. The good supervisor acts as a partner with whom you can always discuss the next stages of work and who tries to solve any problems on an ongoing basis.

It is definitely good to go on all available courses, conferences or training. Talks with employees of scientific centers from around the world allow not only to polish up the language, but also to get to know a completely different point of view on a given issue, or to establish contacts that will result in future cooperation.



Ion collision simulation.
Autorstwo: CERN, Henning Weber

Marcin Moździerz

The interview with the author of the PhD thesis: "Multiscale modeling of Solid Oxide Fuel Cell stack"

How would you explain the issue of fuel cells to someone outside the field?

Currently, the biggest challenge in the field of energy production is achieving an environmentally neutral generation and high efficiency. Fuel cells are devices that combine both these features. Conceptually, they are similar to lithium-ion batteries or lead batteries. However, unlike conventional galvanic cells, they should be supplied with fuel. Fuel cell operation is simple: the continuously supplied fuel is electrochemically oxidized, resulting in a potential difference, i.e. voltage. Water and heat are the only by-products of the process (when using pure hydrogen).

What are the advantages of this solution compared to widely available energy sources?

First of all, these are devices that produce less pollution. Hydrogen is produced on an industrial scale mainly in the process of reforming hydrocarbons, which leads to the creation of certain amounts of CO and CO_2. In spite of that, the amount of compounds created in the production of fuel cells, which are harmful to humans and the environment, calculated per unit of produced energy, is lower. This is due to higher cell efficiency than conventional energy production methods. It is associated with minimizing the energy transformation chain: the chemical energy contained in the fuel is transformed directly into electricity. SOFC fuel cells maintain high efficiency in energy generators of any scale, from small household power systems (approx. 50%) to large combined power plants with gas and steam turbines (approx. 70%). For coal power plants, this efficiency is around 40%. In addition, fuel cells enable the use of renewable energy sources for clean hydrogen production (for example, by means of electrolysis), and thus eliminate supply instability. Hydrogen can be used as an „energy store", while a fuel cell can be used to convert the chemical energy of hydrogen into electricity and possibly heat, depending on current needs. In addition, as hydrogen production can take place locally, fuel cells can be used to the disperse generation of electricity, for example in hard to reach places.

What is innovative in the EAST SOFC cells analyzed in your team?

Fuel cells are divided into many types. One of them is a high-temperature solid electrolyte fuel cell, called SOFC (solid oxide fuel cell) for short. This is the type of fuel cell that is most often referred to as the basis of large energy systems. SOFC fuel cells are characterized by low power density. Therefore, individual units are combined into so-called stacks that allow to generate large power. Merging cells into stacks is connected with various technical problems. For example, sintering large surface electrodes or sealing stacks is a challenge. During the preparation of my doctoral dissertation, I had the pleasure to work on activities related to the implementation of the grant of Grzegorz Brus

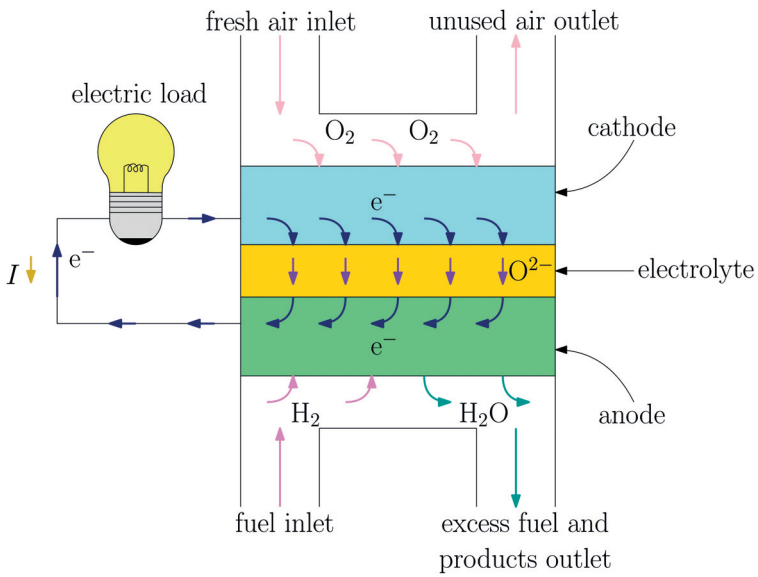
called "Development of a new type of fuel cell stack for the needs of the Polish energy sector", financed by the Foundation for Polish Science. The goal of the project is to develop the so-called EAST (Easy-to-Assemble Stack Type) stack. Compared to classic designs of SOFC-type fuel cell stacks, the EAST stack allows for easy sealing and microstructure design for a specific place in the stack.

Modeling of cells in the multi scale required a lot of calculations. How did you use Cyfronet resources?

During my work, I prepared calculation codes in C++ and Python, allowing simulation of the behavior of individual fuel cells and stacks by methods associated with those known from computational fluid mechanics. Computer simulations are a group of low-cost methods that allow to find weak points and propose improvements on any scale, from quantum-chemical simulations of material properties to simulations of large systems based on SOFC fuel cells. The most computationally demanding problem was building and solving systems of equations with huge rare matrices. Solutions to such problems require powerful computing resources and without modern supercomputers such as Prometheus, this would not be possible.

How do you assess the reality of the wider use of SOFC fuel cells in the economy?

I believe that fuel cells will allow to produce energy in a highly efficient and more environmentally neutral way than at present. In addition, I think they will help to solve the problem of energy storage. Currently, most of the electricity is generated on demand, while fuel cells combined with renewable energy sources allow to store energy in hydrogen and use it when needed. This is not a plan for the future – many companies already sell SOFC fuel cells for home energy production. However, there are still many obstacles to widespread commercialization. The most important challenge is the high operating temperature, which results in a relatively long time needed to start the system and determines a relatively short electrode life time due to thermal stress. The lack of large-scale hydrogen infrastructure is also a problem.



SOFC fuel cell operation scheme

After defending your doctoral dissertation, you closed a certain stage. What tips could you share with those who are just opening this stage?

It was a difficult but extremely interesting path. I could advise persons starting their doctorate not to be discouraged, no matter what administrative obstacles or research problems they encounter during their studies.

Monika Staś

The interview with the author of the PhD thesis: "Conformational properties of oxazole-amino acids"

Since when have you been interested in chemistry? When did you decide to enter the scientific path in this field?

My interest in chemistry began in secondary school, when I read Maria Skłodowska-Curie's biography and observed the experiments carried out during the lesson. However, I really loved chemistry in high school. I will not forget that summer day when we discussed chemical balance in class. To solve arithmetic problems, a quadratic equation was needed. I was delighted that the information about functions, equations or inequalities was useful for solving problems in which the numbers changed into the mass of precipitated sediment or evolved gas. I understood then that mathematics is the language of the world of natural sciences. After high school, I chose bachelor, master and doctoral studies in the field of Chemistry (at the Faculty of Chemistry of the University of Opole). Now, chemistry is a way for me to understand the world, the processes that govern it, as well as people.

In your research, you combined extensive literature research, quantum chemical calculations, molecular modeling, laboratory synthesis and spectroscopic analysis. What was the biggest challenge with a set of such different activities leading to one goal?

Different knowledge and skills were required for each type of action I took. Good theoretical preparation and planning of all stages were also important. In the introduction to my dissertation I have included the formulas and brief characteristics of over 250 peptides described in the literature, produced by living organisms, in which the amino acid residues studied by me were present. It was tedious work, especially since I looked for similarities in each structure to group them together. Laboratory synthesis is time consuming, requires patience, efficiency in the laboratory, as well as knowledge of various laboratory techniques and the identification of obtained products. Spectroscopic analysis requires good sample preparation and analysis of the results obtained. Calculations are not as time consuming as working in a laboratory, but they require regularity, accuracy and critical look at the results obtained. So the biggest challenge for me during the implementation of the intended goals of the dissertation was to combine all the studies that complemented each other.

How can research on the conformational properties of atypical amino acid residues find practical application in medicine or other fields?

The results of the research carried out as part of my work to some extent enrich the knowledge about understanding the mechanism of protein folding. Conformation, i.e. the arrangement of all atoms in space, is an important element of the native form of proteins and peptides. It often decides about their biological activity. Some neurodegenerative diseases (Alzheimer's, Creutzfeldt-Jakob disease) are caused by improper protein formation. The residues I study are produced among others by bacteria and reveal antibacterial properties, also strains resistant to current antibiotics. By incorporating these residues into peptides, their conformation can be modeled, which can translate into their biological activity. My research can be used to design peptides and proteins with specific structure and properties. And these compounds can be used to create new medicines.

How do you assess the current possibilities of designing peptides with the desired properties and what development in this area do you expect in the coming years?

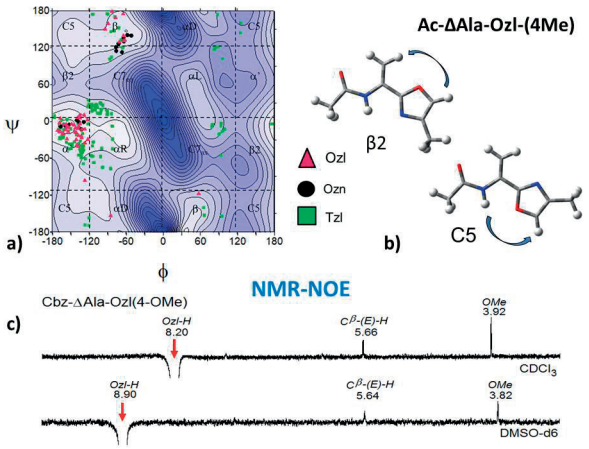
Nowadays, knowledge allows the design of peptides with increased stability, proper folding, changed solubility and resistance to digestion. Peptides can be created by modifying existing compounds or designed from the beginning (*de novo*). Such compounds may be used in biomedicine and bioengineering as receptors, enzymes and detectors. Modeling of peptides and proteins also allows to understand the mechanisms that occur in organisms with their participation, and their functions, including how the immune or endocrine systems work, or the causes of some diseases, so you can create effective medicines or vaccines. I believe that in the future the size of the tested systems, the accuracy of calculations and their sophistication will increase, as well as the number of variables taken into account, to the extent that the studied mechanisms and chemical reactions involving peptides and proteins resemble the cell environment of living organisms.

How did the resources made available by Cyfronet help in your research? What solutions did you use?

I carried out quantum-chemical calculations mainly on the Zeus supercomputer via the PLGrid platform. I used the Gaussian program in version 09. Thanks to computational grants at Cyfronet, I calculated the conformational maps of 6 compounds, and based on them I determined the energy minima for the 13 studied oxazole, oxazoline and thiazole amino acid residues in various solvents. Without theoretical studies I would not be able to interpret NMR-NOE and FT-IR spectra of the obtained compounds and finally give the characteristic conformation of the examined residues.

What, in your opinion, should be particularly considered by those who are just starting PhD studies? What advice would you give them based on your own experience?

In addition to expanding knowledge and skills in one's field and improving English, I would also tell them that you have to be curious, learn from everyone, experiment and not give up, because sometimes failure can lead to discovery. It is also important that participation in conferences is not only about presenting your results, but also allows you to meet many inspiring people. For me, the doctorate was a period that changed my view of science and the world forever. Thanks to it, I learned a lot, traveled to many wonderful places and met amazing people. It is worth undertaking this effort.



a) Conformational map of oxazole-amino acid (Ac-(S)-Ala-Ozl(4-Me)) calculated by the M062X/6-311++G(d,p) method, together with conformers' labels and marked values of torsion angles φ, ψ of the amino acid residues occurring in crystals from the CSD base, b) Arrangement of hydrogen atoms depending on the conformation, c) NOE differential spectra of one of the tested compounds – Cbz-ΔAla-Ozl-(4-COOMe) in various solvents

Piotr Iwo Wójcik

The interview with the author of the PhD thesis: "Random projection in deep neural networks"

What attracted you to neural network algorithms? How did you get interested in them?

Let's start with the fact that neural networks are one example of machine learning models, and it is machine learning and artificial intelligence that interest me most. The main task of artificial intelligence methods is to create software which is able to implement complex, non-algorithmic tasks, usually performed only by people, such as driving a car, playing Go, Starcraft or Quake, or a very important problem of identifying cats in photos. So, in short, artificial intelligence is the most interesting activity which computers can do at present.
I have been interested in machine learning for a long time, while my attention was directed to the neural networks at the beginning of doctoral studies, by the supervisor of my work, prof. Witold Dzwinel. This turned out to be a very good choice, especially in the context of the current popularity and dynamic development of the network.

What effects can we achieve for machine learning by reducing the dimensionality of the input data?

Reduction in dimensionality allows primarily to reduce the computational cost and memory consumption of such methods. In the case of computationally demanding machine learning models, such as neural networks, the reduction in dimensionality even allows the use of such methods on data too large for the model to be directly trained. Learning in a low dimensional space of traits requires fewer learning examples, so it can run faster. In addition, depending on the method of dimensionality reduction, it can also have a regulatory effect, i.e. reduce the risk of the model fitting too much to the data.

Why does the random projection method work for massive data sets with a large number of input features?

The main reason for using the random projection method to analyze huge data sets is simple: it is very effective in both computational and memory terms. In addition, it can work online, i.e. record by record, without the need for loading the entire file into memory.

What was the biggest challenge for you during the research? What obstacles appeared and how did you deal with them?

The biggest challenge came from the very nature of my work – I dealt with teaching models on very large data sets. The size of these sets caused that a special infrastructure was needed to allow efficient performing of calculations. Fortunately, Cyfronet has got just such an infrastructure. It is worth adding that this restriction also prevented other researchers from exploring this topic, thanks to which I entered the relatively unexplored area with my work.

On the other hand, for me personally, the biggest obstacle was the need to divide my time among working on my PhD thesis, professional obligations and other life activities. I overcame this obstacle by doing my PhD almost twice as long as assumed four years :). However, I do not regret it, because the fact that I did not limit myself only to working on my doctoral thesis, allowed me to gather a lot of valuable experience that I am currently using in my professional work.

Learning deep neural networks requires considerable computational time. What support in this respect are Cyfronet's resources?

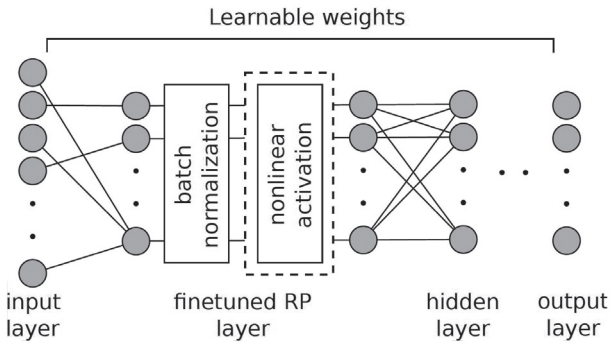
Invaluable. But to give some idea, let's count: during my research I used about 200,000 computational hours and it is rather an underestimated value. The vast majority were neural network training tasks intensively using Tesla K40 graphics cards, with a computing power of about 5 TFlops each, available in the Prometheus supercomputer. If the same calculations had been performed on the popular commercial Google Cloud Platform ML Engine and used the basic „standard_gpu" machine with over 50% faster Tesla K80 graphics card, the calculations would have probably taken over 125,000 hours. Currently, the price of one hour of training on the „standard_gpu" machine costs $ 0.93. So if I were to repeat the GCP calculations today, the total cost would be over PLN 400,000. So, effectively, without the support of Cyfronet's resources I could not carry out my research.

What next steps can be taken based on your results?

Continuation of work could go in two directions. First, the method itself could be improved, for example, by using the recently proposed dense structured random projection matrices that are a particular example of the Toeplitz matrix. The second direction could be the application of the proposed method in areas where high dimensionality and amount of data prevents the use of neural networks, such as for the analysis of genetic data.

In your opinion, what should people pay attention to when entering the scientific path?

First of all, I would like to draw their attention to the fact that a scientific career, at least in the field of machine learning, is not inseparably connected with the university. Many companies have now great R&D teams in which one can work and develop scientifically. These teams publish the results at the best conferences and actively participate in scientific life, also cooperating with universities.
However, regardless of whether someone chooses a company or university, I think that choosing the right team is essential. The one that really allows for development, provides support, but also a good atmosphere. I think this is a tip that works in almost every professional field.



Neural network with a trained layer of random projection

Grzegorz Zając

The interview with the author of the PhD thesis: "Spectroscopic studies of chiral xanthophylls and their supramolecular aggregates"

The analysed xanthophylls are strong antioxidants used in dietetics, medicine and cosmetics. How did you find a research niche, despite of the interest of this topic and expenditures of industry companies?

Xanthophylls, or other carotenoids, are rather a popular research topic. The majority of science papers regard analysing their impact on an organism or the impact diseases have on carotenoids level in the organism. In my research, for analysing the structure of xanthophylls I have focused on the use of a relatively new spectroscopic method: Raman Optical Activity (ROA), which is a chiroptical extension of Raman spectroscopy. So far the ROA method hasn't been used in analysis of the structure of xanthophylls. Additionally, if we take xanthophylls' aggregates into account, from literature we get to know that such exist and that they could lead to change of properties like bioavailability or persistence, whereas their structure hasn't been fully examined yet. Moreover, many xanthophylls are chiral and create supramolecular structures, therefore I decided to make the use of the ROA method and other chiroptical methods like ECD (Electronic Circular Dichroism) in my research.

During the research you have made the use of different spectroscopic analysis techniques, modelling with several calculation methods and also laboratory experiments. What is the biggest challenge related to such a diverse set of tasks?

The biggest challenge, which turns out to be an advantage and a disadvantage at the same time, are frequent changes. Honestly speaking, in my case it was very hard to get to theoretical calculations after months of experimental measurements – and vice versa. However, on the other side, this diversity prevents from routine and it's hard to get bored, since moment after moment we do something completely new, which requires other abilities. Additionally, it gives the possibility to develop many new skills that shall be useful in the future science work.

Could you point out the milestones of your research?

It seems that the first milestone was the moment I got interested in chiroptical methods. As early as at bachelor studies I become fascinated about chirality sensitive methods in determination of absolute configuration of chiral molecules. What interested me, was that with the use of Circular Dichroism or ROA we are able to determine the absolute configuration of the chiral compound's stereocentre, without the need of crystallization, which appears in conventional method used in this case (X-ray crystallography).

Having in mind that ROA spectroscopy has the low sensivity, I got interested in the research on resonant amplification, which could make it possible to analyse the structure of drugs in low

concentration and short time. A research on carotenoids was suggested to me by my supervisor – prof. Małgorzata Barańska. At the beginning, during MSc studies, I studied the structure of carotenoids in unbound form, trying to get resonance spectra of ROA. After many failed attempts, and after I had started PhD studies, I suggested an experiment with aggregates of carotenoids. Eventually, that first experiment finished successfully and became a milestone of the whole research.
The other important achievement was understanding that the obtained results prove not only the resonant amplification of the signal, but also an additional amplification related to aggregation. Then the name AIRROA (Aggregation-Induced Resonance Raman Optical Activity) appeared.
The results of ECD calculations were also very important. They showed that even simple models of carotenoids' aggregates (dimers) and unordered models can give the similar spectroscopic answer to experimental spectra. That additionally changes the previous idea of the structure of analysed aggregates, which in literature are rather described as highly ordered structures composed of many monomers.

What further steps can be made based on the results of your research?

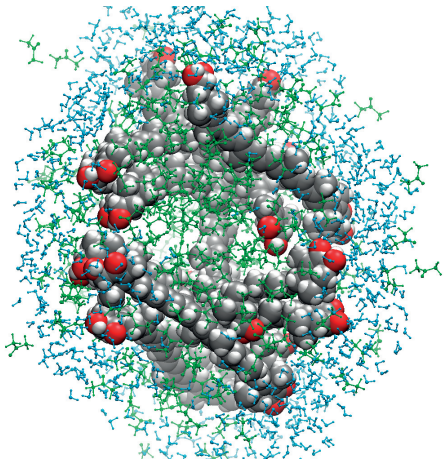
The next step, which is in progress in cooperation with prof. Petr Bouř from the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, are theoretical calculations of the AIRROA effect. Their aim is a deeper recognition of the additional effect of signal amplification, related to aggregation, and proposing a development of resonant ROA spectroscopy theory. After that, I plan to apply the discovered effect to research on different groups of compounds.
If we take practical implementations into consideration, the effect may be used by pharmaceutical industry, similarly to other chiroptical methods. The obtained resonant signal is so strong that it allows to take measurements of compounds with a biological concentration. We can imagine the possible application of this method in analysis of a three-dimensional structure of new drugs in natural environment or in the connection with an active centre of enzyme.

How did you make the use of Cyfronet's infrastructure?

In my research I have mainly used the computing power of Prometheus supercomputer, connecting via the PLGrid infrastructure. I have used Gaussian software in quantum-chemical calculations and Amber for molecular dynamics calculations. The number of computing hours I have used during the PhD research was so high that without support from Cyfronet I wouldn't get results within a reasonable time.

Based on your experience, what advice could you give to people who are at the beginning of their scientific career? What should PhD students focus on?

I think that differentiation of applied methods, diversity of research and working at the border of multiple disciplines is a key to success. It's not always possible, but I think, that if we have the chance to learn something new, what exceeds our research field, we should at least try, as it can become useful in an unexpected moment. PhD students should primarily get to a subject that is fascinating and gives them a lot of joy at the same time.



Astaxanthin aggregate model obtained on the basis of molecular dynamics simulation

CFD and FEA in AGH Solar Boat

Igor Łukasiewicz
AGH Solar Boat Team

AGH Solar Boat is a group of students who share a common goal to build and develop a solar boat that takes part in numerous international competitions such as Solar Sport One and Monaco Solar & Energy Boat Challenge. Our activities are aimed at using modern technologies such as: photovoltaics, lightweight composites, batteries, propellers and hull designs to build boats powered by solar energy. Through this we promote Renewable Energy Sources and the development of electromobility in Europe. Our team is divided into four main groups: electrical, construction, composites and marketing, each of them focusing on their own tasks, which result in our boat "Baśka". The members of the project come from as many as ten faculties of the university.

The resources of ACC Cyfronet AGH have significantly influenced the development of the project. The construction section was able to simulate CFDs related to i.a. testing the resistance of boats and their individual components – year after year we try to reduce the total resistance coefficient of the structure through changes in the design and following the latest trends in world regatta classes. Other simulations concerned behavior during the load of gondola and propeller. The results collected during such simulations led us in the right direction and showed us where the critical stresses could be improved. However, due to the high complexity of these analyses, they required a lot of processing power that our workshop or private computers could not provide. Thanks to ACC Cyfronet AGH, which offers researchers supercomputers designed for complex calculations, the members of our team were able to use Ansys 18.2 and 19.0 software via the Prometheus supercomputer.

A set of hydrodynamic-focused tidal simulations of the fluid above the gondola and the hydro-wings was the exact subject of analyses. The number of commercially used hydrofoils is very large, so the choice of one ideal solution took many hours of computing. The complexity of the designed shape required the use of a very dense grid. This in turn forced a long time of calculations and the need to run them in parallel on all 24 cores available in a single Prometheus supercomputer node.

The obtained results were used to determine the parameters which were the subject of our greatest interest, i.e. the coefficients of water resistance and lift force. The second one is particularly important in the case of hydro-wings, on which the way our boat moves is based. Their task is to completely lift the hull so that only the wings are submerged. This process allows to significantly reduce the friction generated by the movement of the boat, but the effect can only be achieved with a sufficiently high coefficient of lift force. In addition, the level of elevation above water is related to the speed at which the boat is moving. The analyses has also allowed us to calculate the minimum speed needed to emerge the boat, but high enough to maintain it. In the case of solar-powered boats, any change to improve energy efficiency, such as this reduction in friction, is desirable. The more energy is saved, the greater the distance the boat can cover.

AGH Solar Boat "Baśka"

AGH Solar Boat team during Monaco Solar & Energy Boat Challenge

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

Cyfronet's main mission is to serve the scientific community and to actively support also individual researchers, by providing them with the IT resources of the Centre: starting from the computing power of supercomputers, through mass storage, to a super-fast computer network and useful IT tools. We are very pleased and proud that among the wide range of topics and research work carried out with the help of ACC Cyfronet AGH supercomputers, some have been appreciated and awarded by the Polish National Science Centre. The scientific successes of Cyfronet infrastructure's users are the best proof of its usefulness as a tool for conducting scientific and research work.

photo: Michał Niewdana, NSC

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.

Award of the National Science Centre for Adam Rycerz, PhD DSc from the Institute of Physics of the Jagiellonian University*


photo: Michał Niewdana, NSC

On October 10, 2017, Adam Rycerz won the Prize in the field of technical sciences. The prize was awarded for theoretical analysis of quantum transport of a charge in graphene nanosystems. The jury considered it particularly important that the laureate had proposed a mechanism for controlling current polarization in the space of the valley degrees of freedom through electrostatic fields.

– *In the theory of condensed matter, I am particularly fascinated by the possibility of predicting – with relatively little effort and resources – completely new, previously unknown physical phenomena, the practical use of which is actually within reach* – says Adam Rycerz. – *Although the cases where the implementation of such a plan was successful, can be counted on the fingers of one hand*

(usually the experimental discovery preceded the laborious process of building the theory) abandoning such attempts will certainly mean the end of civilization.

Adam Rycerz supports his research with calculations on Cyfronet supercomputers within the CORRCO2 computational grant: "Macroscopic quantum effects and critical phenomena in Dirac materials".

Award of the National Science Centre for Joanna Sułkowska, PhD DSc from the Centre of New Technologies and the Faculty of Chemistry of the University of Warsaw*

On October 10, 2018, Joanna Sułkowska became a laureate of the Award in the field of life sciences for her work on "looped proteins". It is worth emphasizing that the improper operation of "looped proteins" can lead to various civilization diseases, and unraveling their secrets can bring a breakthrough in the treatment of Alzheimer's disease, Parkinson's disease, or obesity.


photo: Michał Niewdana, NSC

Joanna Sułkowska specializes in molecular and theoretical biophysics. Her research is interdisciplinary – she uses physical, mathematical and biological knowledge as well as the node theory.

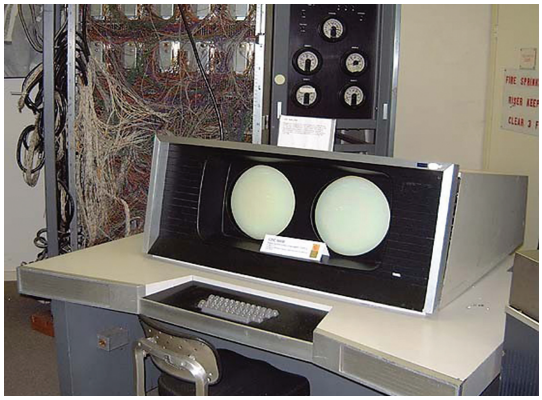
– *The very important thing in my work is the ability to look broadly, not to attach to a given field: analyzing the structure of proteins not only with the physicist's eye, but also taking into account the perspective of a biologist and chemist. You have to imagine things that cannot be seen with the naked eye, although they do exist* – says Joanna Sułkowska.


photo: Michał Niewdana, NSC

Joanna Sułkowska supports her research with calculations on the Prometheus supercomputer located at ACC Cyfronet AGH. The research is carried out within the TRMDMG3 computational grant: "Mechanism of tRNA methylation with the participation of narrowed proteins".

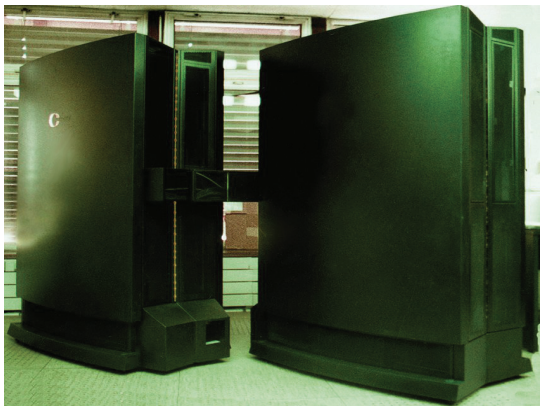
Congratulations to the winners and we wish you further success!

**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 81st 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 project has 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) on the TOP500 list.

Cyfronet exhibition stand at the ISC'19 conference

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

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

ISBN 978-83-61433-33-0