50 lat 1973/2023 CYFRONET

Akademickie Centrum Komputerowe

CYFRONET AGH

2023

ACK CYFRONET AGH jest jednostką wiodącą, ustanowioną przez Komitet Badań Naukowych, w zakresie eksploatacji i rozbudowy komputerów dużej mocy (KDM) oraz miejskiej akademickiej sieci komputerowej (MAN). CYFRONET jest koordynatorem Programu PLGrid oraz uznanym przez NCBiR Centrum Kompetencji w zakresie rozproszonych infrastruktur obliczeniowych typu gridowego i chmurowego.

Szanowni Państwo!

W roku 2023 Cyfronet obchodzi jubileusz 50-lecia swojej działalności. Przez te pięćdziesiąt lat służyliśmy polskiej nauce poprzez udostępnianie zasobów obliczeniowych i sieciowych oraz pomoc w ramach posiadanych kompetencji. Był to czas niezwykle wielkich przełomów technologicznych, które umożliwiły prowadzenie badań naukowych w coraz to bardziej zaawansowany sposób. Zmieniła się rola komputerów: z maszyn dostępnych tylko w wybranych specjalizowanych jednostkach stały się powszechnie dostępnymi i wszechobecnymi narzędziami informatycznymi. Również w Cyfronecie nastąpiło wiele niezwykle ważnych zmian. Jednak istota naszej działalności, nasza misja, którą spełniamy przez te wszystkie lata, pozostała taka sama. Wspieramy naukę, innowacyjną gospodarkę i społeczeństwo informacyjne dostarczając moc obliczeniową, systemy składowania danych, usługi sieciowe, oprogramowanie oraz pomagając i przekazując specjalistyczną wiedzę. Traktując potrzebę zmian jako naturalny ciąg zdarzeń nie tylko w nauce, stale pracujemy nad modyfikacją oferowanych usług poprzez stosowanie najnowszych rozwiązań technologicznych, a te w informatyce są ciągiem nieustających radykalnych zmian.

Lata 2021-2022 przyniosły kilka znaczących owoców naszych stałych starań o zapewnienie polskiej nauce jak najlepszych możliwości obliczeniowych. Uruchomiliśmy dwa nowe superkomputery: Aresa i Athenę, która stała się najszybszym superkomputerem w Polsce. Konfiguracja Atheny obejmuje serwery z procesorami AMD EPYC oraz karty NVIDIA z procesorami GPGPU A100. Dzięki tym zasobom teoretyczna moc obliczeniowa Atheny wynosi 7,7 PetaFlopsa, a na potrzeby obliczeń AI to moc obliczeniowa niemalże 240 PetaFlopsów!

W czerwcu 2022 roku po raz pierwszy w historii na liście TOP500 najszybszych superkomputerów znalazły się jednocześnie 3 maszyny z jednego polskiego centrum obliczeniowego. Są to działające w Cyfronecie: Athena (miejsce 105), Ares (290) i Prometheus (475). Superkomputery z Cyfronetu zajęły również wysokie miejsca na liście Green500 najbardziej efektywnych energetycznie superkomputerów – na czele z Atheną na prestiżowym 9. miejscu. Jednocześnie wspólnie z partnerami tworzącymi kierowane przez Finlandię konsorcjum LUMI świętowaliśmy uzyskanie na liście TOP500 przez superkomputer LUMI trzeciego miejsca na świecie i pierwszego w Europie. To niezwykle ważne osiągnięcie, bowiem polscy naukowcy uzyskali możliwość wykonywania obliczeń na superkomputerze z „najwyższej światowej półki". Dostęp do LUMI odbywa się za pośrednictwem portalu programu PL-Grid, który koordynuje Cyfronet.

Dodatkowo, Europejskie Wspólne Przedsięwzięcie w Dziedzinie Obliczeń Wielkiej Skali (EuroHPC JU) wytypowało Polskę jako miejsce na instalację systemu superkomputerowego nowej generacji, jako części pan-europejskiej infrastruktury przetwarzania danych. Superkomputer w skali *mid-range* o mocy przynajmniej kilkukrotnie przewyższającej moc Atheny zostanie zainstalowany już pod koniec 2023 roku. Operatorem będzie ACK Cyfronet AGH, co potwierdza uznanie naszych kompetencji w zakresie HPC (High Performance Computing) na arenie międzynarodowej.

Jakie wyzwania staną przed światem w ciągu kolejnych 50 lat? Część z nich, jak sprawy energetyki, związane z ochroną środowiska i zagrożeniami antropogenicznymi, możemy już zarysować obecnie. Inne pojawią się niespodziewanie. Jednak doświadczenie półwiecza naszej działalności pokazuje, że bez względu na panujące wokół warunki, kluczowa jest współpraca oparta na zaufaniu i wzajemnym zrozumieniu. Dlatego w imieniu Cyfronetu dziękuję Państwu za ostatnie 50 lat owocnej współpracy i zapraszam do dalszych wspólnie podejmowanych działań!

Z wyrazami szacunku

Prof. Kazimierz Wiatr
Dyrektor ACK Cyfronet AGH

DYREKTOR
prof. dr hab. inż. Kazimierz Wiatr
Tel. +48 12 633 34 26
e-mail: Kazimierz.Wiatr@cyfronet.pl

Zastępca Dyrektora
ds. Infrastruktury Informatycznej
mgr inż. Karol Krawentek
Tel. +48 12 633 34 26
e-mail: Karol.Krawentek@cyfronet.pl

Zastępca Dyrektora
ds. Komputerów Dużej Mocy
mgr Marek Magryś
Tel. +48 12 633 34 26
e-mail: Marek.Magrys@cyfronet.pl

Zastępca Dyrektora
ds. Administracyjnych
mgr Agnieszka Szymańska
Tel. +48 12 633 34 26
e-mail: Agnieszka.Szymanska@cyfronet.pl

Zastępca Dyrektora
ds. Ekonomiczno-Finansowych
mgr Angelika Zaleska-Walterbach
Tel. +48 12 633 80 53
e-mail: Angelika.Zaleska@cyfronet.pl

Dział Sieci Komputerowych – mgr Krzysztof Gaweł (kierownik)
Tel. +48 12 634 10 25, e-mail: Krzysztof.Gawel@cyfronet.pl

Dział Komputerów Dużej Mocy – mgr inż. Patryk Lasoń (kierownik)
Tel. +48 12 632 33 55, e-mail: Patryk.Lason@cyfronet.pl

Dział Bezpieczeństwa Danych – dr inż. Grzegorz Sułkowski (kierownik)
Tel. +48 12 632 33 55, e-mail: Grzegorz.Sulkowski@cyfronet.pl

Dział Pamięci Masowych – mgr inż. Adrian Marszalik (kierownik)
Tel. +48 12 632 33 55, e-mail: Adrian.Marszalik@cyfronet.pl

Dział Oprogramowania HPC – mgr Łukasz Flis (kierownik)
Tel. +48 12 632 33 55, e-mail: Lukasz.Flis@cyfronet.pl

Dział Centrum Operacyjne – dr Andrzej Zemła (kierownik)
Tel. +48 12 632 33 55, e-mail: Andrzej.Zemla@cyfronet.pl

Dział Użytkowników – mgr inż. Robert Pająk (kierownik)
Tel. +48 12 632 33 55, e-mail: Robert.Pajak@cyfronet.pl

Dział Techniczny Nawojki – mgr inż. Damian Trela (kierownik)
Tel. +48 12 632 33 55, e-mail: Damian.Trela@cyfronet.pl

Dział Techniczny Podole – mgr inż. Mariusz Kula (kierownik)
Tel. +48 12 632 33 55, e-mail: Mariusz.Kula@cyfronet.pl

Dział Administracyjny – mgr Agnieszka Szymańska (kierownik)
Tel. +48 12 633 34 26, e-mail: Agnieszka.Szymanska@cyfronet.pl

Dział Ekonomiczno-Finansowy – mgr Angelika Zaleska-Walterbach (kierownik)
Tel. +48 12 633 80 53, e-mail: Angelika.Zaleska@cyfronet.pl

Dział Obsługi Projektów – mgr Aleksandra Pałuk (kierownik)
Tel. +48 12 632 33 55, e-mail: Aleksandra.Paluk@cyfronet.pl

Laboratorium Programu PLGrid – prof. Jacek Kitowski (kierownik)
Tel. +48 12 633 34 26, e-mail: kito@agh.edu.pl

Laboratorium Akceleracji Obliczeń i Sztucznej Inteligencji
dr hab. inż. Paweł Russek, prof. uczelni (kierownik)
Tel. +48 12 633 34 26, e-mail: Pawel.Russek@cyfronet.pl

Laboratorium Technologii Chmurowych – dr Łukasz Dutka (kierownik)
Tel. +48 12 633 34 26, e-mail: Lukasz.Dutka@cyfronet.pl

Laboratorium Przetwarzania Danych – inż. Roksana Wilk (kierownik)
Tel. +48 12 632 33 55, e-mail: Roksana.Wilk@cyfronet.pl

Laboratorium Interdyscyplinarnych Obliczeń Naukowych – mgr inż. Joanna Kocot (kierownik)
Tel. +48 12 632 33 55, e-mail: Joanna.Kocot@cyfronet.pl

Laboratorium Obliczeń Kwantowych – dr Mariusz Sterzel (kierownik)
Tel. +48 12 632 33 55, e-mail: Mariusz.Sterzel@cyfronet.pl

Laboratorium Metod Informatycznych w Medycynie – dr inż. Marian Bubak (kierownik)
Tel. +48 12 633 34 26, e-mail: bubak@agh.edu.pl

Laboratorium Algorytmów Równoległych – prof. Bogusław Cyganek (kierownik)
Tel. +48 12 632 33 55, e-mail: cyganek@agh.edu.pl

Laboratorium Zastosowań Technik Obliczeniowych – dr hab. inż. Łukasz Rauch, prof. uczelni (kierownik)
Tel. +48 12 632 33 55, e-mail: lrauch@agh.edu.pl

Laboratorium Technik Wizualnych – mgr Jacek Przybylski (kierownik)
Tel. +48 12 632 33 55, e-mail: Jacek.Przybylski@cyfronet.pl

ATHENA moc dla polskiej nauki

ATHENA

Cyfronet a zadania ery sztucznej inteligencji i komputerów kwantowych
ATHENA

Kadencja 2021-2024

prof. dr hab. Marek Jarnicki *Przewodniczący*	Uniwersytet Jagielloński
prof. dr hab. Tadeusz Lesiak *Zastępca Przewodniczącego*	Instytut Fizyki Jądrowej PAN
mgr Jakub Badowski	Instytut Nafty i Gazu Państwowy Instytut Badawczy
prof. dr hab. inż. Marek Gorgoń	Akademia Górniczo-Hutnicza
mgr inż. Jarosław Górski	Akademia Wychowania Fizycznego
dr Robert Gryboś	Instytut Katalizy i Fizykochemii Powierzchni PAN
prof. dr hab. Tomasz Hachaj	Uniwersytet Pedagogiczny
prof. dr hab. inż. Marek Kisiel-Dorohinicki	Akademia Górniczo-Hutnicza
mgr inż. Mariusz Kwinta-Pudełko	Uniwersytet Rolniczy
mgr inż. Wojciech Majka	Uniwersytet Jagielloński Collegium Medicum
mgr inż. Tomasz Masier	Instytut Ekspertyz Sądowych
ks. dr Roman Mazur SDB	Uniwersytet Papieski Jana Pawła II
mgr Paweł Mikłaszewski	Akademia Muzyczna Akademia Sztuk Pięknych Akademia Sztuk Teatralnych
dr Wojciech Palacz	Uniwersytet Jagielloński
prof. dr hab. Dariusz Put	Uniwersytet Ekonomiczny
mgr Elżbieta Tomczyk	Sieć Badawcza Łukasiewicz Krakowski Instytut Technologiczny
mgr inż. Dariusz Żelasko	Politechnika Krakowska
Z-ca Dyrektora Jacek Barski	Urząd Marszałkowski Województwa Małopolskiego
Dyrektor Paweł Schmidt	Urząd Miasta Krakowa

CYFRONET AGH

Athena – mocne wsparcie dla obliczeń naukowych

Athena osiąga teoretyczną moc obliczeniową ponad **7,7 PetaFlopsów** (7709 TeraFlopsów), co zapewniło maszynie zajęcie 105 miejsca w czerwcu 2022 r. na liście TOP500 oraz sprawia, że jest to obecnie najszybszy superkomputer w Polsce. Zainstalowany w 2021 roku w Cyfronecie system ma dostarczyć polskiemu środowisku naukowemu i gospodarce najnowocześniejszych zasobów obliczeniowych opartych o procesory i akceleratory GPGPU najnowszej generacji wraz z niezbędnym podsystemem składowania danych opartym o bardzo szybkie pamięci flash. Konfiguracja Atheny obejmuje: 48 serwerów z procesorami AMD EPYC i 1 TB pamięci RAM (w sumie 6144 rdzenie obliczeniowe CPU) oraz 384 karty GPGPU NVIDIA A100.

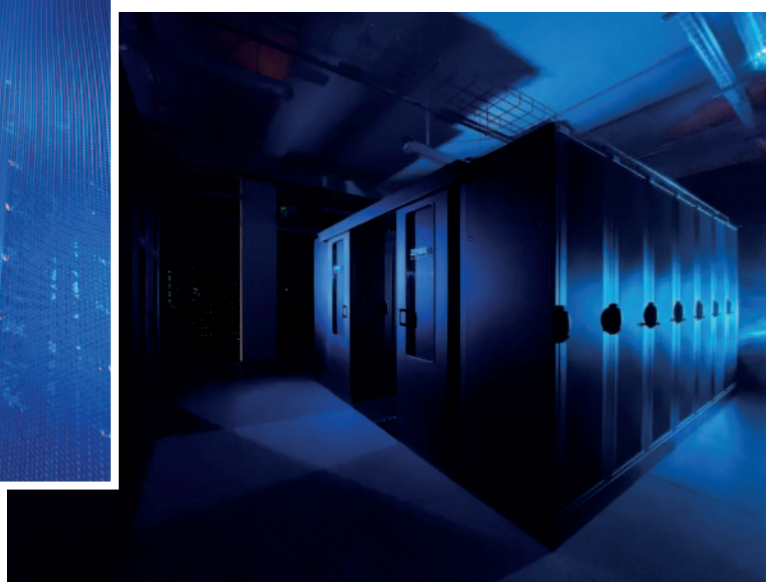
Athena w liczbach	
Liczba rdzeni obliczeniowych	6 144
Liczba kart z procesorami graficznymi GPGPU	384
Moc obliczeniowa	7,7 PFlops
Ranking TOP500 najszybszych komputerów świata (czerwiec 2022)	105 lokata

Niezbędnym elementem umożliwiającym wykorzystanie tak dużej mocy obliczeniowej w efektywny sposób jest zapewnienie wysokowydajnej sieci wewnętrznej superkomputera (Infiniband HDR o przepustowości 4 x 200 Gb/s na serwer) oraz bardzo szybkiego podsystemu dyskowego. Jest on zbudowany w oparciu o otwarte oprogramowanie Lustre, używane również w systemach dyskowych superkomputerów Ares i Prometheus, oraz dedykowane serwery dyskowe wyposażone w pamięci flash w standardzie NVMe. System został zainstalowany w istniejącym centrum danych Cyfronetu oraz zintegrowany z infrastrukturą PLGrid.

Tego typu infrastruktura wychodzi naprzeciw potrzebom użytkowników superkomputerów Cyfronetu, którzy wykorzystują infrastrukturę obliczeniową zarówno do wykonywania standardowych wysokowydajnych symulacji naukowych (HPC), jak i do aplikowania metod sztucznej inteligencji (AI) i uczenia maszynowego (ML) w badaniach z zakresu medycyny, farmakologii, biologii, chemii, fizyki oraz wielu innych dziedzin nauki. **Moc obliczeniowa Atheny dla obliczeń AI to prawie 240 PetaFlopsów!**

Spodziewanym efektem dostarczenia specjalizowanych zasobów obliczeniowych Atheny będzie poszerzenie zakresu prowadzonych prac badawczych, możliwość podejmowania zaawansowanych symulacji i analiz oraz zwiększenie możliwości przetwarzania spływających w sposób ciągły danych z laboratoriów z całego świata. Bezpośrednim spodziewanym efektem prowadzonych prac będą artykuły i opracowania naukowe, patenty, a w dalszej perspektywie nowatorskie rozwiązania, które mogą być podstawą dla powstawania nowych rozwiązań w gospodarce.

ATHENA
ATHENA
ZPAS
ZPAS
AGH



Ares – w kierunku krótszego czasu obliczeń

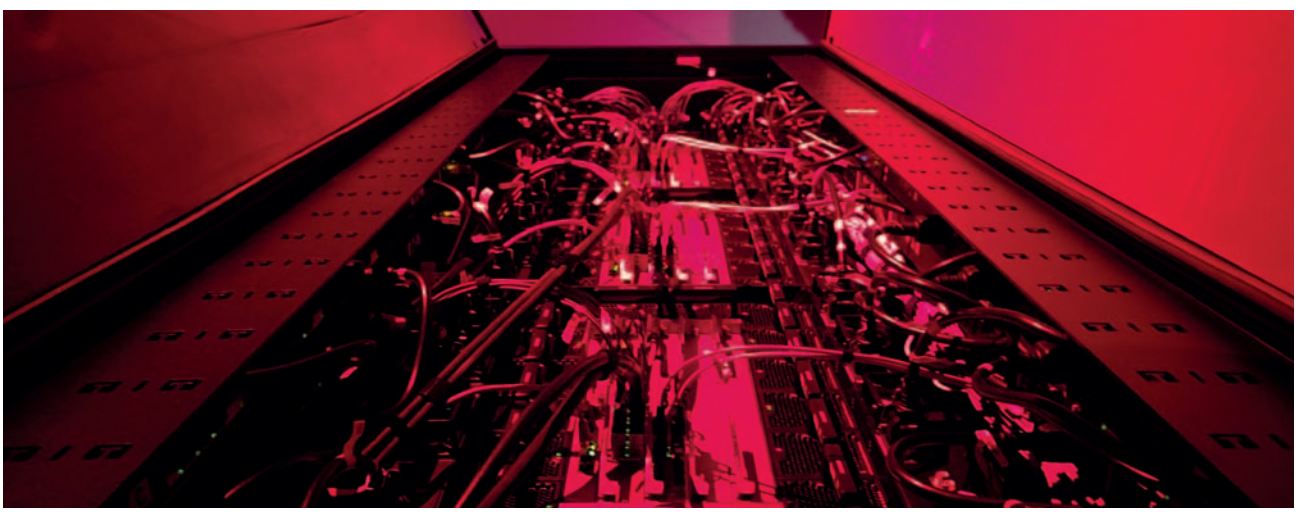
W roku 2021 został uruchomiony w Cyfronecie superkomputer Ares. Jest on zbudowany z serwerów obliczeniowych z procesorami firmy Intel (model Xeon Platinum), podzielonych na trzy grupy:

- 532 serwery wyposażone w 192 GB pamięci RAM,
- 256 serwerów wyposażonych w 384 GB pamięci RAM,
- 9 serwerów, każdy posiadający 8 kart NVIDIA Tesla V100.

Ares w liczbach	
Liczba rdzeni obliczeniowych	37 824
Pamięć operacyjna	147,7 TB
Liczba kart z procesorami graficznymi GPGPU	72
Moc obliczeniowa	4 PFlops
Ranking TOP500 najszybszych komputerów świata (czerwiec 2022)	290 lokata

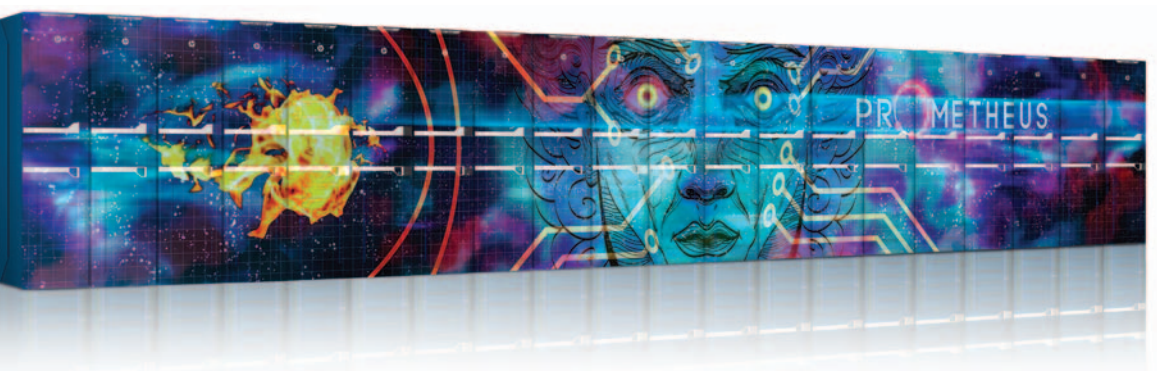
Sumaryczna, teoretyczna wydajność części CPU to ponad 3,5 PFlops, a części GPU to ponad 500 TFlops. Aresa wspomaga również system dyskowy o pojemności ponad 11 PB. Do przesyłania danych używana jest sieć typu InfiniBand EDR. Superkomputer posiada 37 824 rdzenie obliczeniowe oraz 147,7 TB pamięci RAM. Został też wyposażony w system chłodzenia cieczą.

Ares stanowi uzupełnienie zasobów obliczeniowych Cyfronetu poprzez dostarczenie nowszej generacji procesorów i serwerów z większą ilością pamięci. Umożliwi to skrócenie czasu obliczeń zadań naukowych oraz zaadresowanie problemów, które do tej pory nie mogły być uruchamiane w dużej skali ze względu na niewystarczającą ilość pamięci. Dodatkowo, umiejscowienie Aresa w innej geograficznie lokalizacji niż Prometheus gwarantuje zachowanie ciągłości świadczenia usług obliczeniowych.



Prometheus – PetaFlops'owa moc

Bardzo szybko następujące zmiany w świecie nauki wpływają na tempo rozwoju zaplecza informatycznego, oferowanego naukowcom przez Cyfronet. Rosnące coraz wyraźniej zapotrzebowanie środowisk naukowych na moc obliczeniową i coraz większe magazyny danych jest wyraźnie zauważalne ze strony dziedzin niemalże tradycyjnie związanych z Komputerami Dużej Mocy: chemii, fizyki, astronomii, szeroko rozumianych nauk o życiu oraz dziedzinach im pokrewnych. Astronomia, astrofizyka czy fizyka przestrzeni kosmicznej z jednej strony opierają się na akwizycji i analizie danych, a z drugiej na złożonych symulacjach komputerowych. Nauki biologiczne, chemiczne, medyczne, jak i wyżej wymienione, cechują się dynamicznym rozwojem oraz wprowadzaniem nowych, coraz bardziej wyrafinowanych metod badawczych, np. technik molekularnych opierających się na wysokowydajnym sekwencjonowaniu DNA. Multidyscyplinarna dziedzina, jaką jest medycyna, mierzy się z wieloma czasochłonnymi analizami, m.in. genomu ludzkiego. W efekcie wzrasta zapotrzebowanie na automatyczne zbieranie, przechowywanie i analizowanie sygnałów oraz obrazów biomedycznych, a do realizacji tych procesów niezbędne stają się zasoby superkomputerowe. Możliwość powiązania ze sobą wielu unikatowych danych, to jest danych klinicznych, genetycznych oraz środowiskowych i społecznych, niesie ze sobą wiele korzyści, ale i tu konieczne są dedykowane usługi, które mogą zaoferować jedynie centra superkomputerowe.



To właśnie z tego typu zadaniami mierzy się Prometheus – najszybszy polski superkomputer. Jako następca Zeusa, został włączony do infrastruktury PLGrid i nieodpłatnie służy naukowcom, także w ramach międzynarodowych projektów badawczych. Wykorzystanie superkomputera obejmuje: analizę wyników badań, symulacje numeryczne, przetwarzanie danych z sieci, zaawansowane wizualizacje oraz równoległą analizę ogromnych zbiorów danych (*big data*).

Prometheus składa się z ponad 2239 serwerów platformy HP Apollo 8000, połączonych superszybką siecią InfiniBand FDR o przepustowości 56 Gbit/s. Superkomputer posiada 53 748 rdzeni obliczeniowych (procesorów Intel Haswell oraz Intel Skylake) oraz 283,5 TB pamięci operacyjnej w technologii DDR4. Został wyposażony w dwa systemy plików o łącznej pojemności 10 PB oraz bardzo dużej szybkości dostępu wynoszącej 180 GB/s. Prometheus jest również wyposażony w 144 karty NVIDIA z procesorami graficznymi GPGPU Tesla K40 XL oraz w 32 karty NVIDIA Tesla V100. Teoretyczna moc obliczeniowa superkomputera to 2,7 PFlops (PetaFlops)!

Co ważne, dzięki innowacyjnej technologii bezpośredniego chłodzenia cieczą procesorów i modułów pamięci operacyjnej, Prometheus zaliczany jest do najbardziej energooszczędnych systemów obliczeniowych tej klasy na świecie. Zostało to osiągnięte dzięki wykorzystaniu do chłodzenia mikroprocesorów cieczy o temperaturze 28°C, do wychłodzenia której w naszym klimacie wystarczą tzw. *dry-coolery,* zamiast konsumujących duże ilości energii elektrycznej generatorów wody lodowej oraz klasycznych klimatyzatorów technologicznych. Dzięki zastosowaniu chłodzenia cieczą, elementy elektroniczne pracują w niższych niż zwykle

temperaturach, co pozytywnie wpływa nie tylko na ich mniejszą awaryjność, ale także pozwala uzyskać wydajność o ponad 5% większą niż dla analogicznej instalacji opartej na klasycznym chłodzeniu powietrzem. Ponadto, chłodzenie cieczą umożliwia osiągnięcie ekstremalnie wysokiej gęstości instalacji – 144 serwerów obliczeniowych, dzięki czemu ważąca ponad czterdzieści ton część obliczeniowa Prometheusa mieści się w zaledwie dwudziestu szafach. Ma to istotny wpływ także na wewnętrzną transmisję danych, bowiem odległości połączeń są tu krytyczne.

Prometheus w liczbach	
Liczba rdzeni obliczeniowych	53 748
Pamięć operacyjna	283,5 TB
Liczba kart z procesorami graficznymi GPGPU	144
Moc obliczeniowa	2,7 PFlops
Ranking TOP500 najszybszych komputerów świata (czerwiec 2022)	475 lokata

Cały system obliczeniowy, wraz z niezbędnymi elementami towarzyszącymi, w tym m.in. systemem gwarantowanego zasilania z dodatkowym agregatem prądotwórczym oraz nowoczesnymi systemami klimatyzacji technologicznej i gaszenia gazem, został zainstalowany w nowoczesnym Centrum Przetwarzania Danych (Data Center) Cyfronetu, specjalnie przystosowanym do eksploatacji Prometheusa.

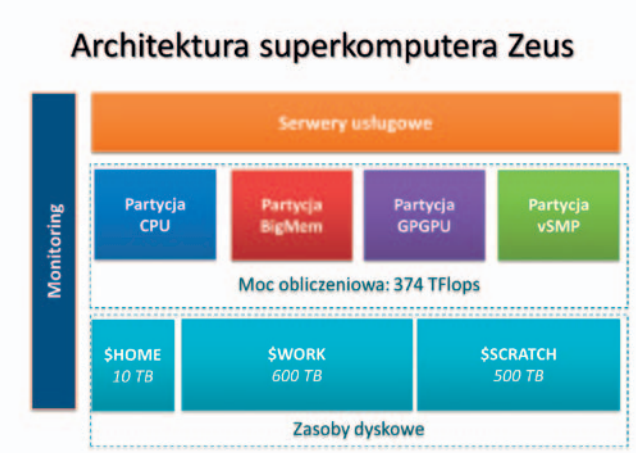
Prometheus po raz kolejny znalazł się w zestawieniu 500 najszybszych komputerów świata (lista TOP500, edycja z czerwca 2022 r.) zajmując **475 miejsce, jako jeden z pięciu superkomputerów z Polski**.

Ponieważ zastosowany w superkomputerze Zeus podział na partycje o różnej funkcjonalności został bardzo dobrze wykorzystany przez użytkowników, architektura superkomputera Prometheus składa się również z kilku klas węzłów o zróżnicowanym przeznaczeniu:

- klasycznej partycji ogólnego przeznaczenia z wydajnymi dwuprocesorowymi węzłami obliczeniowymi (Intel Xeon E5-2680v3 Haswell i Intel Xeon Gold 6128),
- zestawu serwerów wyposażonych w akceleratory graficzne GPGPU NVIDIA Tesla K40 XL,
- partycji akceleracyjnej wyposażonej w zestaw urządzeń uzupełniających konfigurację Prometheusa o kilka rodzajów akceleratorów (m.in.: karty GPGPU – NVIDIA K80, karty Intel Xeon Phi 7120P oraz karty Nallatech z układami FPGA Altera Stratix V: P395-AB2-330A-30 i P395-D8-330A-30),
- partycji dedykowanej dla obliczeń związanych ze sztuczną inteligencją, wyposażonej w akceleratory graficzne GPGPU NVIDIA Tesla V100. Warto wspomnieć, że **partycja ta stanowi system o mocy obliczeniowej ponad 4 PFlops dla operacji tensorowych i 256 TFlops dla standardowych obliczeń wykonywanych na liczbach podwójnej precyzji, co czyni go najszybszym dedykowanym rozwiązaniem dla sztucznej inteligencji dostępnym dla potrzeb nauki w Polsce**.

Dzięki uruchomieniu superkomputera Prometheus użytkownicy otrzymali ponad siedmiokrotnie większe możliwości w porównaniu z wcześniej wykorzystywanym Zeusem. Znacznie wydajniejsze procesory, szybsza sieć połączeń wewnętrznych oraz większa ilość pamięci operacyjnej pozwalają realizować obliczenia w skali niemożliwej do osiągnięcia na wcześniejszych zasobach Cyfronetu.

Zeus – ponad 60 000 lat obliczeń



Nieprzerwanie od roku 2008 w Cyfronecie działa superkomputer Zeus – długoletni lider wydajnościowy rankingów najszybszych systemów obliczeniowych w Polsce. Superkomputer ten posiada aktualnie 25 468 rdzeni obliczeniowych o łącznej teoretycznej mocy obliczeniowej 374 TFlops, 60 TB pamięci operacyjnej RAM oraz pamięć dyskową o pojemności 2,3 PB. Wyposażony dodatkowo w ponad 200 kart GPGPU, stanowi bardzo duże wsparcie dla obliczeń wykonywanych przez społeczności naukowe.

Od momentu instalacji Zeus nieprzerwanie notowany był na liście **TOP500**, najszybszych superkomputerów świata (12 razy), zajmując bardzo wysokie miejsca: cztery razy w pierwszej setce, a najwyżej – **na 80 pozycji**. Zeus był 10-krotnie najszybszy w Polsce.

Architektura

Zeus jest heterogenicznym superkomputerem obliczeniowym. Składa się z czterech klas węzłów, zróżnicowanych pod względem architektury zasobów obliczeniowych, specjalnie dostosowanych do wymagań środowisk naukowych:

- **klasycznego klastra** serwerów obliczeniowych, z wysoko wydajnymi węzłami CPU wyposażonymi w dwa procesory Intel Xeon i 16-24 GB pamięci na węzeł,
- klastra serwerów z dużą ilością pamięci – złożonego z węzłów z czterema procesorami AMD Opteron i 256 GB pamięci na węzeł,
- zestawu serwerów wyposażonych w akceleratory **GPGPU** (procesory Intel Xeon i karty NVIDIA M2050 oraz karty NVIDIA M2090) i **FPGA** (moduły Pico Computing M-503 z układami Xilinx Virtex-6 LX240T),
- **„wirtualnego" komputera SMP** z dużą ilością pamięci współdzielonej, wykorzystującego dodatkowe oprogramowanie vSMP firmy ScaleMP, które pozwala na uruchamianie maszyn posiadających do 768 rdzeni i 6 TB pamięci RAM.

Zróżnicowanie rodzajów węzłów umożliwia dopasowanie aplikacji użytkowników do sprzętu, który jest najlepiej dobrany do cech i specjalnych wymagań tych aplikacji. Na przykład, grupa

klasycznych węzłów CPU jest najczęściej wykorzystywana do realizacji szeregowych i równoległych (MPI) zadań obliczeniowych, podczas gdy druga grupa („grube" węzły) jest idealna dla zadań wymagających dużej ilości pamięci operacyjnej. Węzły GPU pozwalają niektórym aplikacjom na korzystanie z akceleratorów GPGPU, a z kolei węzły vSMP umożliwiają uruchamianie zadań wymagających ogromnej pamięci lub pozwalają na skalowanie aplikacji, które nie korzystają

z żadnej biblioteki komunikacji między węzłami, takiej jak MPI, w celu zrównoleglenia obliczeń.

Użytkownicy

Od początku swego istnienia, superkomputer Zeus służył całemu polskiemu środowisku naukowemu. W 2021 roku, tylko sam Zeus wykonał ponad **1,6 miliona zadań** dla naukowców z wielu uczelni i instytutów badawczych. Większość z nich realizowanych było na więcej niż jednym procesorze, a typowe obliczenia użytkownika żądają nawet **kilku tysięcy rdzeni dla jednego zadania!**

3 superkomputery Cyfronetu na liście TOP500 superkomputerów o najwyższej mocy obliczeniowej

Superkomputery z Cyfronetu na liście TOP500

Superkomputer Zeus

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

Superkomputery Zeus i Prometheus

- 2015 - VII, 269 i 49 miejsce, 374 + 1659 TF
- 2015 - XI, 387 i **38** miejsce, 374 + 2399 TF

Superkomputer Prometheus

- 2016 - VI, 49 miejsce, 2399 TF
- 2016 - XI, 60 miejsce, 2399 TF
- 2017 - VI, 72 miejsce, 2399 TF
- 2017 - XI, 78 miejsce, 2399 TF
- 2018 - VI, 103 miejsce, 2399 TF
- 2018 - XI, 131 miejsce, 2399 TF
- 2019 - VI, 174 miejsce, 2399 TF
- 2019 - XI, 241 miejsce, 2399 TF
- 2020 - VI, 288 miejsce, 2399 TF
- 2020 - XI, 324 miejsce, 2399 TF

Superkomputery Prometheus i Ares

- 2021 - VI, 373 i **216** miejsce, 2399 + 3510 TF
- 2021 - XI, 440 i 267 miejsce, 2399 + 3510 TF

Superkomputery Prometheus, Ares i Athena

- 2022 - VI, 475, 290 i **105** miejsce, 2399 + 3510 + 7709 TF

TOP 500 CERTIFICATE
The List.
Athena - FormatServer THOR ERG21, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100 SXM4 40 GB, Infiniband HDR
Cyfronet, Poland
is ranked
among the World's TOP500 Supercomputers
in the 59th TOP500 List published at the ISC22
Conference on June 01, 2022.
Congratulations from the TOP500 Editors
Erich Strohmaier NERSC/Berkeley Lab
Jack Dongarra University of Tennessee
Horst Simon NERSC/Berkeley Lab
Martin Meuer Prometeus

30 maja 2022 roku, w czasie konferencji ISC High Performance w Hamburgu ogłoszono najnowszą listę TOP500 superkomputerów o największej mocy obliczeniowej na świecie. **Po raz pierwszy w historii na liście znalazły się jednocześnie 3 superkomputery z jednego polskiego centrum obliczeniowego. Są to działające w Cyfronecie: Athena (105 lokata), Ares (290) i Prometheus (475).**

Historia obecności zainstalowanych w Cyfronecie maszyn na liście TOP500 sięga roku 1996, kiedy to komputer SPP1200/XA-32 zajął 408 miejsce. Po dłuższym okresie bez notowań, w 2010 roku na listę trafił Zeus i pozostał tam aż do 2015 roku. Od tamtej pory Cyfronet jest w zestawieniu reprezentowany przez Prometheusa, do którego w czerwcu 2021 dołączył Ares, a w 2022 roku Athena.

Obok zaprezentowano zestawienie miejsc, jakie superkomputery Zeus, Prometheus, Ares i Athena zajmowały na liście TOP500 w ciągu ostatnich lat.

Obecność superkomputerów Cyfronetu na liście TOP500 jest ważnym potwierdzeniem kompetencji w zakresie: dostępności zasobów, jakości usług, zaufania użytkowników, niezawodnej pracy i większej dostępności do programów UE.

Czołówka światowej wydajności energetycznej

3 superkomputery Cyfronetu znalazły się również na liście Green500 – superkomputerów najbardziej ekologicznych, czyli o najwyższej wydajności energetycznej, obliczanej jako współczynnik ilości operacji zmiennoprzecinkowych na sekundę (mocy obliczeniowej superkomputera) do poboru energii: Gflops/W. Szczególnym sukcesem jest zajęcie przez Athenę 9. miejsca na liście Green500 najbardziej efektywnych energetycznie superkomputerów świata. Ta pozycja świadczy o doskonałym stosunku udostępnianej mocy obliczeniowej do poboru energii elektrycznej. Na liście Green500 nie zabrakło też Aresa i Prometheusa. Ares znalazł się w pierwszej setce, na 83 miejscu, natomiast Prometheus zajął 162 miejsce.

The GREEN 500 CERTIFICATE

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

Cyfronet, Poland

is ranked

among the World's TOP500 Supercomputers

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

Congratulations from the Green500 Editors

Wu-chun Feng Virginia Tech

Kirk Cameron Virginia Tech

Wykorzystanie superkomputerów

Krakowskie superkomputery są częścią europejskiej chmurowej i gridowej infrastruktury w ramach Europejskiej Infrastruktury Gridowej (EGI). Jednocześnie, są również ważnymi superkomputerami w ogólnopolskiej infrastrukturze obliczeniowej PLGrid – platformie umożliwiającej prowadzenie badań naukowych *in silico* i wykonywanie obliczeń na komputerach dużej mocy, także w architekturze chmurowej i gridowej.

Dzięki infrastrukturze PLGrid możliwe jest intensywne wykorzystanie mocy obliczeniowych superkomputerów przez naukowców. Dedykowane środowiska obliczeniowe, tzw. gridy dziedzinowe, oraz specjalistyczne platformy informatyczne umożliwiają efektywną realizację coraz bardziej złożonych problemów obliczeniowych. Tematyka badań naukowych realizowanych przy pomocy superkomputerów ACK Cyfronet AGH jest bardzo bogata. Oto wybrane przykłady tematów obliczeń:

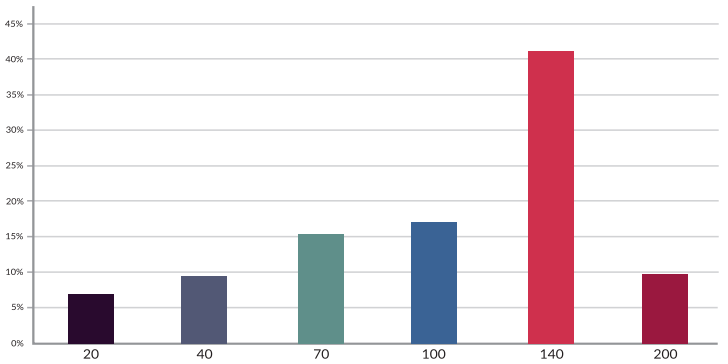
- planowanie wieloetapowych obliczeń z zastosowaniem akceleratorów kwantowych,
- badanie właściwości fizykochemicznych makrocząsteczek,
- modelowanie procesów polimeryzacji rodnikowych,
- badania nowych materiałów nadprzewodzących,
- analizy numeryczne procesu spalania w komorze hybrydowego silnika rakietowego,
- predykcja fal grawitacyjnych za pomocą uczenia maszynowego,
- procesy mikrofizyczne w plazmie kosmicznej,
- uczenie maszynowe w medycynie weterynaryjnej,
- projektowanie sond fluorescencyjnych wykorzystywanych w obrazowaniu medycznym.

Szeroki zakres tematów badawczych jest dowodem ciągle rosnącej liczby naukowców, którzy są świadomi korzyści płynących z wykorzystania superkomputerów. Z ich pomocą można uzyskać wyniki złożonych symulacji wiele, wiele razy szybciej niż na zwykłym komputerze stacjonarnym. Obliczenia, które przy użyciu pojedynczych komputerów często zajęłyby wiele lat (w konkretnych rzeczywistych przypadkach ponad 150, 700 czy nawet 1000 lat), tu mogą być wykonane najczęściej w ciągu zaledwie kilku dni. Co ważne, użytkownicy Cyfronetu mogą korzystać z profesjonalnego wsparcia – od pełnej dokumentacji, poprzez szkolenia, aż po indywidualne konsultacje z ekspertami.

Poza indywidualnymi naukowcami i małymi grupami badawczymi obliczenia w ramach wielu różnych dyscyplin naukowych prowadzą z pomocą superkomputerów nawet międzynarodowe konsorcja, oczywiście z udziałem polskich naukowców. Dzięki krakowskim superkomputerom polscy uczeni mogą uczestniczyć w ważnych międzynarodowych projektach naukowych, w tym w eksperymentach z listy ESFRI i PMIB: CTA, LOFAR, EPOS, Wielkiego Zderzacza Hadronów w CERNie oraz niedawno odkrytych fal grawitacyjnych w detektorach LIGO i VIRGO.

Rok	Liczba zadań	Czas CPU w latach
Superkomputer Zeus		
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
Superkomputery Zeus i Prometheus łącznie		
2015	7 505 763	15 952
2016	7 748 677	24 653
2017	9 066 892	39 232
2018	8 342 686	42 436
2019	4 993 639	44 027
2020	5 696 919	41 761
Superkomputery Zeus, Prometheus i Ares łącznie		
2021	5 549 582	43 409

Punktacja artykułów naukowych według MEiN, powstałych w 2021 roku dzięki zasobom ACK Cyfronet AGH



Oczywiście, nawet najwyższe lokaty w zestawieniu TOP500, czy też najnowocześniejsze technologie wykorzystane do budowy komputerów dużej mocy nie oddają w pełni znaczenia tego rodzaju zasobów obliczeniowych dla polskiego środowiska naukowego. O użyteczności superkomputerów udostępnianych przez ACK Cyfronet AGH jako narzędzia prowadzenia prac naukowo-badawczych, najlepiej świadczą dane statystyczne dotyczące ich wykorzystywania.

W tabeli przedstawione zostały zagregowane najważniejsze dane dotyczące liczby zadań obliczeniowych oraz czasu ich trwania, wykonanych przez Cyfronet dla innych jednostek.

Warto podkreślić, że ogromne zapotrzebowanie użytkowników na moc obliczeniową i przestrzeń do przechowywania danych nie zostałoby zaspokojone bez ciągłego powiększania zasobów obliczeniowych i pamięci dyskowej. Dlatego uważnie analizujemy sugestie użytkowników i dane statystyczne dotyczące prowadzonych obliczeń, nie zapominając jednocześnie o ciągłej obserwacji światowych trendów w informatyce.

Poziom naukowy zadań zrealizowanych z wykorzystaniem infrastruktury udostępnianej przez ACK Cyfronet AGH jest bardzo wysoki. Dowodem tego są wyniki prac naukowo-badawczych, prowadzonych w roku 2021 przy wykorzystaniu tej infrastruktury, które zostały zaprezentowane w wielu publikacjach.

Kompleksowa infrastruktura wydajnego i bezpiecznego składowania danych cyfrowych

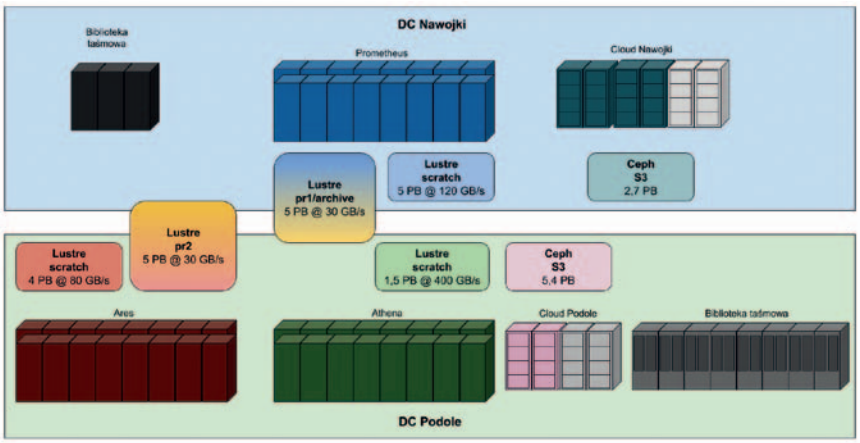
Obserwowane obecnie zjawisko lawinowo rosnącej ilości wytwarzanych informacji w postaci cyfrowej dotyczy także środowiska naukowego. Dostęp do bardzo wydajnych superkomputerów umożliwia podejmowanie analiz wielkoskalowych problemów badawczych, w efekcie czego powstają olbrzymie zbiory danych. Wymagają one całkowicie nowego podejścia do procesów przetwarzania i przechowywania informacji. Problem ten, stanowiący obecnie jedno z najważniejszych wyzwań współczesnego świata cyfrowego, określany jest pojęciem BigData. Również i w ACK Cyfronet AGH wyraźnie widoczna jest korelacja pomiędzy wzrostem oczekiwań w zakresie dostępnej pojemności, szybkości oraz dodatkowych funkcjonalności zasobów pamięci masowej, spowodowanym oferowaniem coraz wydajniejszych systemów obliczeniowych. Architektura Systemu Składowania Danych Cyfronetu – głównej platformy pamięci masowej dla Komputerów Dużej Mocy, złożona jest z kilku podstawowych elementów:

- sieci SAN – wydajnej i wysokodostępnej sieci przeznaczonej do komunikacji urządzeń wchodzących w skład Systemu Składowania Danych oraz klientów wykorzystujących udostępniane zasoby lub usługi,
- macierzy oraz serwerów dyskowych oferujących przestrzeń przeznaczoną na składowanie danych użytkowników o dużej rozpiętości parametrycznej – począwszy od szybkich lecz równocześnie drogich i mało pojemnych rozwiązań, a skończywszy na urządzeniach bardzo pojemnych i relatywnie tanich, ale zarazem o ograniczonej wydajności,
- serwerów usługowych wraz ze specjalistycznym oprogramowaniem usługowym i wirtualizacyjnym, udostępniających dla użytkowników funkcjonalności takie jak: automatyczne kopie zapasowe i archiwalne, hierarchiczne systemy składowania danych, wydajne sprzętowe platformy plików lub rozproszone sieciowe systemy plików,
- bibliotek taśmowych oraz specjalizowanego oprogramowania służącego do zapisywania krytycznych danych użytkowników na nośnikach magnetycznych,
- infrastruktury dodatkowej, w tym sieci Ethernet, Infiniband, rozwiązań wspomagających zarządzanie infrastrukturą IT oraz umożliwiających bezpieczne przechowywanie nośników magnetycznych.

Aktualnie całkowita fizyczna pojemność zasobów dyskowych i taśmowych Cyfronetu wynosi ponad **70 PB**.

Pamięć masowa dla superkomputerów

Odpowiednie skorelowanie projektu architektury infrastruktury obliczeniowej z właściwym doborem technologii rozwiązań składowania danych pozwala zoptymalizować usługi świadczone na rzecz użytkowników naukowych. Skala problemów w tym obszarze rośnie wraz ze złożonością i wydajnością eksploatowanych superkomputerów – obecnie systemy składowania danych dołączone do superkomputerów Cyfronetu przechowują miliardy plików o rozmiarach sięgających terabajtów. Szeroki zakres tematyczny prac badawczych realizowanych na zasobach udostępnianych przez Cyfronet znajduje swoje odbicie w różnorodności konfiguracji kluczowych superkomputerów Centrum, a przez to również w konfiguracji dedykowanych im zasobów pamięci masowej. Zasoby systemu składowania danych Cyfronetu znajdują się w dwóch lokalizacjach.



W superkomputerach stosuje się m.in. wydajną przestrzeń tymczasową, tzw. *scratch*. Tu kluczowym elementem jest szybkość działania, dlatego jest ona oparta na bardzo szybkiej architekturze rozproszonego systemu plików – Lustre. Zaletą Lustre jest możliwość skalowania pojemności jak i wydajności przestrzeni dyskowej. W wyniku łączenia pojemności wielu serwerów, przepustowość I/O jest również zagregowana i skaluje się z dodatkowymi serwerami. Ponadto, przepustowość i/lub pojemność może być łatwo zwiększona poprzez dynamiczne dodawanie większej liczby serwerów, bez konieczności przerywania obliczeń użytkowników. Obecnie w Cyfronecie wszystkie superkomputery korzystają z przestrzeni *scratch* realizowanych przez Lustre. W przypadku Prometheusa przestrzeń ta charakteryzuje się pojemnością 5 PB i szybkością 120 GB/s. Ares dysponuje przestrzenią o pojemności 4 PB i szybkością 80 GB/s. W obu tych komputerach przestrzeń scratch jest realizowana przy pomocy dysków mechanicznych. W przypadku Atheny dane użytkowników przechowywane są na dyskach półprzewodnikowych. Zastosowanie tego typu rozwiązania znacząca zwiększa wydajność systemu, planowana pojemność tego typu przestrzeni dla Atheny ma wynieść 1,5 PB oraz osiągać przepustowość 400 GB/s.

Większa część zasobów pamięci dyskowej Cyfronetu jest przeznaczona dla potrzeb użytkowników usług dziedzinowych rozwijanych w programie PLGrid. Infrastruktura PLGrid oferuje dedykowaną przestrzeń roboczą dla grup korzystających z usług dziedzinowych – funkcjonalność niezbędną do umożliwienia współpracy między naukowcami pracującymi w rozproszonych geograficznie lokalizacjach. Funkcjonalność ta jest realizowana za pomocą systemu plików Lustre. Maksymalna pojemność zasobu */pr1* w superkomputerze Prometheus to 5 PB, a sumaryczna szybkość operacji zapisu i odczytu sięga 30 GB/s. W przypadku Aresa i Atheny zasób */pr2* charakteryzuje się pojemnością 5 PB i szybkością 30 GB/s.

Dodatkowymi zasobami do przechowywania zasobów danych użytkowników oraz projektów w Cyfronecie jest system obiektowego przechowywania danych. Jest on oparty o oprogramowanie CEPH. Dane w tym systemie dostępne są poprzez protokół S3 oparty na REST API i znajdują się w unikalnych globalnie kontenerach (bucket), w których użytkownicy przechowują swoje dane w formie obiektów.

Szczególnym przypadkiem pamięci masowej są zasoby dla dużych projektów i kolaboracji międzynarodowych, w których uczestniczy Cyfronet, takich jak WLCG (Worldwide LHC Computing Grid), analizujący dane z detektora LHC w CERN czy CTA (Cherenkov Telescope Array), badający promieniowanie gamma za pomocą sieci radioteleskopów. Tego typu projekty wymagają ogromnych zasobów dyskowych, dostępnych najczęściej za pomocą nietypowych protokołów takich jak SRM, xroot czy GridFTP. Cyfronet dostarcza tego typu przestrzeń dyskową przy użyciu kilku instancji dedykowanego oprogramowania DPM (Disk Pool Manager) i za pomocą dedykowanych sieci, takich jak LHCone. Łączna pojemność systemów DPM w Cyfronecie przekracza 2 PB.

Aktualnie całkowita udostępniana pojemność dyskowa eksploatowana przez ACK Cyfronet AGH wynosi około **45 PB**.

Kopie zapasowe i archiwizacja

ACK Cyfronet AGH zapewnia swoim użytkownikom szerokie portfolio usług związanych z zabezpieczaniem informacji przechowywanej w postaci cyfrowej. Oprócz zaawansowanych rozwiązań technologicznych, takich jak: dedykowane systemom pamięci masowej sieci komunikacyjne, nowoczesne macierze dyskowe, czy też sprzętowe serwery plików, Centrum realizuje również usługi kopii zapasowych i archiwalnych, opartych o nośniki magnetyczne. Wbrew przewidywaniom mówiącym o nieuchronnym zmierzchu rozwiązań wykorzystujących składowanie danych na taśmach magnetycznych, technologia ta stale się rozwija, oferując w kolejnych generacjach nie tylko coraz większe pojemności nośników, lecz również zdecydowanie lepsze przepustowości oraz mechanizmy wspierające bezpieczeństwo i efektywność zapisu informacji (np. wbudowane w napędy taśmowe algorytmy szyfrujące i kompresujące dane).

Cyfronet dysponuje aktualnie dwoma bibliotekami taśmowymi posiadającymi ponad 5 tysięcy slotów na taśmy magnetyczne standardu LTO oraz 25 napędów generacji 6 i 7. Pojedynczy nośnik magnetyczny LTO-7 posiada pojemność fizyczną 6 TB i pozwala na zapis z prędkością sięgającą 300 MB/s, co teoretycznie umożliwia przechowywanie w bibliotekach taśmowych prawie 30 PB nieskompresowanych danych. Opisane zasoby wykorzystywane są do realizacji bieżących kopii zapasowych oraz archiwalnych istotnych zasobów informacyjnych użytkowników Centrum.

Kopia zapasowa wykonywana jest na aktywnych danych poprzez proces replikacji z lokalizacji źródłowej do wyodrębnionej i wyizolowanej lokalizacji docelowej. Procedura wykonania kopii zapasowej zapewnia spójność danych źródłowych i zapasowych, zarówno na poziomie pojedynczego pliku, jak i całego środowiska (co ma miejsce w przypadku złożonych systemów IT, takich jak serwery baz danych lub poczty oraz systemy wirtualne). Fizycznie, proces klonowania jest zwykle wykonywany przez kopiowanie danych źródłowych z dysku klienta na zasoby dyskowe / taśmowe docelowego serwera archiwizacyjnego, przy użyciu dedykowanego lub współdzielonego medium dostępowego, takiego jak Ethernet lub sieć SAN. Celem utworzenia archiwum jest natomiast zapewnienie bezpieczeństwa nieużywanych danych i zwolnienie zajmowanych zasobów pamięci masowej. W odróżnieniu od kopii zapasowej, archiwum jest tworzone tylko jeden raz, poprzez migrację danych z lokalizacji źródłowej do docelowej.

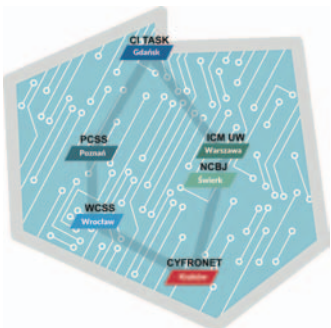
Cyfronet oferuje szeroki zakres usług do tworzenia kopii zapasowych, zarówno dostępnych bezpośrednio dla użytkowników, jak i działających automatycznie, bez ich ingerencji. Wśród tych do dyspozycji użytkowników, są usługi oparte na protokołach sieciowych FTP, NFS i SCP, działające w ramach dedykowanych serwerów kopii zapasowych. Udostępniają one rozwiązania służące do wykonywania kopii zapasowych, umożliwiając bezpośredni dostęp do zapasowych danych. W gestii użytkowników pozostaje natomiast decyzja, które dane mają pełnić rolę kopii zapasowych, a które archiwum.

W szczególnych przypadkach, Centrum oferuje usługę kopii zapasowych i archiwalnych, zwaną Systemem Powszechnej Archiwizacji. W ramach tej usługi, użytkownik otrzymuje dedykowaną przestrzeń dyskową, zabezpieczoną na wielu poziomach. Dane są składowane na macierzach dyskowych wyposażonych w mechanizmy nadmiarowości typu RAID 6, współpracujące z systemem hierarchicznego składowania danych HSM. W celu zapewnienia jeszcze wyższego poziomu bezpieczeństwa danych, kopie zapasowe danych użytkowników są dodatkowo chronione przez geograficzną replikację danych do jednostek stowarzyszonych. Użytkownicy Systemu Powszechnej Archiwizacji mogą także zwiększyć bezpieczeństwo swoich danych poprzez ich szyfrowanie przy pomocy certyfikatów. Obecnie, całkowita pojemność zasobów taśmowych w Cyfronecie przekracza **25 PB**.

Program PLGrid

Infrastruktura i projekty

W ciągu ostatnich kilkunastu lat Cyfronet aktywnie uczestniczył w wielu krajowych i ogólnoeuropejskich projektach z zakresu technologii gridowych i cloudowych. Uzyskane w ten sposób wiedza i doświadczenie, a także podejmowane w wielu krajach Europy inicjatywy gridowe, przyczyniły się do sformułowania założeń i rozpoczęcia Programu PLGrid. Początkiem tego Programu było podpisanie porozumienia powołującego Konsorcjum PLGrid w 2007 roku. Obecnie w skład Konsorcjum wchodzą następujące instytucje: Centrum Informatyczne Trójmiejskiej Akademickiej Sieci Komputerowej PG, Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego UW, Centrum Informatyczne Świerk NCBJ w Otwocku-Świerku, Poznańskie Centrum Superkomputerowo Sieciowe IChB PAN, Wrocławskie Centrum Sieciowo-Superkomputerowe PWr. oraz **Akademickie Centrum Komputerowe Cyfronet AGH jako inicjator i lider Programu i Konsorcjum PLGrid**. Praca podjęta przez pięciu partnerów Konsorcjum PLGrid zaowocowała zbudowaniem niezawodnej, działającej 24 godziny na dobę, przez cały rok, rozproszonej infrastruktury obliczeniowej dla nauki, w ramach której udostępniana jest użytkownikom wielka moc obliczeniowa, a także duże zasoby pamięci masowej oraz dedykowane narzędzia umożliwiające tworzenie zaawansowanych aplikacji naukowych.

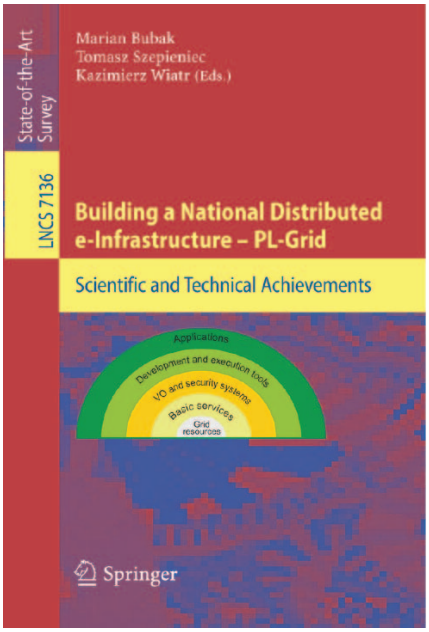


Proces budowy infrastruktury PLGrid rozpoczął się w trakcie projektu PL-Grid (Polska Infrastruktura Informatycznego Wspomagania Nauki w Europejskiej Przestrzeni Badawczej) i stanowił odpowiedź na rosnące potrzeby polskich środowisk naukowych, w których coraz istotniejszą rolę odgrywały zasoby obliczeniowe. Głównym celem zbudowanej infrastruktury było wsparcie badań naukowych poprzez integrację danych eksperymentalnych i wyników zaawansowanych obliczeń komputerowych, prowadzonych przez geograficznie rozproszone zespoły naukowców, przy użyciu zasobów centrów superkomputerowych. Cel ten został osiągnięty między innymi poprzez rozbudowę zasobów obliczeniowych we wszystkich instytucjach wchodzących w skład Konsorcjum. Co więcej, dzięki pracom realizowanym w projekcie PL-Grid, jesienią 2011 r. superkomputery wszystkich polskich akademickich centrów superkomputerowych znalazły się na TOP500 – liście najszybszych superkomputerów świata. Tego samego roku, superkomputer Zeus, zlokalizowany w Cyfronecie, znalazł się na 80 pozycji listy TOP500, będąc najszybszym superkomputerem w Polsce.

Kolejnym etapem realizacji Programu PLGrid było zapewnienie naukowcom profesjonalnego wsparcia informatycznego poprzez opracowanie specyficznych środowisk obliczeniowych, czyli usług i oprogramowania oraz udzielanej pomocy użytkownikom w planowaniu, uruchamianiu i analizie wyników złożonych eksperymentów naukowych. Przygotowanie dedykowanych środowisk informatycznych, nazywanych gridami dziedzinowymi, dopasowanych do potrzeb 13 różnych grup naukowców, było najważniejszym zadaniem projektu PLGrid Plus

(Dziedzinowo zorientowane usługi i zasoby infrastruktury PL-Grid dla wspomagania Polskiej Nauki w Europejskiej Przestrzeni Badawczej).

Przystosowanie infrastruktury do potrzeb naukowców konkretnych dziedzin było wielkim sukcesem projektu PLGrid Plus. Dlatego kierunek tych prac kontynuowano w ramach kolejnego projektu – PLGrid NG (Dziedzinowe Usługi Nowej Generacji w Infrastrukturze PL-Grid dla Polskiej Nauki), w ramach którego utworzone zostały usługi domenowo-specyficzne, spełniające wymagania kolejnych grup badawczych, reprezentujących 14 dziedzin nauki (łącznie w obu projektach zbudowano narzędzia informatycznego wsparcia dla 27 dziedzin nauki).



Uczestnicy Programu PLGrid nie poprzestali jednak na dostarczeniu domenowo-zorientowanych rozwiązań. Dzięki wieloletniemu zaangażowaniu w rozwój gridowych infrastruktur obliczeniowych, Cyfronet jest obecnie rozpoznawany jako Centrum Kompetencji w zakresie usług cloudowych i gridowych. Potwierdzeniem tego faktu było przyznanie ACK Cyfronet AGH środków na realizację nowego przedsięwzięcia: „Centrum kompetencji w zakresie rozproszonych infrastruktur typu gridowego – PLGrid Core". Będąc kolejnym krokiem implementacji Programu PLGrid, projekt ten miał na celu rozwój możliwości infrastruktury, m.in. w zakresie obliczeń chmurowych i na dużych zasobach danych. Oprócz rozbudowy sprzętu komputerowego czy tworzenia nowego oprogramowania, prace w projekcie koncentrowały się także na infrastrukturze towarzyszącej. Jedną z ważnych inwestycji była budowa nowego zapasowego centrum danych, zlokalizowanego w pewnej odległości od dotychczasowej serwerowni Cyfronetu, co przełożyło się bezpośrednio na zwiększenie bezpieczeństwa przechowywanych danych naukowych.

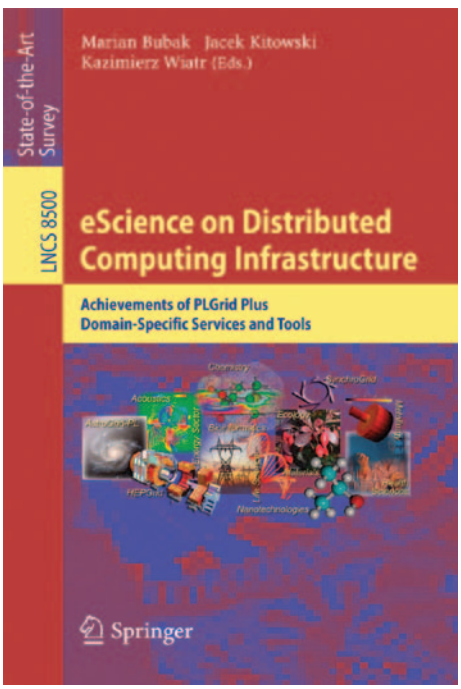
W ramach prac prowadzonych w projekcie PLGrid Core, w Cyfronecie uruchomiono nowy **superkomputer Prometheus – w edycji listy TOP500 z listopada 2015 roku został on sklasyfikowany na 38 miejscu, najwyższym w historii polskich superkomputerów!**

Obecnie w ramach infrastruktury PLGrid udostępniane są olbrzymie zasoby mocy obliczeniowych i pamięci masowych. W infrastrukturze zaimplementowanych zostało wiele narzędzi wspomagających organizację skomplikowanych eksperymentów obliczeniowych, projektowanie i uruchamianie aplikacji, wspierających obliczeniowo badania naukowe oraz wizualizację otrzymywanych wyników. Udostępniono także nową usługę – obliczenia w „chmurze".

Wszystkie projekty realizujące cele Programu PLGrid uzyskały finansowanie m.in. ze środków Europejskiego Funduszu Rozwoju Regionalnego w ramach Programu Operacyjnego Innowacyjna Gospodarka. ACK Cyfronet AGH miał zaszczyt pełnić odpowiedzialną rolę koordynatora tychże projektów. Szeroki wachlarz oferowanych usług przyczynia się do umacniania współpracy pomiędzy polskimi i międzynarodowymi grupami specjalistów z wielu różnych dziedzin nauki – także humanistycznych i społecznych. Istotnym jest fakt, iż każda osoba prowadząca badania naukowe może być użytkownikiem infrastruktury. Dostęp do wielkich mocy obliczeniowych, ogromnych zasobów pamięci oraz zaawansowanych usług na światowym poziomie jest nieodpłatny dla polskich naukowców i wszystkich osób prowadzących działalność naukową, związaną z uczelnią lub instytutem naukowym w Polsce. Jedynym warunkiem jest założenie konta przez Portal PLGrid.

Usługi zorientowane dziedzinowo w infrastrukturze PLGrid

Infrastruktura PLGrid powstała dzięki inicjatywie Cyfronetu oferuje jednolity dostęp do zasobów wszystkich pięciu polskich centrów Komputerów Dużej Mocy. Unifikacja ma miejsce na wielu poziomach, poczynając od jednego loginu i hasła użytkownika w całej infrastrukturze, aż po dostęp do aplikacji naukowych.



Mając na uwadze sygnalizowaną przez użytkowników potrzebę implementacji dodatkowych mechanizmów wsparcia prowadzenia badań naukowych, infrastruktura PLGrid została rozszerzona o szereg specyficznych środowisk, rozwiązań i usług, opracowanych zgodnie z wymaganiami zdefiniowanymi przez 13 pilotażowych grup naukowców w ramach projektu PLGrid Plus (2011-2015). Głównym celem projektu było ułatwienie naukowcom korzystania z infrastruktury, a tym samym, przyciągnięcie nowych społeczności użytkowników, którzy potrzebują dużej mocy obliczeniowej i przestrzeni dyskowej superkomputerów, ale posiadają zbyt małe umiejętności w posługiwaniu się nimi. Aby umożliwić i ułatwić tworzenie środowisk domenowo-specyficznych, w ramach projektu miała miejsce szeroka współpraca z przedstawicielami różnych dyscyplin, pracującymi często w konsorcjach dziedzinowych.

Dedykowane usługi dziedzinowe z jednej strony ukrywają złożoność infrastruktury, a z drugiej, w tym samym czasie, oferują funkcje istotne z punktu widzenia potrzeb badaczy z danego obszaru tematycznego. W ten sposób użytkownicy otrzymują dokładnie taką funkcjonalność, jakiej potrzebują. Co więcej, jest ona im dostarczona w sposób specyficzny dla danej dziedziny, tak aby osiągnąć maksymalną intuicyjność i użyteczność.

Osiągnięcia naukowe i techniczne projektu PLGrid Plus zostały przedstawione w książce opublikowanej w wydawnictwie Springer, we wrześniu 2014 roku. Książka ta jest ważnym źródłem informacji dla naukowców, programistów i administratorów systemów, którzy w swojej pracy korzystają ze środowisk sieciowych i chmurowych. Książka zawiera 36 rozdziałów i jest podzielona na trzy części: pierwsza (rozdziały 1 do 8) zawiera ogólny przegląd prac wykonanych w ramach projektu i opis obecnego stanu infrastruktury PLGrid, w tym nowych rozwiązań w dziedzinie bezpieczeństwa i oprogramowania pośredniczącego. Druga część (rozdziały od

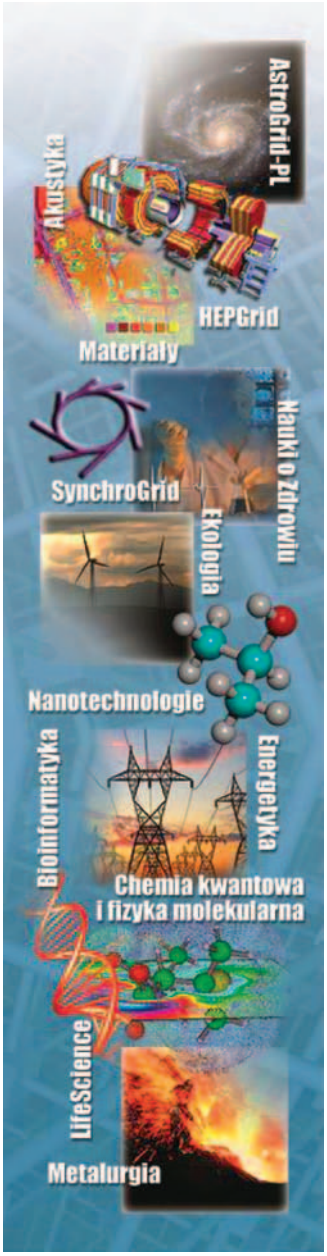
9 do 13) prezentuje nowe środowiska i usługi IT, które mogą być wykorzystywane przez wszystkie wyżej wymienione grupy naukowców. Trzecia część (rozdziały 14 do 36) opisuje, jak specyficzne środowiska, narzędzia i usługi, przygotowane w ramach projektu PLGrid Plus, są wykorzystywane w zaawansowanych obliczeniach i symulacjach komputerowych przeprowadzonych przez różne grupy naukowców. Rozdziały te przedstawiają modele obliczeniowe, nowe algorytmy i metody ich implementacji przy użyciu dostępnych narzędzi i usług.

Sukces projektu PLGrid Plus, a w szczególności rosnąca popularność specjalistycznych narzędzi i platform przygotowanych dla użytkowników z pierwszych 13 strategicznych obszarów nauki, doprowadziły do szybkiego wzrostu zapotrzebowania na podobne usługi dla naukowców z innych dziedzin. Dlatego, Konsorcjum PLGrid rozpoczęło projekt PLGrid NG (2014-2015), którego głównym celem było wdrożenie i udostępnienie, w ramach infrastruktury PLGrid, kilkunastu dodatkowych usług informatycznych dla grup naukowców reprezentujących 14 kolejnych obszarów nauki.

Nowe, zorientowane dziedzinowo usługi objęły szeroki zakres aktywności: począwszy od dostarczenia specjalistycznego oprogramowania, poprzez mechanizmy przechowywania danych, aż po nowoczesne platformy zintegrowane z nowym rodzajem narzędzi oraz dedykowanymi bazami danych, które przyspieszyły prowadzenie badań naukowych, jak również usprawniły i zautomatyzowały pracę grup badawczych.

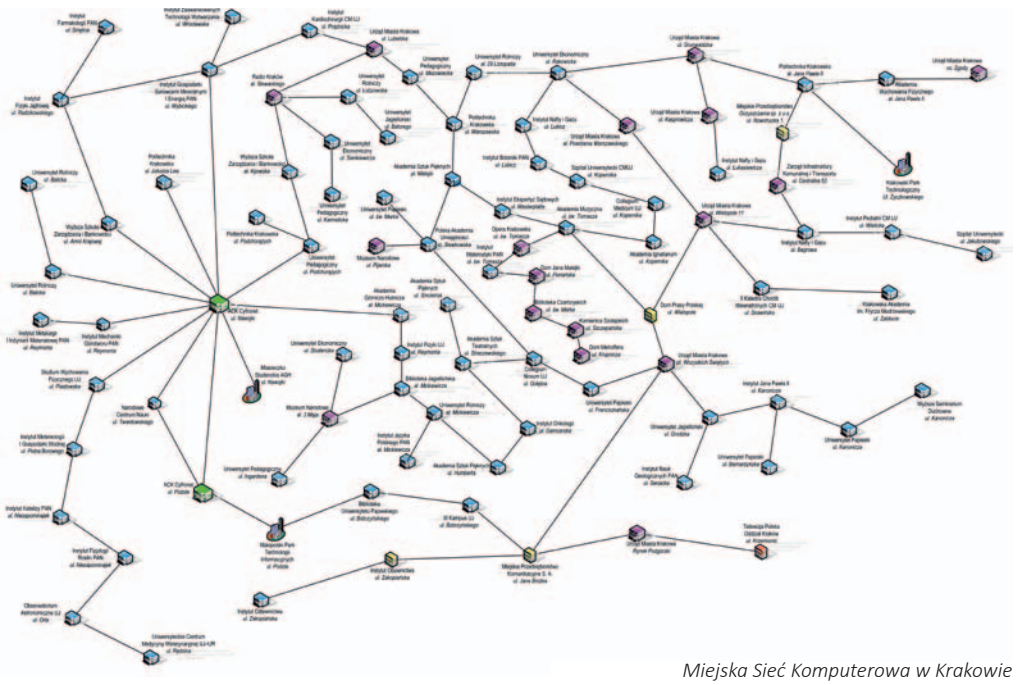
Przygotowanie i wdrożenie zestawu domenowo-specyficznych usług bardzo dobrze wpisało się w potrzeby wynikające z rozwoju zaawansowanej infrastruktury IT, przeznaczonej do realizacji nowoczesnych badań naukowych. Optymalnie dopasowana e-infrastruktura PLGrid spełnia nie tylko potrzeby naukowców dotyczące odpowiednich zasobów i usług obliczeniowych, ale także umożliwia polskim jednostkom naukowym współpracę z międzynarodowymi organizacjami badawczymi.

Rozbudowa istniejącej infrastruktury obliczeniowej w kierunku domenowo-specyficznych rozwiązań dla zespołów badawczych pozwoliła w efekcie końcowym na bardziej efektywne prowadzenie badań naukowych.



Miejska Sieć Komputerowa

ACK Cyfronet AGH stara się jak najlepiej służyć środowisku naukowemu, uczelniom i jednostkom naukowym w zakresie dostępu do krajowej i światowej sieci Internet oraz do mocy obliczeniowych na potrzeby badań naukowych. Dlatego ważnym zadaniem stawianym przed Cyfronetem jest zapewnienie bezawaryjnego funkcjonowania sieci oraz usług teleinformatycznych i obliczeniowych 7 dni w tygodniu przez 24 godziny na dobę.



Miejska Sieć Komputerowa w Krakowie

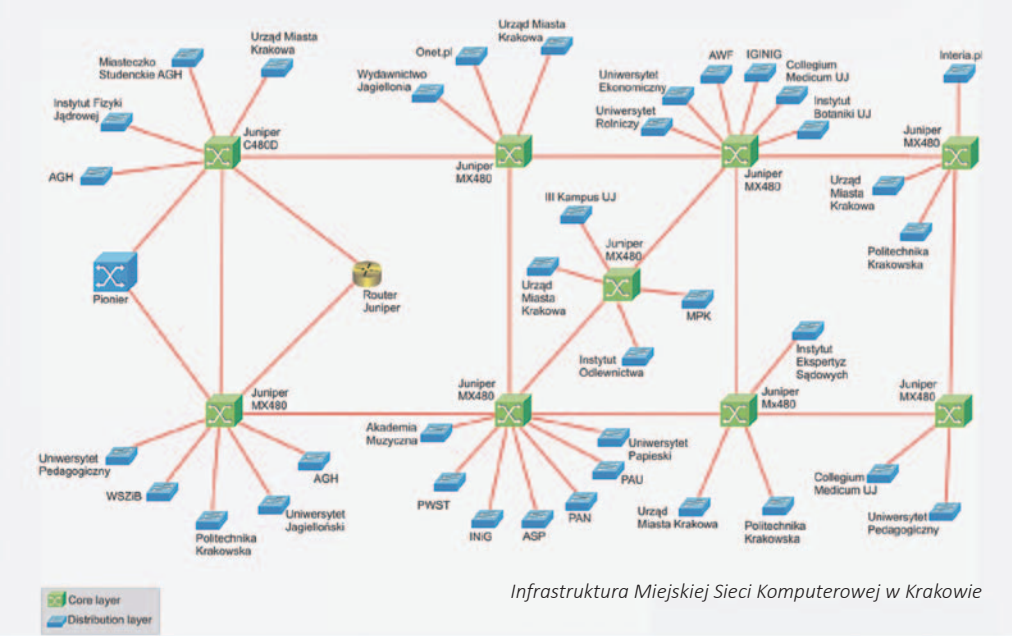
Charakterystyka MSK

ACK Cyfronet AGH, jako administrator Miejskiej Akademickiej Sieci Komputerowej MSK w Krakowie, stale rozbudowuje i modernizuje sieć, dostosowując ją do potrzeb i oczekiwań użytkowników. Obecnie, MSK eksploatuje własną infrastrukturę światłowodową o łącznej długości ponad 200 km. Światłowody MSK są ułożone w rejonie Starego Miasta, w okolicach kampusu AGH, docierają do Bronowic, Krowodrzy oraz do Czyżyn i Nowej Huty. Zrealizowane zostały także przyłączenia odległych instytutów w Prokocimiu, w rejonie Borku Fałęckiego oraz na terenie III Kampusu Uniwersytetu Jagiellońskiego w Pychowicach. Światłowody są układane w kanalizacji własnej, w dzierżawionej kanalizacji teletechnicznej oraz napowietrznie.

Infrastruktura światłowodowa jest podstawą funkcjonowania Miejskiej Akademickiej Sieci Komputerowej. ACK Cyfronet AGH dąży do tego, aby obejmowała ona jak największą liczbę obiektów uczelnianych i instytucji badawczych. Jednocześnie, z uwagi na stale rosnącą rolę

nowoczesnych środków łączności, w codziennej pracy bardzo ważnym jest, aby infrastruktura światłowodowa, oprócz dużej przepustowości, zapewniała również bezpieczną komunikację, uzyskiwaną w praktyce poprzez wykorzystanie łączy zapasowych pozwalających zachować ciągłość działania w sytuacjach zerwania tras podstawowych.

W warstwie rdzeniowej zastosowano nowoczesne przełączniki wyposażone w interfejsy 1 i 10 Gb Ethernet. Zapoczątkowano stopniowe doposażanie kluczowych przełączników w interfejsy 100 Gb. Wszystkie urządzenia warstwy rdzeniowej są połączone przynajmniej z dwoma lub trzema sąsiednimi, co pozwala automatycznie przełączyć ruch w przypadku awarii jednego z urządzeń lub awarii linii światłowodowej. Warstwę dystrybucyjną sieci oparto o przełączniki, z których każdy jest wyposażony w dwa interfejsy 1 Gb Ethernet (przyłączenie do warstwy rdzeniowej) oraz szereg interfejsów Ethernet typu 10/100/1000 Mbps służących do przyłączania użytkowników.



Infrastruktura Miejskiej Sieci Komputerowej w Krakowie

Miejska Akademicka Sieć Komputerowa w Krakowie jest ważnym węzłem Akademickiej Sieci PIONIER – ogólnopolskiej szerokopasmowej sieci optycznej stanowiącej istotną bazę dla badań naukowych i prac rozwojowych w Polsce. MSK jest połączona w kierunku Warszawy, Katowic, Bielska-Białej i Rzeszowa łączami sieci PIONIER, każde o przepływności 2x10 Gbps. Łączność pomiędzy Centrami Komputerów Dużej Mocy w Polsce (Gdańsk, Kraków, Poznań, Warszawa i Wrocław) realizowana jest z przepustowością 2x100 Gbps. Za pośrednictwem sieci PIONIER realizowana jest komunikacja z wieloma ośrodkami krajowymi oraz zagranicznymi. Łączność zagraniczna odbywa się poprzez europejską naukową sieć GEANT o przepustowości 100 Gbps. Oprócz głównego połączenia do sieci GEANT, realizowane jest połączenie rezerwowe o przepustowości 5 Gbps do sieci operatorów Telia Carier Poland oraz Lumen Technologies.

Usługi sieciowe dedykowane użytkownikom

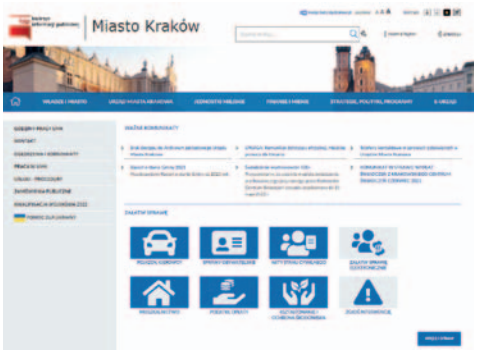
Akademickie Centrum Komputerowe Cyfronet AGH od początku powstania polskiego Internetu (w połowie 1991 roku) aktywnie uczestniczy w rozwoju zarówno infrastruktury telekomunikacyjnej, jak też, co bardzo istotne, w rozwoju szeroko rozumianych usług internetowych. Realizowane na bardzo wydajnych komputerach usługi obejmują między innymi:

- **e-mail** – internetowa poczta elektroniczna, z dostępem poprzez protokół SMTP lub interfejs webowy *http://poczta.cyfronet.pl*,
- **www** – udostępniane serwisy, obok nowości ze świata nauki, prezentują informacje dotyczące kultury, zabytków oraz wielu innych dziedzin,
- **news** – forum dyskusyjne podzielone na grupy tematyczne z wielu dziedzin, w tym dyskusje naukowe przeznaczone dla wąskiego grona specjalistów,
- **dns** – serwery systemu nazw domen – obsługujące usługę tłumaczenia nazw sieciowych na adres IP dla użytkowników Miejskiej Akademickiej Sieci Komputerowej,
- **ftp** – serwer przechowuje i stale aktualizuje kopie zagranicznych archiwów oprogramowania (tzw. *mirror*), a także zawiera oprogramowanie typu shareware i freeware pracujące pod systemami MS-Windows oraz UNIX; dzięki uruchomieniu serwera ftp zostało znacznie ograniczone obciążenie łącza międzynarodowego, a także zwiększono wygodę użytkowników Miejskiej Akademickiej Sieci Komputerowej,
- **eduroam** – umożliwia łączność z siecią akademicką w każdym miejscu na świecie objętym zasięgiem eduroamu, przy użyciu jednego autoryzowanego konta,
- **box** – dysk sieciowy (*http://box.cyfronet.pl*) umożliwiający wymianę plików i ich synchronizację; dzięki dedykowanym aplikacjom może być również używany na urządzeniach przenośnych.

Usługi sieciowe w liczbach (w roku 2021)	
Liczba wysłanych i odebranych wiadomości e-mail	> 17 000 000
Liczba sesji nawiązanych z serwerem pocztowym	> 50 000 000

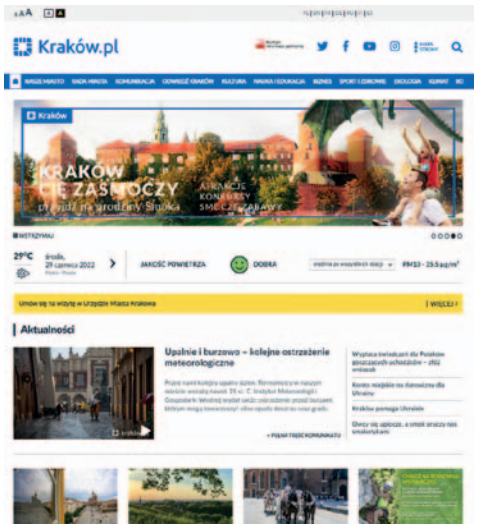
Portale i aplikacje mobilne

Centrum nie ogranicza swojej działalności jedynie do obszarów powiązanych z nauką, ale przyczynia się także do rozwoju społeczeństwa informacyjnego. Serwer WWW w ACK Cyfronet AGH pełni funkcję węzła internetowego dla całej krakowskiej społeczności naukowej. Centrum wciąż ulepsza i rozbudowuje swój portal internetowy, który na przestrzeni ostatnich lat zyskał dużą popularność. W roku 2021 została opublikowana nowa wersja portalu, wraz z wersją dla urządzeń mobilnych.



Szczególne znaczenie dla Centrum ma współpraca z władzami Krakowa. Umowa pomiędzy Gminą Kraków i Cyfronetem, dotycząca promocji miasta, zaowocowała stworzeniem portalu internetowego, który oprócz prezentowania informacji naukowych zapoznaje czytelników z kulturą, zabytkami, turystyką, transportem lokalnym i wieloma innymi aspektami życia w Krakowie.

We współpracy z Urzędem Miasta Krakowa został przygotowany i uruchomiony Biuletyn Informacji Publicznej dla regionu krakowskiego. W roku 2005, współpraca ta została rozszerzona o świadczenie usług w zakresie publikacji treści dla miejskich instytucji, bibliotek, szkół, itp.



Najnowsza wersja oficjalnego serwisu internetowego miasta Krakowa została opublikowana jako Miejska Platforma Internetowa „Magiczny Kraków". W roku 2007 portal ten – dostępny pod adresem *www.krakow.pl* – został nominowany do World Summit Award jako najlepszy serwis e-administracji w Polsce. Mobilna wersja portalu została natomiast nagrodzona podczas konferencji Mobile Trends w roku 2012 jako najlepsza mobilna strona miejska w Polsce.

Współpraca z Urzędem Miasta wchodzi również w obszar urządzeń mobilnych – Cyfronet opracował m.in. mobilną aplikację *Kraków.pl*, która może być wykorzystana jako przewodnik miasta Krakowa i źródło ważnych informacji, takich jak numery telefonów, lokalizacje punktów informacyjnych, konsulatów lub aptek. Co ważne, aplikacja umożliwia znalezienie wszystkich tego typu miejsc w trybie mapy off-line. Aplikacja *Kraków.pl* jest dostępna w kilku wersjach językowych.

www.CYFRONET.krak

Archiwum KSAF

Archiwum KSAF

Zasoby obliczeniowe

ACK Cyfronet AGH dysponuje dojrzałą i unikatową infrastrukturą obliczeniową dostępną dla polskich naukowców. Jest ona oparta na pięciu filarach, rozbudowanych o szereg usług oraz rozwiązań przygotowanych wraz z naukowcami na potrzeby badaczy z różnych dziedzin nauki. Wraz ze szkoleniami i kompleksowym wsparciem, zasoby dostępne w naszym centrum obliczeniowym pozwalają na efektywne przeprowadzanie wszelkich badań naukowych wykorzystujących infrastrukturę komputerową.

Zasoby obliczeniowe

Superkomputery Athena, Ares i Prometheus dostarczają:
15 PFlops
100 000+ rdzeni
600 GPGPU
950+ TB RAM.

Zasoby dyskowe

45 PB przestrzeni dyskowej i 25 PB zasobów taśmowych na dane wraz z szybkim systemem plików Lustre umożliwia analizę wielkich zbiorów danych.

Profesjonalne oprogramowanie

Narzędzia i programy oraz biblioteki, które umożliwiają efektywne wykorzystanie zasobów obliczeniowych Cyfronetu, a także usprawniają pracę w różnych dziedzinach nauki.

Narzędzia pracy zespołowej

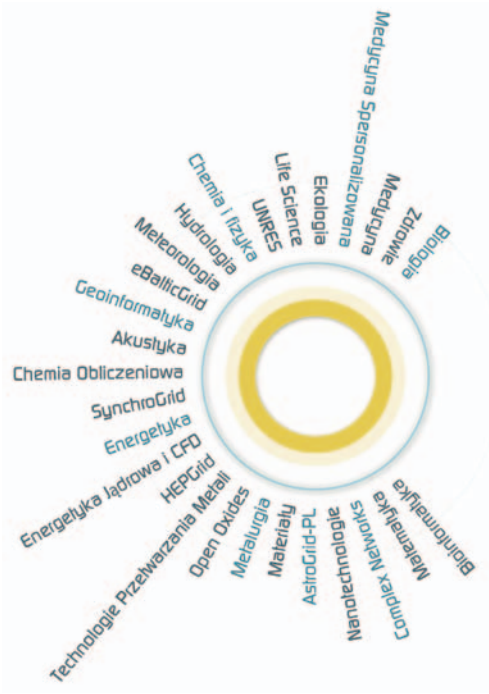
Serwisy i narzędzia takie jak serwer repozytoriów Bitbucket Git oraz środowisko do zarządzania projektami JIRA ułatwiają realizację i koordynację złożonych projektów oraz współpracę zespołów badawczych.

Chmura obliczeniowa

Elastyczne rozwiązanie oparte o platformę OpenStack, pozwalające na łatwe dostosowanie środowiska obliczeniowego do potrzeb grupy badawczej.

Zaawansowane platformy obliczeniowe oraz usługi dla domen badawczych

Naukowcy prowadzący prace badawcze z wykorzystaniem superkomputerów i dużych zasobów pamięci masowej mają różne wymagania oraz sposoby pracy, zatem potrzebują różnego typu dostępu do komputerów oraz innych zasobów infrastruktury IT. Odpowiadając na potrzeby użytkowników, Cyfronet udostępnia szereg zaawansowanych platform informatycznych i dedykowanych usług. Ukrywają one złożoność infrastruktury i jednocześnie dostarczają funkcjonalności istotne z punktu widzenia badaczy z danej dziedziny, precyzyjnie dopasowane do ich potrzeb.



Wraz z infrastrukturą obliczeniową udostępniamy zbiór narzędzi, który umożliwia badaczom wykonywanie złożonych eksperymentów o wielkiej skali oraz efektywne i łatwe zarządzanie ich rezultatami. Efektywność obliczeń i analiz oraz bezpieczeństwo analizowanych danych są gwarantowane poprzez odpowiedni dobór rozwiązań IT rozwijanych oraz konfigurowanych przez doświadczonych programistów zatrudnionych w Cyfronecie. Platformy prezentowane poniżej zostały z powodzeniem zastosowane w gridach dziedzinowych dostępnych w Programie PLGrid. Wszystkie narzędzia, usługi oraz serwisy są udostępniane w ramach infrastruktury PLGrid, co pozwala polskim naukowcom i ich zagranicznym współpracownikom na dostęp do zasobów oraz danych w najbardziej dla nich wygodny sposób.

W Centrum oferujemy m.in. zaawansowane narzędzia i interfejsy graficzne umożliwiające budowę dedykowanych środowisk do prowadzenia badań naukowych, przeprowadzanie wirtualnych eksperymentów, budowę portali webaplikacji, wizualizację wyników obliczeń, uruchamianie złożonych scenariuszy obliczeniowych, jak również wspierające jednolity i efektywny dostęp do danych. Wszystkie te usługi są ważnym wsparciem dla osób prowadzących badania naukowe, wpływają bowiem na usprawnienie i, tam gdzie to możliwe, zautomatyzowanie pracy grup badawczych, co znacznie przyspiesza uzyskanie wyników badań naukowych. Na kolejnych stronach znajdują się opisy ich możliwości.

Zaproszenie do współpracy

Poszukujemy osób zainteresowanych rozwojem usług obliczeniowych, w tym tych zorientowanych na konkretne dziedziny badawcze. Oferujemy również wsparcie w prowadzeniu badań naukowych.

Zachęcamy naukowców do udostępnienia swoich autorskich kodów ekspertom naszego Centrum, którzy będą mogli pomóc w ich efektywnej kompilacji na naszych superkomputerach. Równocześnie zapewniamy pomoc w ich użytkowaniu i efektywnym wykorzystaniu. Umożliwiamy również wykorzystanie własnych licencji oprogramowania naukowego posiadanych przez grupy badawcze, gdy umożliwiają one obliczenia na superkomputerach Cyfronetu.

Chmura obliczeniowa w infrastrukturze PLGrid

W trakcie tworzenia infrastruktury PLGrid w znacznej mierze kierowano się potrzebami naukowców. Dlatego parametry infrastruktury pozwalają na jej łatwe dostosowanie do rozwiązywania nawet skomplikowanych problemów badawczych przez różne grupy badaczy – począwszy od małych zespołów, a skończywszy na międzynarodowych konsorcjach naukowych. Aby spełnić ich wymagania, obok typowych platform obliczeniowych i przechowywania danych, oferujemy „chmurę obliczeniową".

– *Do tej pory, infrastruktura PLGrid dostarczała szeregu platform informatycznych, wraz z zasobami obliczeniowymi i pamięcią masową. Chmura obliczeniowa nie stanowi jedynie ich rozszerzenia. Traktujemy ją jako nowy poziom jakości w prowadzeniu badań* – mówi Kazimierz Wiatr, Dyrektor Centrum. – *Użytkownik może w łatwy sposób połączyć się z żądanym zestawem maszyn wirtualnych (VM), z pełnymi prawami dostępu do systemu operacyjnego. W celu zapewnienia wysokiego poziomu bezpieczeństwa, wszystkie maszyny wirtualne działają w dedykowanej sieci lokalnej. Dostęp do poszczególnych usług można uzyskać z całego świata, co ułatwia współpracę między naukowcami* – dodaje Dyrektor.

Należy podkreślić następujące zalety płynące z korzystania z chmury obliczeniowej:

- Chmura zwiększa elastyczność badań, ponieważ naukowcy mogą dostroić maszyny wirtualne do swych specyficznych potrzeb. Do tej pory, aby skonfigurować „wirtualne laboratorium" służące do rozwiązania określonego problemu naukowego, użytkownik potrzebował pomocy ze strony ekspertów PLGrid. Obecnie, każdy naukowiec może samodzielnie stworzyć i łatwo rozbudować takie wirtualne laboratorium.
- Katalog maszyn wirtualnych oferowanych przez PLGrid wspiera wiele systemów operacyjnych. Dzięki temu użytkownicy mogą uruchamiać swoje aplikacje w systemach operacyjnych innych niż Scientific Linux, np. w Windows lub innych systemach linuxowych.
- Chmura obliczeniowa pozwala na łatwe zbudowanie i uruchomienie środowiska testowego. Ta funkcjonalność jest bardzo przydatna dla naukowców tworzących swoje własne oprogramowanie. Mogą w prosty sposób uruchomić każde zadanie testowe i szybko przeanalizować jego wyniki.
- Możliwe jest utrzymanie komunikacji z już uruchomionym zadaniem obliczeniowym. Ponadto, każda maszyna wirtualna może być łatwo powielana, nawet w tysiącach lub więcej egzemplarzy. Uruchomienie nowej maszyny wirtualnej zajmuje niespełna 30 sekund.
- Chmura obliczeniowa jest także najlepszym, a w wielu przypadkach jedynym, rozwiązaniem do uruchamiania zadań za pośrednictwem starszych pakietów oprogramowania. W bezpiecznym środowisku sieci LAN można korzystać nawet z przestarzałych, wycofanych już z powszechnego użycia, systemów operacyjnych. Ta funkcjonalność jest również dobrym rozwiązaniem dla rozproszonych grup międzynarodowych, wykorzystujących w swych badaniach wiele różnych pakietów. Każda grupa może prowadzić własne obliczenia i w prosty sposób udostępniać swoje wyniki innym.

– *Chmura obliczeniowa w PLGrid i Cyfronecie to innowacyjne rozwiązanie na skalę europejską. Mamy silne przekonanie, że wprowadzi ona nowy poziom jakości dla badań prowadzonych przez naszych użytkowników* – podsumowuje Profesor Wiatr.

Obecnie, w ramach zasobów Cyfronetu eksploatowanych jest **ponad 200 różnych typów maszyn wirtualnych**.

Onedata jest globalnie skalowalnym systemem zarządzania danymi, który pozwala na przezroczysty dostęp do danych przechowywanych w rozproszonych systemach składowania danych. Onedata nadaje się równie dobrze dla małych grup użytkowników jak i dla dużych międzynarodowych środowisk naukowych, dzięki czemu mogą korzystać z tego systemu zarówno federacje użytkowników naukowych jak i komercyjnych. Onedata umożliwia użytkownikom korzystanie z jednorodnego systemu zarządzania danymi zarówno do przechowywania danych osobistych jak i związanych z pracą zawodową, np. wyników badań, oraz daje możliwość dostępu do nich w wydajny sposób z dowolnego urządzenia.

Onedata pozwala na budowanie federacji zaufanych użytkowników oraz dostawców przestrzeni dyskowej, co pozwala użytkownikom z jednej strony na korzystanie z rozproszonej przestrzeni dyskowej a z drugiej na kontrolowanie ilości zasobów udostępnianych poszczególnym użytkownikom lub grupom roboczym przez administratorów, bez konieczności stosowania certyfikatów.

Zalety dla użytkowników

- Zunifikowany dostęp do danych rozproszonych pomiędzy dostawców przestrzeni dyskowej opartych o heterogeniczne systemy. Korzystając z Onedata, użytkownicy mają dostęp do swoich danych w każdym miejscu i z dowolnego urządzenia.
- Dane użytkowników zorganizowane są w wirtualne foldery (ang. *Space*), które mogą być rozproszone pomiędzy dostawców przestrzeni dyskowej i dostępne dla użytkowników z przeglądarki WWW lub poprzez protokół POSIX.
- Łatwy w użyciu Graficzny Interfejs Użytkownika pozwalający na dostęp i wyszukiwanie danych oraz zarządzanie nimi.
- Wsparcie dla łatwego dzielenia się i współpracy z innymi użytkownikami w oparciu

o dane, przy zachowaniu wysokiego poziomu bezpieczeństwa opartego o listy kontroli dostępu oraz złożony system grup i uprawnień do nich.

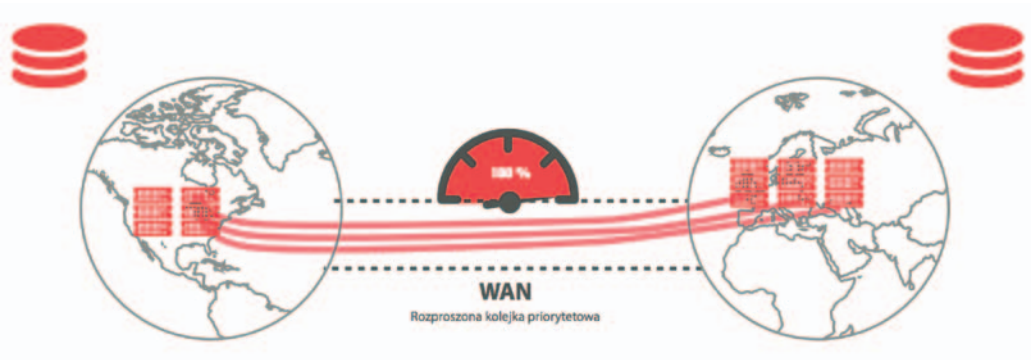
- Funkcjonalność publikowania i dostępu do danych otwartych oraz możliwość łatwej rejestracji identyfikatorów typu DOI dla zbiorów danych.

Zalety dla administratorów

- Łatwość instalacji i konfiguracji poprzez kontenery Docker i klientów powłoki.
- Uproszczony mechanizm przy zachowaniu pełnej kontroli administratorów w oparciu o system tokenów.
- Złożony system monitorowania dostarczający informacji o wszystkich aspektach systemu, dostępny poprzez interfejs REST oraz w postaci graficznej w panelu administracyjnym.
- Wsparcie dla wielu systemów składowania danych, w tym POSIX (np. Lustre), Amazon S3, Ceph, OpenStack Swift, czy GlusterFS.

Zalety dla twórców aplikacji

- Łatwa integracja z systemem poprzez interfejsy REST i CDMI.
- Elastyczna autentykacja i autoryzacja dostępna o technologie OpenID Connect i Macaroons.
- Kompletna dokumentacja interfejsów w postaci pozwalającej na automatyczne generowanie kodu aplikacji klienckich odpowiedzialnego za komunikację z usługami Onedata.

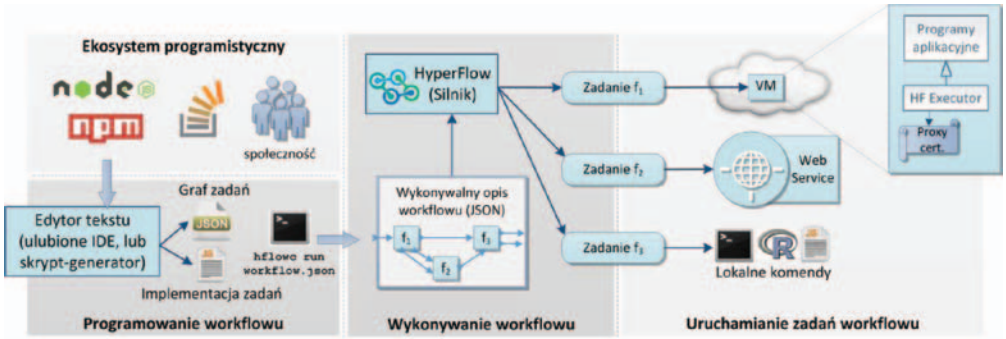


Użytkownicy Onedata

Onedata jest aktualnie uruchomione oraz testowane w ramach kilku infrastruktur europejskich, w tym w projektach: PLGrid, INDIGO-DataCloud, EGI DataHub, Open Nebula Science Cloud oraz Human Brain Project. W projekcie HBP Onedata wykazało możliwość dostarczenia systemu zarządzania danymi pozwalającemu na spełnienie wymagań przepustowości dla aplikacji wizualizacji przekrojów mózgu w czasie rzeczywistym.

Więcej informacji można znaleźć pod adresem *https://onedata.org*.

HyperFlow jest narzędziem, które umożliwia kompozycję aplikacji naukowych w złożone przepływy, tzw. *workflowy naukowe*. HyperFlow wspomaga tworzenie przepływów, rozmieszczanie ich w chmurze obliczeniowej oraz ich uruchamianie.



Programowanie workflowów naukowych

W HyperFlow workflow jest opisywany jako graf zadań (nazywanych *procesami*) w prostym formacie opartym o JSON. Zadania workflowu reprezentują procedury naukowe – kolejne kroki w przepływie naukowym. Zadania te mogą być albo zaimplementowane w języku JavaScript, albo odwzorowane na programy wykonywalne. W pierwszym przypadku zadania workflowu są wykonywane przez `Node.js`, szeroko wykorzystywane środowisko uruchomieniowe dla języka JavaScript. Dzięki temu programowanie workflowu osadzone jest w dojrzałym ekosystemie programistycznym (społeczność, narzędzia, biblioteki, zasoby sieciowe), nie zaś w środowisku specyficznym dla danego systemu, mało znanym i wspieranym w ograniczonym stopniu. Dzięki możliwości niskopoziomowego programowania zadań workflowu, mogą one z łatwością wywoływać zewnętrzne Web Service'y, czy też wykonywać lokalne komendy, w ramach jednego przepływu.

W drugim przypadku twórca workflowu nie musi implementować kodu JavaScript, a jedynie przypisać każde zadanie workflowu do uprzednio utworzonej maszyny wirtualnej, na której zainstalowane są aplikacje naukowe, oraz określić komendę, która zostanie wykonana w przypadku uruchomienia danego zadania.

Dzięki wsparciu tych dwóch podejść do programowania workflowu, HyperFlow jest przydatny zarówno doświadczonym programistom / inżynierom, którzy pragną mieć możliwość programowania niskopoziomowego, jak i naukowcom dziedzinowym, którzy chcą jedynie komponować przepływy z uprzednio istniejących modułów.

Rozmieszczanie w chmurze

HyperFlow automatycznie rozmieszcza workflow w chmurze obliczeniowej. Aby to było możliwe, użytkownik musi jedynie przygotować plik konfiguracyjny, w którym określone jest odwzorowanie zadań workflowu na dostępne obrazy maszyn wirtualnych. Następnie użytkownik wykonuje prostą komendę *hflowc setup*, w wyniku której tworzone są instancje maszyn wirtualnych w chmurze. Instancje te zawierają komponenty środowiska uruchomieniowego HyperFlow oraz aplikacje naukowe uruchamiane z zadań workflowu.

Wykonanie workflowu

Po utworzeniu instancji workflowu w chmurze, użytkownik może ją uruchomić przy pomocy komendy *hflow run <workflow_directory>*. Każdy workflow uruchomiony jest ze swoją własną instancją środowiska uruchomieniowego HyperFlow. Powoduje to, że poszczególne workflowy są od siebie izolowane, co zwiększa bezpieczeństwo i niezawodność.

Środowisko uruchomieniowe HyperFlow (tzw. *HyperFlow Executor*) automatycznie przesyła dane wejściowe z katalogu użytkownika na instancje maszyn wirtualnych, wykonuje aplikacje naukowe, a następnie przesyła dane wyjściowe z powrotem do katalogu użytkownika. HyperFlow obsługuje kilka mechanizmów transferu danych, m.in. sieciowy system plików, gridftp oraz Amazon S3.

Aplikacje

HyperFlow jest częścią kilku większych systemów, w których jest wykorzystywany do implementacji szeregu aplikacji. W infrastrukturze PLGrid HyperFlow jest zintegrowany jako system zarządzania workflowami naukowymi, który umożliwia kompozycję naukowych workflowów i ich uruchamianie w chmurze. Przykładową aplikacją wspieraną przez HyperFlow jest solwer dla problemów opartych na metodzie elementów skończonych, który może być użyty do rozwiązywania wielu różnorodnych problemów naukowych. W projekcie Paasage (*http://www.paasage.eu*) HyperFlow jest integrowany jako silnik wykonawczy dla aplikacji naukowych uruchamianych w kontekście środowiska *multi-cloud*. W projekcie ISMOP (*http://www.ismop.edu.pl*) HyperFlow jest częścią systemu podejmowania decyzji w sytuacjach zagrożenia powodziowego. Hyperflow będzie miał również zastosowanie w obliczeniach parametrycznych i przetwarzaniu typu workflow, jako zamiennik technologii Scalarm.

Kontakt

HyperFlow jest rozwijany i utrzymywany przez zespół DICE (*http://dice.cyfronet.pl*). Zapraszamy do zgłaszania wszelkich pytań oraz sugestii.

Kod źródłowy HyperFlow oraz podręczniki instalacji i użytkowania są dostępne na stronie *https://github.com/hyperflow-wms/hyperflow*.

Model Execution Environment

Model Execution Environment (MEE) to stos technologiczny umożliwiający wykonywanie strumieni zadań na infrastrukturach komputerowych wielkiej mocy, w tym na zasobach udostępnianych przez ACK Cyfronet AGH. Celem platformy jest umożliwienie łatwego prowadzenia obliczeń przez naukowców dziedzinowych, tj. przez osoby posiadające podstawową wiedzę z zakresu interakcji z klastrami obliczeniowymi i innymi rodzajami zasobów HPC.

Strumienie, kroki i modele obliczeniowe

W ramach MEE zadania obliczeniowe przyjmują formę tzw. strumieni (*pipelines*), tj. zbiorów obliczeń, z których każde nosi nazwę kroku (*step*), gdzie rezultaty każdego kroku stanowią dane wejściowe dla kolejnych kroków. MEE udostępnia szeroką gamę narzędzi pozwalających użytkownikom na projektowanie kroków, zestawianie ich w strumienie zadań oraz zlecanie wykonywania tych strumieni na dostępnych zasobach obliczeniowych.

Każdy krok jest zdefiniowany w oparciu o zestaw modułów obliczeniowych (tj. wykonywalnego kodu) przechowywanych w repozytorium GitLab. Moduły te określamy zbiorczo mianem modelu. Gdy pojawia się potrzeba wykonania danego kroku, platforma MEE automatycznie wczytuje dany model na zasoby HPC, a następnie monitoruje jego wykonywanie w oparciu o wybrane dane wejściowe. Przed uruchomieniem strumienia zadań użytkownik jest w stanie wybrać konkretną wersję modelu (dzięki mechanizmom wersjonowania udostępnianym przez Git), co z kolei umożliwia śledzenie rezultatów oraz gwarantuje powtarzalność obliczeń.

Strumienie mogą być uruchamiane w trybie automatycznym lub ręcznym. Strumień automatyczny jest wykonywany w całości, od początku do końca, natomiast w strumieniach ręcznych po każdym kroku następuje przerwa, zaś użytkownik musi ręcznie zlecić wykonanie kolejnego kroku.

Mechanizm ten umożliwia weryfikację rezultatów pośrednich i pozwala np. na zaoszczędzenie zasobów poprzez anulowanie strumienia, który nie generuje pożądanych wyników.

MEE jest dostępne poprzez wygodny webowy interfejs UI. Platforma udostępnia również interfejsy programistyczne pozwalające na jej integrację z zewnętrznymi narzędziami.

Zarządzanie danymi badawczymi

MEE zleca i monitoruje wykonanie strumieni zadań na zasobach HPC, a zatem musi też dbać o zarządzanie związanymi z nimi danymi naukowymi, tak aby każdy model operował na właściwych danych i by istniała możliwość pobrania rezultatów z infrastruktury HPC. W tym celu MEE udostępnia zestaw interfejsów zarządzania danymi – użytkownicy mogą przesyłać dane wejściowe i pobierać rezultaty. Platforma zarządza komunikacją z zasobami storage'owymi HPC, umożliwiając automatyczne wprowadzanie danych wejściowych i pobieranie wyników. Opisywana funkcjonalność obejmuje również dedykowane elementy interfejsu użytkownika.

Bezpieczeństwo

Platforma MEE jest zintegrowana z mechanizmami uwierzytelniania i bezpieczeństwa infrastruktury PLGrid. Użytkownicy posiadający konta w infrastrukturze PLGrid mogą uzyskiwać dostęp do MEE za ich pośrednictwem oraz zlecać obliczenia w ramach swoich przydziałów. Każde obliczenie wykonywane jest w ramach konkretnego grantu badawczego, który może zostać skonfigurowany w ramach platformy.

Organizacje

MEE udostępnia poszczególnym zespołom badawczym wyodrębnione środowiska pracy. Każda organizacja posiada unikalny URL dostępowy i może definiować własne strumienie oraz kroki, a także zarządzać własnymi danymi badawczymi. MEE dba o rozgraniczenie organizacji – każdą z nich można zarządzać w autonomiczny sposób, umożliwiając dostęp wybranym grupom użytkowników.

Aplikacje

Platforma Model Execution Environment znalazła zastosowanie w szeregu projektów badawczych, do których zaliczamy europejskie projekty EurValve, PRIMAGE oraz InSilicoWorld, konsorcjum POLVAS, Centrum Medycyny Obliczeniowej Sano, a także szereg zespołów powstających na zasadzie ad hoc i dysponujących wyodrębnionymi organizacjami w ramach MEE.

https://mee.cyfronet.pl

fot. Adam Frączek, KSAF AGH

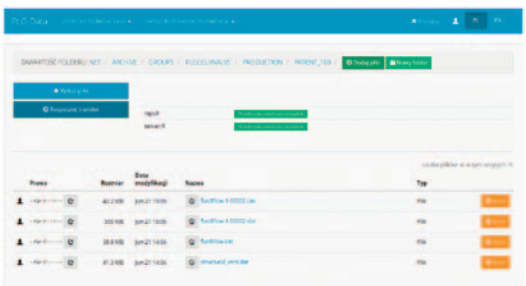
fot. Krzysztof Mastalski, KSAF AGH

Proste narzędzie zarządzania plikami na klastrze obliczeniowym

PLG-Data

Usługa PLG-Data pozwala, za pośrednictwem przeglądarki WWW, zarządzać swoimi zbiorami danych, zgromadzonymi na zasobach dyskowych infrastruktury PLGrid. Dzięki niej można oglądać, pobierać, dodawać i usuwać pliki oraz foldery. Ułatwia także zarządzanie prawami dostępu dla innych osób, członków zespołu badawczego lub współpracowników spoza tej grupy. Obecnie posiada wsparcie dla klastra Prometheus, a w przyszłości planowane jest rozszerzenie dostępności innych superkomputerów poprzez tę usługę.

Zestaw funkcjonalności, na które pozwala narzędzie, to m.in.:



- pobieranie plików z klastra na dysk lokalny,
- dodawanie nowych i usuwanie istniejących zbiorów,
- zmiana nazw plików i folderów, jak również zmiana uprawnień dostępu do nich,
- szybka nawigacja pomiędzy katalogami: domowym, roboczym i zespołów, dzięki poręcznemu menu skrótów,
- sprawny podgląd plików graficznych bez konieczności pobierania ich na dysk.

Dzięki odpowiedniej konstrukcji adresu URL do poszczególnych zbiorów, narzędzie pozwala łatwo przekazać ich lokalizację innej osobie, np. poprzez wklejenie wprost treści URL z paska adresu przeglądarki do wiadomości e-mail czy komunikatora. Adresat znajdzie współdzielony katalog lub pobierze plik jednym kliknięciem – pod warunkiem, że posiada odpowiednie uprawnienia dostępu.

Usługa jest zabezpieczona szyfrowanym protokołem HTTPS (na odcinku połączenia komputera użytkownika z portalem PLG-Data) oraz protokołem GridFTP – pomiędzy portalem usługi oraz klastrem obliczeniowym. Zastosowanie tych technik pozwala użytkownikom korzystać z niej w sposób bezpieczny. Osoba korzystająca z usługi PLG-Data nie nabywa żadnych dodatkowych uprawnień w dostępie do zbiorów danych poza tymi, które już posiada.

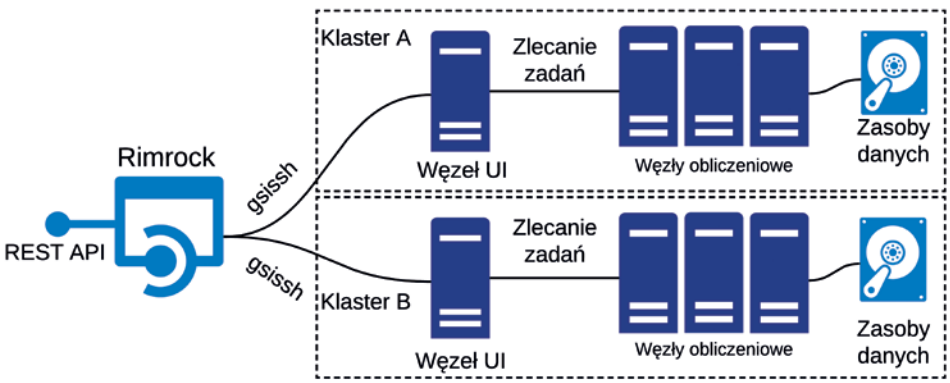
Logowanie do narzędzia odbywa się poprzez parę login-hasło PLGrid lub certyfikat p12 załadowany w przeglądarce. Usługa jest dostępna w polskiej i angielskiej wersji językowej, posiada także rozbudowany interfejs programistyczny (API), który pozwala na jej użycie przy budowie platform, narzędzi czy innych usług wykorzystujących zasoby infrastruktury PLGrid.

Adres narzędzia PLG-Data: *https://data.plgrid.pl*

Rimrock to usługa infrastruktury PLGrid umożliwiająca kontrolę obliczeń naukowych i zarządzanie otrzymywanymi za ich pomocą rezultatami dzięki nowoczesnym interfejsom dostępowym opartym o REST. REST (ang. *Representational State Transfer*) jest swego rodzaju wzorcem projektowania architektur rozproszonych aplikacji. Dzięki wykorzystaniu REST możliwy jest łatwy dostęp do usług, aplikacji i zaawansowanych skryptów.

Dostępność dla różnorodnych zastosowań

Zastosowanie interfejsów REST przy budowie usługi rimrock umożliwia wykorzystanie funkcjonalności usługi niezależnie od języka programowania, który został wybrany do tworzenia systemów wykorzystujących zasoby obliczeniowe. Można więc w oparciu o rimrock budować aplikacje internetowe i desktopowe, jak również tworzyć zaawansowane skrypty obliczeniowe (np. za pomocą języka powłoki *Bash* i polecenia *curl*). Ciekawym rozwiązaniem oferowanym przez usługę jest możliwość tworzenia aplikacji internetowych uruchamianych w całości w przeglądarce użytkownika, przy zminimalizowaniu udziału strony serwerowej.



Integracja z wieloma systemami zarządzania obliczeniami

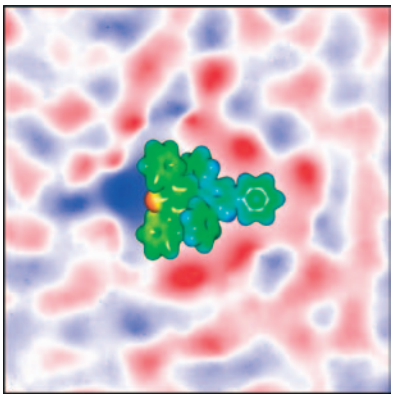
Usługa rimrock wykorzystuje system zarządzania zadaniami Slurm, dzięki czemu zapewnia wsparcie dla jego unikatowych własności. Pozwala to na łatwą integrację aplikacji już istniejących w rozwijanych systemach. Dostęp do rezultatów obliczeń jest ułatwiony poprzez ukrycie wewnętrznego protokołu przesyłania plików (*GridFTP*) i grupowanie wyników zadań.

Bezpieczeństwo danych

Wymiana danych z usługą jest realizowana za pomocą bezpiecznego połączenia HTTPS, a do autoryzacji użytkowników wykorzystywany jest tymczasowy certyfikat użytkownika (tzw. *proxy)*.

https://submit.plgrid.pl

Chemia i Biologia – struktura elektronowa i dynamika molekularna



A. Eilmes, P. Kubisiak: Potencjał elektrostatyczny cieczy jonowej w otoczeniu rozpuszczonej cząsteczki barwnika

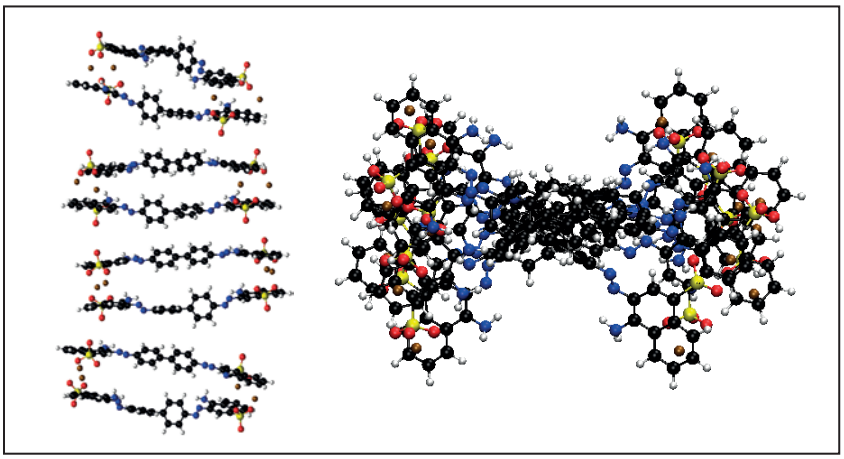
Współczesna chemia obliczeniowa wymaga ciągłego zwiększania zasobów, które umożliwiają przeprowadzenie symulacji dla coraz większych układów (ważnych m.in. w nanotechnologii i naukach biologicznych) oraz poprawę dokładności uzyskanych wyników. Na szczęście, ciągły rozwój technologii w połączeniu z profesjonalnym oprogramowaniem oferowanym przez Cyfronet wychodzi naprzeciw rosnącym wymaganiom i umożliwia przeprowadzanie różnorodnych obliczeń z zakresu chemii kwantowej.

Superkomputery dostępne w naszym Centrum składają się z węzłów obliczeniowych dysponujących nawet 1,5 TB RAM i 48 rdzeniami, co umożliwia wykonywanie obliczeń kwantowochemicznych wymagających ogromnych zasobów pamięci lub dużej liczby rdzeni ze współdzieloną pamięcią operacyjną. Co więcej, wszystkie węzły superkomputera połączone są szybkim interfejsem InfiniBand o wysokiej przepustowości i krótkim czasie dostępu, co przekłada się na zwiększenie szybkości wykonywania obliczeń, jeśli są one rozproszone pomiędzy węzły obliczeniowe. Wiele programów do symulacji chemicznych wymaga również szybkiego dostępu do systemów dyskowych, który jest możliwy dzięki wykorzystaniu rozproszonego systemu plików Lustre.

Wydajność zadań z chemii obliczeniowej zależna jest także od odpowiedniej konfiguracji oprogramowania naukowego oraz umiejętności jego użycia. Nasz zespół administratorów posiada niezbędną wiedzę, umiejętności oraz duże doświadczenie w instalacji oraz efektywnym wykorzystaniu dostępnych aplikacji i narzędzi. W naszym szerokim portfolio można odnaleźć wiele programów umożliwiających różnorodne obliczenia chemiczne. Wśród nich:

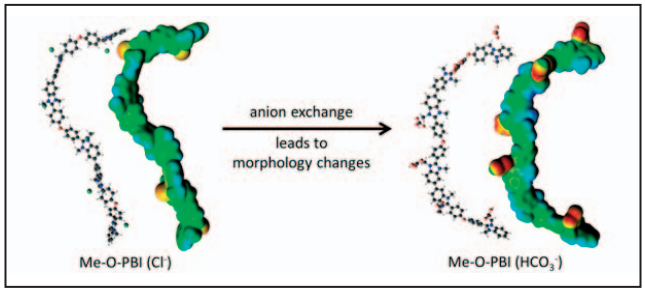
- Wszechstronne i powszechnie używane programy kwantowochemiczne takie jak **Gaussian**, **GAMESS US**, **NWChem**, **Schrödinger**, **Q-Chem, Psi4, ORCA** czy **TURBOMOLE,** dzięki którym można przeprowadzić obliczenia struktury elektronowej oraz wielu właściwości różnorodnych systemów molekularnych wykorzystując metody *ab initio,* pół-empiryczne oraz teorie funkcjonałów gęstości.
- Pakiety **Molpro**, **CFOUR** i **Dalton** do analizy systemów molekularnych z dużą dokładnością, w oparciu o zaawansowane schematy obliczeniowe, takie jak metody sprzężonych klastrów (CCSD(T)) oraz multireferencyjne (MCSCF).
- Pakiet Amsterdam Modeling Suite **(AMS** (dawniej **ADF**)**, DFTB, MOPAC, COSMO-RS)** udostępniający metody do analizy różnorodnych właściwości systemów molekularnych (w szczególności spektroskopowych, takich jak widma NMR czy ESR) poprzez m.in. wykorzystanie relatywistycznego przybliżenia ZORA oraz metody COSMO-RS. Dodatkowo, rozbudowane GUI aplikacji **AMS** (**AMSInput**, **AMSViev**, etc.) ułatwia pracę naszym użytkownikom.

- Znacząca część dostępnych aplikacji może być używana do obliczeń właściwości systemów periodycznych. Wśród nich dostępne są **BAND**, **Quantum ESPRESSO** oraz **SIESTA**.
- Środowisko **AlphaFold** wykorzystujące uczenie maszynowe do analizy struktur geometrycznych białek.
- **Desmond**, **Gromacs**, **Amber**, **LAMMPS**, **NAMD**, **Tinker-HP, CPMD**, **CP2K** i **Terachem** przeznaczone do symulacji wykorzystujących mechanikę molekularną oraz dynamikę molekularną dla systemów zbudowanych nawet z setek tysięcy atomów.



O. Klimas: Obliczona optymalna struktura klastra ośmiu cząsteczek czerwieni Kongo widziana z różnych perspektyw

Wykorzystanie kart graficznych do przeprowadzania obliczeń (**GPGPU**) w wielu dziedzinach nauki pozwala na wielokrotne przyspieszenie wykonywanych zadań. W naszym Centrum część węzłów umożliwia takie obliczenia poprzez użycie jednocześnie nawet ośmiu kart GPGPU (per węzeł) wykorzystujących środowisko programistyczne **CUDA**. Wiele pakietów dostępnych w Cyfronecie umożliwia obliczenia GPGPU, wśród nich oprogramowanie chemiczne: **GAMESS**, **Terachem**, **NAMD**, **Quantum ESPRESSO, Tinker-HP**. Eksperci z ACK Cyfronet AGH współpracują z wieloma grupami tworzącymi i rozwijającymi powyższe pakiety. Warto dodać, że pracownicy Cyfronetu włożyli wiele wysiłku, aby dopracować środowiska obliczeniowe dla naszych użytkowników.



Potencjał elektrostatyczny cząsteczki tworzącej membranę jonowymienną. Opublikowane przez W. Germer, J. Leppin, C. Kirchner, H. Cho, H. Kim, D. Henkensmeier, K. Lee, M. Brela, A. Michalak oraz A. Dyck w Macromol. Mater. Eng. 2015, 300, 497–509

Uczenie maszynowe (ML) i sztuczna inteligencja (AI)

Analiza danych przyspieszona przez sztuczną inteligencję czyni wielkie postępy w wielu dziedzinach badań, w tym w naukach o materiałach, naukach przyrodniczych, językoznawstwie i naukach społecznych. Zdolność sieci neuronowych do uczenia się na podstawie złożonych danych może znacznie poprawić analizę danych, klasyfikację i wykrywanie wzorców, z potencjalnymi zastosowaniami w wielu systemach, m.in. w rozpoznawaniu obrazów, przetwarzaniu języka i optymalizacji.

Centrum superkomputerowe Cyfronet wychodzi naprzeciw tym wyzwaniom i przygotowało kilka pakietów:

PyTorch to pakiet, specjalna biblioteka dla języka Python, oparta na bibliotece Torch. Pozwala na implementację złożonych algorytmów Uczenia Głębokiego z obszaru Przetwarzania Języka Naturalnego, przetwarzania wideo i obrazów oraz wielu innych. Nadaje się bardzo dobrze do modelowania nowych architektur z dziedziny uczenia maszynowego z nastawieniem na eksperymenty.

TensorFlow pozwala podobnie jak Pytorch na implementację modeli bazujących na paradygmacie przepływu tensorów. Ze względu na swój charakter oraz statyczny graf reprezentacji umożliwia wydajną optymalizację treningu oraz inferencji modeli z uwzględnieniem platformy obliczeniowej.

Keras to biblioteka do projektowania modeli neuronalnych. Stanowi zewnętrzne API dla silników bazujących na TensorFlow, Microsoft Cognitive Toolkit, Theano lub PlaidML. Keras został zaprojektowany tak, aby był przyjazny dla użytkownika – jest aplikacją modułową i rozszerzalną.

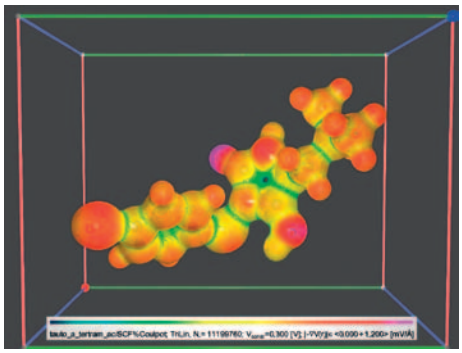
Scikit-learn to biblioteka dla języka Python. Posiada różne algorytmy klasyfikacji, regresji i grupowania, w tym obsługę maszyn wektorowych, losowych lasów, zwiększania gradientu, k-średnich i DBSCAN. Został zaprojektowany do współdziałania z bibliotekami numerycznymi i naukowymi Pythona: NumPy i SciPy.

SchNet to architektura głębokiego uczenia (deep learning), która umożliwia modelowanie molekularne materiałów na poziomie atomowym.

Horovod to zestaw narzędzi dla TensorFlow, Keras, PyTorch i MXNet. Głównym założeniem Horovod jest uczynienie procesu głębokiego uczenia (deep learning) rozproszonym, aby zwiększyć wydajność obliczeń i równocześnie ułatwić korzystanie z niego.

Wizualizacja danych, POVRay/ScPovPlot3D

Wizualizacja danych umożliwia analizę i zrozumienie wyników nawet bardzo złożonych obliczeń, szczególnie, jeśli ich forma finalna jest wielowymiarowa bądź uwzględnia czynnik czasu. Większość aplikacji do obliczeń numerycznych posiada jakiś moduł generujący ich wizualizację. W Python'ie jest dostępny moduł matplotlib czy VTK+, biblioteki graficzne posiadają także Matlab czy R. Podobnie wygląda sytuacja w przypadku programów do geowizualizacji (GIS) czy programów do obliczeń chemicznych. Niestety, niezależnie od tego, jak bardzo programy te są dopracowane, to wynik ich działania jest ograniczony iloczynem kartezjańskim dostępnych (a przy tym zgodnych) opcji.

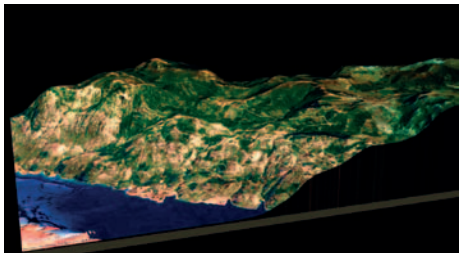


Janusz Opiła: konfiguracja pola elektrostatycznego wokół molekuły polimeru. Ukazano powierzchnię stałego potencjału z aproksymacją trilinearną, kolor koduje moduł natężenia pola elektrostatycznego

Usunięcie tego ograniczenia – choćby tylko w celu stworzenia prototypu stylu wizualizacji do późniejszego wdrożenia w dedykowanym pakiecie – jest możliwe, wymaga jednak skorzystania z programu graficznego ogólnego przeznaczenia. Dobrymi przykładami są tu 3DMax, Blender czy POVRay. Jednak tylko ten ostatni wyposażony jest w domenowy język opisu sceny (SDL) o charakterze skryptowym, który pozwala na programowe, nieinteraktywne tworzenie wizualizacji. Posługiwanie się niezliczoną liczbą opcji języka SDL wymaga dość długiej nauki, dlatego napisano dedykowane API, w formie zestawu specjalizowanych modułów – pakiet ScPovPlot3D. Nie jest to projekt zamknięty, wciąż dopisywane są kolejne rozszerzenia, można zatem mówić o dojrzałej, działającej wersji beta. Aktualnie projekt znajduje się w wersji 4.0 i jest hostowany na GitHub'ie (URL: *https://github.com/JustJanush/Plot3Dv4*) – wieloplatformowe API wymaga POVRay'a w wersji min. 3.7.

Najważniejsze moduły to:

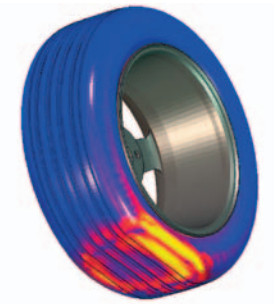
- VectorField.inc – hybrydowa wizualizacja pól wektorowych z użyciem widżetów i/lub rurek strumienia pola,
- Potential.inc – hybrydowa wizualizacja pól skalarnych, na siatkach regularnych i nieregularnych z interpolacją trilinearną albo sześcienną Catmull-Rom'a,
- BPatchSurf.inc – hybrydowa wizualizacja powierzchni w oparciu o dane na siatkach regularnych i nieregularnych z zaimplementowanym prostym krigingiem (KDE),
- Mesh2Surf.inc – hybrydowa wizualizacja danych określonych na regularnych siatkach 2D ($z = f(x, y)$),
- TextExt.inc – rozszerzone formatowanie tekstu 3D, zorientowane na prezentację równań.



Janusz Opiła: Wizualizacja terenu uzyskana w oparciu o dane wysokościowe zebrane na nieregularnej siatce i tekstury pozyskane z aplikacji GoogleEarthPro (okolice Karlobagu, Chorwacja). Opracowanie własne, DOI: 10.23919/MIPRO.2018.8400037

W razie potrzeby developer pakietu służy wsparciem technicznym. Informacje kontaktowe: *https://skos.agh.edu.pl/osoba/janusz-opila-2390.html* .

CAD/CAE



Aplikacje typu CAD (projektowanie wspomagane komputerowo) oraz CAE (komputerowe wspomaganie prac inżynieryjnych) stanowią dziś nieodłączną część procesu projektowania i budowy niemal wszystkiego – od części samochodowych po budynki. Dzięki symulacjom komputerowym inżynierowie mogą dokonywać oceny wytrzymałościowej elementów maszyn i konstrukcji, wykonywać liniowe i nieliniowe analizy strukturalne z uwzględnieniem odkształceń, zjawisk kontaktowych, plastyczności, hipersprężystości, itp. Istotne miejsce w dziedzinie nauk ścisłych zajmują również symulacje dotyczące: spalania, turbulencji, przepływów wielofazowych, reakcji chemicznych, radiacji i mechaniki płynów.

Wszystkie powyższe zagadnienia mogą być badane przez użytkowników dzięki specjalistycznemu oprogramowaniu, np. ANSYS, ABAQUS, FLUENT, MARC czy OPERA.

ANSYS jest rozbudowanym środowiskiem, posiadającym intuicyjny interfejs graficzny, wspierającym naukowców z prawie wszystkich nauk ścisłych oraz przemysłu. Uzyskiwane wyniki charakteryzują się dużą dokładnością. Rezultaty prezentowane są w postaci tekstowej i graficznej, np. diagramy, izopowierzchnie, deformacje. Możliwości obliczeniowe ANSYS-a są bardzo duże, obejmują one m.in. analizę harmoniczną i spektralną, lepkosprężystość, drgania własne, statykę i dynamikę.

ABAQUS znajduje szerokie zastosowanie w przemyśle, a wykonywanie obliczeń opiera się o metodę elementów skończonych. Użytkownik opracowuje kombinacje elementów skończonych, materiałów, procedur analizy i sekwencji obciążeń, zgodnie ze swoimi wymaganiami.

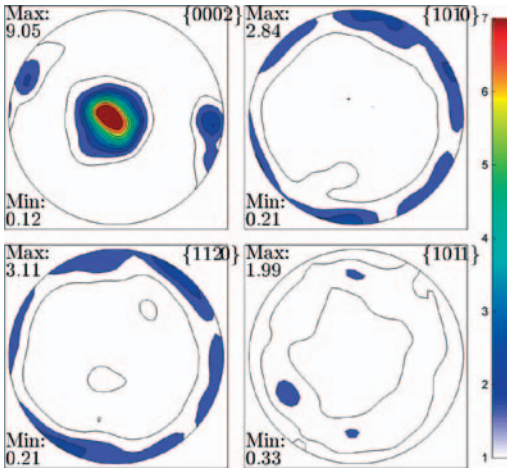
FLUENT umożliwia badanie i symulację zachowania cieczy, co znajduje zastosowanie przy projektowaniu urządzeń wykorzystujących lub transportujących płyny. Może być wykorzystywany w dziedzinie chemii, elektroniki, metalurgii, energetyki cieplnej, biomedycyny i innych.

MARC jest środowiskiem ogólnego zastosowania przeznaczonym do wykonywania obliczeń nieliniowej analizy elementów skończonych w celu wykonania symulacji zachowania produktu przy zmiennych warunkach statycznych, dynamicznych oraz z uwzględnieniem wszystkich rodzajów nieliniowości, w tym geometrycznej, materiałowej oraz warunków brzegowych i kontaktu. Jest to rozwiązanie posiadające możliwość symulacji produkcji i testowania wytworzonego produktu w celu wykrycia zniszczeń, uszkodzeń, awarii oraz propagacji pęknięć. Wszystko to można połączyć z możliwą analizą termiczną, elektryczną, magnetyczną i strukturalną.

OPERA jest oprogramowaniem używającym metody elementów skończonych do projektowania i optymalizacji urządzeń elektromagnetycznych, zarówno w dwóch jak i trzech wymiarach. Pozwala na wyszukiwanie rozwiązań numerycznych dla problemów z wielu dziedzin nauki: elektrostatyki, magnetostatyki, elektromagnetyzmu. Oprogramowanie umożliwia projektowanie różnych urządzeń elektrycznych: transformatorów, silników, przełączników, skanerów MRI i lamp rentgenowskich. To potężne narzędzie do wirtualnego prototypowania, dzięki któremu proces projektowania trwa o wiele krócej.

Aplikacje matematyczne

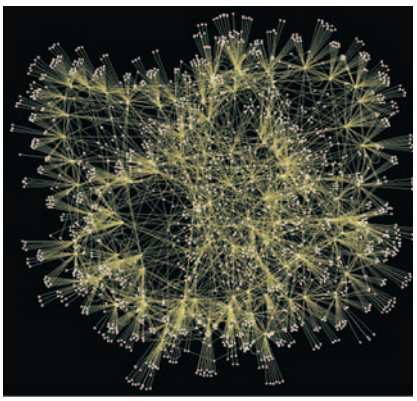
Dzięki aplikacjom matematycznym nawet bardzo rozbudowane i skomplikowane obliczenia mogą zostać wykonane w krótkim czasie. Użytkownicy ACK Cyfronet AGH mają dostęp do programów wspomagających wykonywanie zaawansowanych zadań z algebry, analizy, kombinatoryki, statystyki, teorii liczb, geometrii jak i innych działów matematyki. Wykonywanie działań takich jak całkowanie, różniczkowanie, obliczenia symboliczne, aproksymacja i interpolacja, Transformata Fouriera i Laplace'a, operacje macierzowe, przetwarzanie sygnałów cyfrowych, itp. staje się o wiele łatwiejsze. Przy pomocy odpowiednich narzędzi możliwa jest nawet wizualizacja wyników działań na zbiorach liczb rzeczywistych i/lub zespolonych. Niektóre programy pozwalają dodatkowo na tworzenie wykresów interaktywnych 2D bądź 3D. Wszystkie te funkcjonalności są bardzo istotne, ponieważ w pracy naukowców kluczowe jest często przygotowanie precyzyjnego modelu, jak najdokładniej opisującego analizowane zagadnienia.



Bartosz Sułkowski: Wyniki symulacji tekstury Zn po wytłaczaniu hydrostatycznym w 250 °C

Przykładem środowiska znajdującego zastosowanie w wyżej opisanych zagadnieniach jest **MATLAB**. Dzięki swoim dodatkowym modułom pozwala na wykonywanie obliczeń z zakresu modelowania finansowego, równań różniczkowych cząstkowych (PDE), czy optymalizacji wielokryterialnej, liniowej oraz nieliniowej. Inne zastosowania obejmują m.in. analizę danych geograficznych i generowanie na ich podstawie map, modelowanie biologicznych systemów dynamicznych i systemów hydraulicznych, czy analizę i wizualizację danych genomicznych oraz proteomicznych. Możliwe jest również użycie środowiska symulacyjnego Simulink, służącego do budowania modeli symulacyjnych z bloków, wizualizacji i analizy danych, bez potrzeby tradycyjnego programowania.

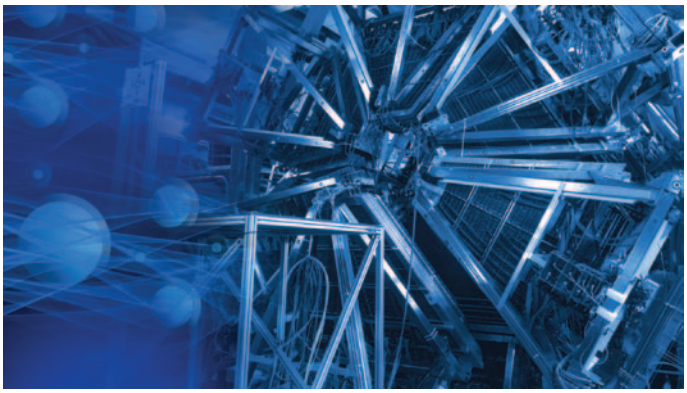
Obok tego rozbudowanego środowiska, użytkownicy znajdą w oprogramowaniu oferowanym przez Centrum także użyteczną aplikację, **MATHEMATICA**, dedykowaną do prowadzenia obliczeń równoległych ze zdefiniowaną dokładnością, do obliczeń symbolicznych i numerycznych. Zaletą tego pakietu jest także posiadanie narzędzia do automatycznego wykrywania i poprawiania błędów oraz możliwość zlecania zadań na kilku komputerach jednocześnie.



Rafał Rak: Sieć połączeń jednominutowych stóp zwrotu dla spółki KGHM (polska spółka giełdowa)

Innym uniwersalnym i interakcyjnym pakietem matematycznym jest **MAPLE**. Stosuje się go m.in. do upraszczania wyrażeń i dokładnej reprezentacji wielkości matematycznych. Oferuje możliwość korzystania z baz danych, generowania kodu innych języków programowania, tworzenia pokazów slajdów zawierających wykonywane zadania wraz z własnymi komentarzami, czy komunikację z aplikacjami MATLAB i systemami CAD. Interesującą aplikacją jest również **SageMATH,** pozwalająca na zaawansowane obliczenia matematyczne z użyciem przyjaznego użytkownikowi środowiska Jupyter Notebook. Pozwala ona na użycie wielu dostępnych bibliotek i narzędzi numerycznych, jak NumPy, SciPy, Sympy, Maxima, GAP, Flint oraz R.

Intensywnie rozwijające się badania naukowe wymagają coraz bardziej zaawansowanych narzędzi. Wśród nich ogromną rolę pełnią narzędzia informatyczne, wspomagające prowadzenie efektywnych badań od momentu ich projektowania, aż po opracowywanie wyników. Cyfronet, poprzez śledzenie najnowszych rozwiązań i tworzenie własnych opracowań, stara się wypełniać ważny obszar swojej misji wsparcia nauki. Na te potrzeby powstały specjalistyczne laboratoria.



Laboratorium Obliczeń Kwantowych

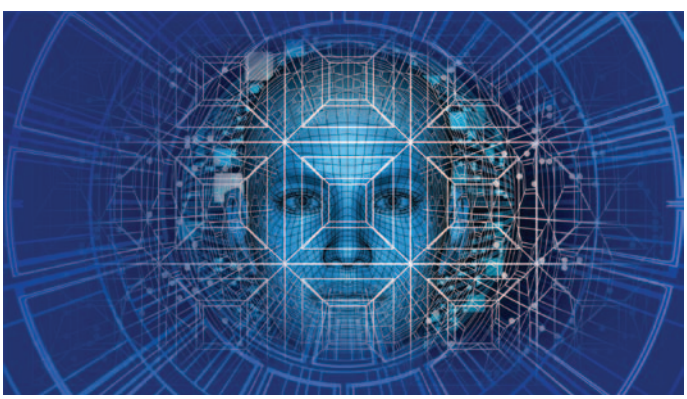
Laboratorium powstało w celu prowadzenia prac badawczych w zakresie wykorzystania komputerów kwantowych w obliczeniach i wsparcia obliczeń klasycznych akceleratorami kwantowymi.

Jednym z kluczowych zadań jest śledzenie rozwoju technologii obliczeń kwantowych oraz dostępnych platform akceleratorów kwantowych celem ich wykorzystania w dedykowanych usługach ACK Cyfronet AGH. Współpracujemy z innymi podmiotami badawczymi i partnerami przemysłowymi, zarówno jako konsorcjanci wspólnych inicjatyw, jak i realizując prace zlecone.

W oparciu o własne kompetencje oraz wymianę wiedzy eksperckiej z siecią partnerów, nasz zespół pracuje nad rozwiązaniem problemów stojących na drodze do szerszego i bardziej efektywnego wykorzystania akceleratorów kwantowych w obliczeniach na rzecz nauki i gospodarki.

Działamy również na rzecz popularyzacji obliczeń wykorzystujących akceleratory kwantowe oraz merytorycznego wsparcia użytkowników. W tym zakresie przygotowujemy niezbędną dokumentację i materiały, prowadzimy szkolenia oraz publikujemy wyniki prowadzonych prac badawczych.

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



Laboratorium Algorytmów Równoległych

Laboratorium skupia swoją działalność na szeroko pojętym aspekcie obliczeniowym równoległych algorytmów ze szczególnym uwzględnieniem algorytmów uczenia maszynowego, a w szczególności możliwości ich efektywnego wykorzystania na dużych klastrach obliczeniowych.

Ze względu na rosnącą z każdym dniem ilość dostępnych danych, tradycyjne algorytmy uczenia maszynowego stają się niewystarczające, zaś szeregowy paradygmat przetwarzania jest obliczeniowo nieefektywny. Aby sprostać nowym wyzwaniom związanym z rosnącą ilością danych, w wielu przypadkach konieczne jest wykorzystanie dużych klastrów obliczeniowych, co wymusza dostosowanie stosowanych algorytmów do pracy w trybie równoległym. W Laboratorium Algorytmów Równoległych rozpatrujemy zarówno teoretyczne, jak i praktyczne aspekty związane z tym zadaniem. W szczególności skupiamy się na następujących dziedzinach: wizji komputerowej, obliczeniach tensorowych, sieciach głębokich, przetwarzaniu obrazów o niskiej jakości, rozpoznawaniu obrazów podwodnych, klasyfikacji danych hiperspektralnych, klasyfikacji danych histopatologicznych, niezbalansowaniu danych.

W obrębie rozpatrywanych dziedzin rozwijamy nowe algorytmy wykorzystujące obliczenia w trybie równoległym, w szczególności badamy aspekty teoretyczne oraz praktyczne związane z tym zjawiskiem.

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

Laboratorium Metod Informatycznych w Medycynie

Główne zadania Laboratorium skupiają się w dwóch sferach. Pierwszą z nich jest działalność badawcza, obejmująca gruntowną analizę i weryfikację dostępnych oraz potencjalnych odpowiedzi na wyzwania znajdujące się na styku medycyny i technologii informatycznych. Druga obejmuje projektowanie, opracowywanie i późniejszą eksploatację dedykowanych aplikacji i platform dla zastosowań w medycynie. Ten zakres obejmuje również monitorowanie stanu bezpieczeństwa opracowywanego oprogramowania oraz mechanizmów składowania i przetwarzania danych.

Dzięki kompleksowemu ujęciu procesów: od rozpoznania problemu badawczego, poprzez analizę potrzeb użytkowników, aż po ostateczne wdrożenie, Laboratorium skutecznie realizuje swoją misję wsparcia środowiska naukowego i medycznego. W ramach upowszechniania wiedzy eksperckiej, członkowie Zespołu publikują wyniki badań w czasopismach naukowych, biorą udział w przygotowywaniu materiałów informacyjnych i prowadzą konsultacje dla użytkowników.

Pracownicy Laboratorium nawiązują współpracę z renomowanymi krajowymi i zagranicznymi instytutami badawczymi i ośrodkami informatyki medycznej. Efektami są m.in. trwające oraz już zrealizowane projekty z istotnym udziałem członków Zespołu:

- Sano: Centrum Doskonałości Nowych Metod Diagnostyki Obliczeniowej i Spersonalizowanej Terapii,
- PRIMAGE: Wykorzystanie metod symulacji komputerowej i analizy obrazów w spersonalizowanej diagnostyce raka,
- Virolab: A Virtual Laboratory for Decision Support in Viral Disease Treatment,
- Gliomed: Diagnostyka glejaków na podstawie wolnokrążącego DNA guza,

- Eurvalve: Wspomaganie spersonalizowanej terapii dla przypadków choroby zastawek serca (Personalised Decision Support for Heart Valve Disease),
- CECM: Centrum Doskonałości Nowych Metod Diagnostyki Obliczeniowej i Spersonalizowanej Terapii (A Centre for New Methods in Computational Diagnostics and Personalised Therapy).

Wcześniejsza działalność obecnego zespołu Laboratorium jest przedstawiona szczegółowo na stronie: *http://dice.cyfronet.pl*.

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

Laboratorium Przetwarzania Danych

Laboratorium projektuje i wdraża dedykowane aplikacje i platformy software'owe dla zastosowań w różnych dziedzinach nauki. Laboratorium tworzą specjaliści w zakresie architektur oprogramowania, programowania Front-end i Back-end, projektowania interfejsów użytkownika i user experience, DevOps, testowania, analityki wymagań. Kierunkami specjalizacji zespołu są przede wszystkim:

- opracowywanie nowatorskich metod pozyskiwania wiedzy z dostępnych danych,
- rozwój technologii wspierających przetwarzanie danych otwartych,
- integracja systemów przetwarzania danych i wiedzy z istniejącymi repozytoriami i e-infrastrukturami.

Laboratorium nawiązuje współpracę z renomowanymi jednostkami naukowymi i badawczymi w ramach polskich i międzynarodowych projektów. Efektami współpracy są między innymi:

- **Budowa Portalu Sat4Envi** (*https://dane.sat4envi.imgw.pl*), udostępniającego dane satelitarne z programu Copernicus. Portal umożliwia wyszukiwanie, oglądanie, zamawianie i pobieranie danych satelitarnych oraz ich produktów pochodnych z poziomu przeglądarki internetowej.
- **Opracowanie i utrzymanie Portalu EOSC** (*https://eosc-portal.eu/*) w ramach serii projektów związanych z European Open Science Cloud (EOSC). Portal zapewnia dostęp do zasobów wielu europejskich e-infrastruktur i infrastruktur badawczych za pomocą zunifikowanego systemu uwierzytelniania użytkownika. Działania EOSC skupiają się wokół realizacji paradygmatu Open Science (otwartej nauki).

Zespół Laboratorium był również zaangażowany w tworzenie **Portalu PLGrid** (*https://portal.plgrid.pl*), udostępniającego naukowcom wiele pakietów oprogramowania, bibliotek i narzędzi naukowych. W 2022 roku z LPD wyodrębniono Laboratorium Interdyscyplinarnych Obliczeń Naukowych.

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

Laboratorium Interdyscyplinarnych Obliczeń Naukowych

Laboratorium Interdyscyplinarnych Obliczeń Naukowych (LION) prowadzi prace badawcze i rozwojowe dotyczące procesów prowadzenia i wspomagania obliczeń naukowych oraz organizacji danych naukowych. Tworzone przez pracowników Laboratorium portale Science Gateways stwarzają możliwość nawiązywania współpracy z podmiotami zewnętrznymi – polskimi i zagranicznymi.

Zadania LION ukierunkowane są także na:

- realizację grantów naukowych i zleceń przemysłowych,
- wykorzystywanie wiedzy eksperckiej do rozwiązywania problemów wymagających użycia różnych zasobów obliczeniowych, np. technologii uczenia maszynowego,
- popularyzację obliczeń wykorzystujących narzędzia tworzone przez Laboratorium wśród użytkowników,
- wsparcie merytoryczne użytkowników, m.in. poprzez monitorowanie potrzeb, opracowanie dokumentacji, prowadzenie szkoleń.

Zespół LION, współtworząc wcześniej Laboratorium Przetwarzania Danych, opracował autorskie środowisko programistyczne InSilicoLab, obejmujące zestaw zaawansowanych narzędzi i bibliotek programistycznych pozwalających na budowę i rozwój dedykowanych portali naukowych. Portale bazujące na InSilicoLab są zaprojektowane w taki sposób, aby gromadzić w jednym miejscu całość narzędzi, których badacze potrzebują dla obliczeń *in silico*. Zaletami są przede wszystkim:

- łatwe uruchamianie nawet skomplikowanych, długich i wymagających wielu obliczeń, eksperymentów użytkownika,
- możliwość wygodnego opisu, kategoryzacji i wyszukiwania danych wejściowych lub wynikowych.

Technologię InSilicoLab wyróżnia dążenie do jak największej użyteczności budowanych za pomocą środowiska narzędzi. Ta sfera obejmuje zarówno użyteczność dla rozwiązywania problemów naukowych w danej dziedzinie, jak również przyjazność portalu dla jego końcowego odbiorcy.

Efektami współpracy LIONu z renomowanymi jednostkami naukowymi i badawczymi w ramach polskich i międzynarodowych projektów jest między innymi rozwój **EPISODES Platform** (*https://tcs.ah-epos.eu/*) w ramach serii projektów związanych z European Plate Observing System - EPOS (*https://www.epos-eu.org/*). Portal oraz zorganizowane wokół niego narzędzia są zorientowane na badanie i analizę sejsmiczności oraz innych zjawisk spowodowanych działalnością człowieka (np. eksploatacja zasobów w ramach kopalni, tworzenie sztucznych zbiorników wodnych). Portal jest zintegrowany z europejską infrastrukturą EPOS.

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

Laboratorium Technologii Chmurowych

Laboratorium zajmuje się projektowaniem i eksploatacją chmury dla nauki, a także narzędzi służących efektywnemu korzystaniu z niej. Zespół opracowuje kompleksowe środowiska dostępu do rozproszonych danych, uwzględniające zarówno kwestie bezpiecznego składowania i przetwarzania danych w chmurze, jak i wygodne interfejsy dostępowe (portale, aplikacje) dla użytkownika końcowego.

Mając na uwadze dynamiczny rozwój nowych technologii przetwarzania i składowania danych w chmurze, Laboratorium w trybie ciągłym prowadzi prace badawcze i publikuje wyniki. Korzystając z wiedzy eksperckiej zespołu, aktywnie wspiera inicjatywy naukowe, w tym międzynarodowe projekty i e-infrastruktury.

Flagowym produktem Laboratorium jest **Onedata**: globalnie skalowalny system zarządzania danymi, ujednolicający dostęp do danych przechowywanych w rozproszonych systemach. Onedata równie dobrze odpowiada na potrzeby zarówno małych grup użytkowników, jak i dużych, międzynarodowych środowisk naukowych. System umożliwia użytkownikom korzystanie z jednorodnego systemu zarządzania danymi zarówno do przechowywania danych osobistych jak i związanych z pracą zawodową, np. wyników badań, oraz daje możliwość dostępu do nich w wydajny sposób z dowolnego urządzenia.

Onedata jest aktualnie uruchomione w ramach kilku infrastruktur europejskich, w tym w projektach: PLGrid, INDIGO-DataCloud, EGI Data Hub oraz Human Brain Project. W projekcie HBP Onedata wykazało możliwość dostarczenia systemu zarządzania danymi pozwalającemu na spełnienie wymagań przepustowości dla aplikacji wizualizacji przekrojów mózgu w czasie rzeczywistym.

Więcej informacji na stronie: *https://onedata.org*.

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

Laboratorium Zastosowań Technik Obliczeniowych

Kluczowym aspektem działania Laboratorium jest opracowywanie modeli numerycznych oraz symulacji komputerowych w oparciu o architektury superkomputerowe, ze szczególnym uwzględnieniem wykorzystania praktycznego w przemyśle. Laboratorium prowadzi własne badania dotyczące technik obliczeniowych, jak również śledzi i analizuje nowe rozwiązania w ich zastosowaniu.

Zespół Laboratorium nawiązuje współpracę w ramach polskich i międzynarodowych konsorcjów, wspólnie z partnerami realizując projekty ukierunkowane na opracowanie lub udoskonalenie procesów produkcji wyrobów metalowych. Wśród najważniejszych efektów tych działań znajdują się:

- **VirtROLL – Virtual strip rolling mill**: głównym celem projektu było stworzenie systemu komputerowego wspomagającego elastyczne projektowanie technologii walcowania wyrobów płaskich na podstawie wyników obliczeń dedykowanych modułów do symulacji numerycznych.
- **PROTEUS-RS – Strategie optymalizacji procesów wytwarzania produktów długich w celu podniesienia jakości wyrobu gotowego poprzez minimalizację naprężeń rezydualnych**: projekt przewiduje wykonanie modeli oraz szeregu symulacji numerycznych w oparciu o architektury komputerowe wysokiej wydajności na potrzeby projektowania procesu wytwarzania metalowych elementów długich.

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

Laboratorium Technik Wizualnych

Laboratorium Technik Wizualnych (Media Lab 5.0) jest studiem filmowo-telewizyjnym, którego głównym zadaniem jest produkcja materiałów dokumentujących i promujących osiągnięcia ACK Cyfronet AGH.

Realizowane filmy i programy telewizyjne popularno-naukowe i szkoleniowe pokazują efekty prac badawczych użytkowników infrastruktury obliczeniowej Cyfronetu.

W Laboratorium prowadzone są warsztaty dziennikarskie, sztuki operatorskiej, realizacji dźwięku, sztuki montażu, grafiki komputerowej i obsługi studia telewizyjnego. To również miejsce prowadzenia prac badawczych z technologii wizyjnych, akustyki oraz technik transmisji on-line.

Laboratorium to studio filmowo-telewizyjne o powierzchni około 100 m^2 wyposażone w ruszt z oświetleniem, nagłośnione, ze sztucznym horyzontem i scenografią. Studio połączone jest z salą audytoryjną na około 50 miejsc. Studio skomunikowane jest z reżyserkami: WIZJI i ŚWIATŁA (ze stanowiskami: regulacji torów kamerowych, realizacji wizji, nagrywania i odtwarzania, grafiki komputerowej, światła), DŹWIĘKU z konsolą Yamaha LS 9 oraz LEKTORKĄ. W Laboratorium znajdują się dwie MONTAŻOWNIE z zestawami komputerowymi z oprogramowaniem Adobe CS6. Laboratorium posiada kamery Panasonic AG-HPX 371 (3 szt.) oraz inny sprzęt filmowo-telewizyjny do realizacji obrazu, światła i dźwięku w produkcjach studyjnych i plenerowych.

Studio współpracuje z profesjonalnymi twórcami i realizatorami oraz studentami krakowskich uczelni.

Filmy zrealizowane przez MediaLab 5.0 można obejrzeć między innymi:

- na kanale ACK Cyfronet AGH w serwisie YouTube (*https://www.youtube.com/user/CyfronetAGH/videos*),
- na platformie Pionier.tv (*https://pionier.tv/wideo-tag/cyfronet/*).

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

Laboratorium Akceleracji Obliczeń i Sztucznej Inteligencji

Nasza misja

Dzięki kompetencjom w dziedzinie algorytmów sztucznej inteligencji (SI) i znajomości nowoczesnych technik obliczeniowych wspieramy naukowców w ich pracach badawczych. Nasza wiedza i doświadczenie pozwala na skuteczną realizację uczenia maszynowego przy wykorzystaniu dostępnej w ACK Cyfronet AGH, dedykowanej do obliczeń neuronowych, partycji klastra komputerowego Prometheus. Partycja do obliczeń SI zbudowana została w oparciu o cztery wydajne serwery wyposażone w osiem kart GPGPU nVidia Volta V100 każdy. Całkowita moc obliczeniowa partycji to ponad 4 PetaFlops, czyli ponad cztery biliardy ($4x10^{15}$) obliczeń dla SI na sekundę.

Specjalizacje domenowe

Poniżej przedstawiono wybrane zastosowania sztucznej inteligencji w przetwarzaniu języka naturalnego, przetwarzaniu obrazów i analizie szeregów czasowych w różnych problemach badawczych i zastosowaniach praktycznych.

Przetwarzanie języka naturalnego

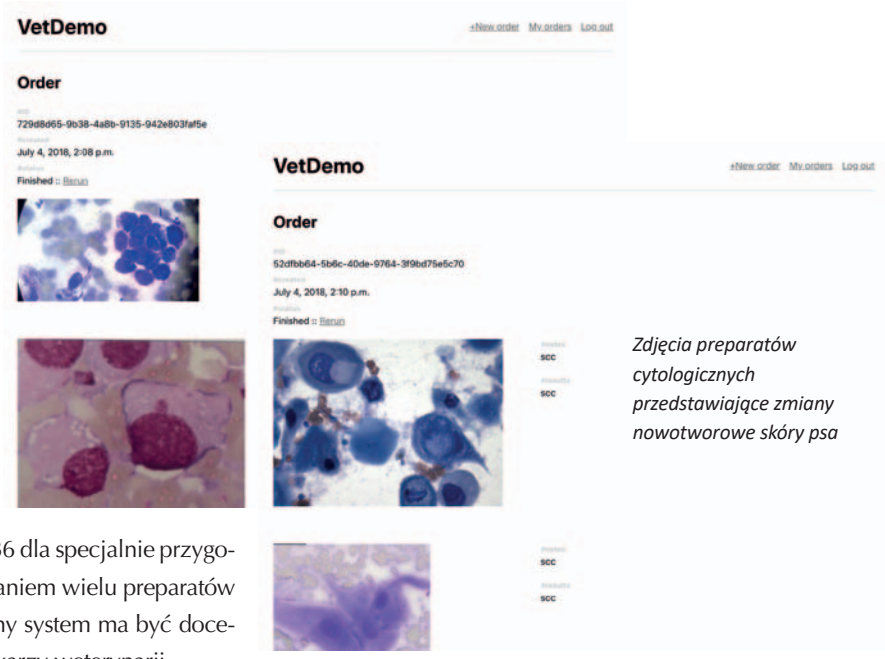
Przetwarzanie języka naturalnego to zagadnienie, w którym Laboratorium posiada wieloletnie doświadczenie. Od wielu lat tworzymy i rozwijamy narzędzie umożliwiające przeszukiwanie, porównywanie oraz klasyfikację dokumentów tekstowych. Praktycznym owocem tych prac jest usługa Scholar, dostępna w ACK Cyfronet AGH na platformie usług PLGrid. Jednym z ważnych zagadnień podejmowanych badawczo jest analiza efektywności metod redukcji dokładności reprezentacji danych tekstowych i ich wpływu na wyniki działania algorytmów. Udało nam się opracować modyfikacje pozwalające na 10-krotne zmniejszenie zużycia energii obliczeń w stosunku do oryginalnej implementacji.

Pojawienie się rozwiązań opartych o sieci neuronowe zrewolucjonizowało dziedzinę. Zajmujemy się badaniami nad możliwością kompresji i implementacji sprzętowej sieci do oceny sentymentu, które pokazały, że możliwa jest redukcja dokładności reprezentacji wag sieci do 8 a nawet 4 bitów przy zachowaniu niemal niezmienionej skuteczności działania sieci. Dodatkowo zajmujemy się zagadnieniami semi-nadzorowanego uczenia, w których ilość dostępnych, oznaczonych danych była bardzo ograniczona, a docelowe kategorie zmieniały się w trakcie działania systemu. Nasze badania pokazały, że możliwe jest opracowanie rozwiązania osiągającego dokładność nawet 98.9%.

Przetwarzanie obrazów

Przetwarzanie obrazów przy pomocy sieci neuronowych skupia się wokół prac Laboratorium, które związane są z rozpoznawaniem i detekcją obiektów na potrzeby aplikacji medycznych. W ich trakcie opracowany został system do klasyfikacji zmian nowotworowych w pobranych od

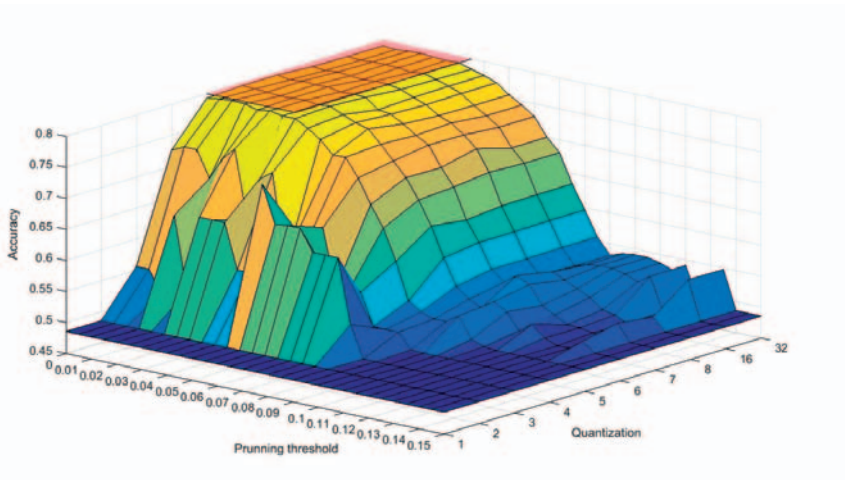
zwierząt próbkach z badania cytologicznego. Zaproponowany system osiągnął skuteczność około 96% dla wybranych trzech typów nowotworów. W pracach zostały wykorzystane modele głębokich sieci typu Resnet-50 oraz Resnet-152 i opracowany został specjalny schemat trenowania z wykorzystaniem algorytmów genetycznych dedykowany do klasyfikacji, który pozwala na selektywne wybieranie obszaru percepcji dla operacji klasyfikacji. Jednym z najnowszych osiągnięć w tym obszarze jest model do detekcji bazujący na architekturze Yolo3, który osiąga bardzo dobrą wartość współczynnika mAP, równą 0.86 dla specjalnie przygotowanych obrazów wykonanych z wykorzystaniem wielu preparatów cytologicznych niskiej jakości. Opracowywany system ma być docelowo wykorzystywany w codziennej pracy lekarzy weterynarii.



Zdjęcia preparatów cytologicznych przedstawiające zmiany nowotworowe skóry psa

Analiza szeregów czasowych

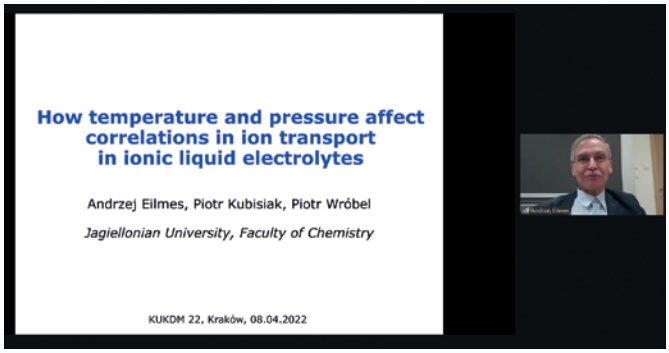
Prace Laboratorium dotyczą także modelowania szeregów czasowych służącego detekcji sytuacji nietypowych. Szczegółowe prace obejmują aplikacje praktyczne, takie jak predykcja anomalii w celu ochrony przed katastrofalnym uszkodzeniem magnesów Wielkiego Zderzacza Hadronów (LHC) oraz innych urządzeń stowarzyszonych znajdujących się w CERN. Opracowany system do detekcji quenchy w magnesach nadprzewodzących wykorzystuje sieci neuronowe GRU oraz LSTM oraz dedykowany post-processing, który pozwolił dodatkowo na klasyfikację wykrytych anomalii. Co więcej, opracowany algorytm został zrealizowany sprzętowo w architekturze FPGA Xilinx Zynq UltraScale+ MPSoC 285, co pozwoliło na uzyskanie bardzo niskiej latencji odpowiedzi.



W obszarze medycyny podejmowany jest problem detekcji zasłabnięć u osób hospitalizowanych, które długi czas są unieruchomione w łóżku szpitalnym. Analizy dotyczą znanego problemu zaniku funkcji wspomagania układu krążenia przez pracę mięśni nóg i są realizowane we współpracy z Medical University of Graz.

Konferencja Użytkowników Komputerów Dużej Mocy (KU KDM)

Podstawowym założeniem konferencji KU KDM było zainicjowanie corocznych spotkań naukowych gromadzących użytkowników, którzy wykonują obliczenia z wykorzystaniem komputerów dużej mocy i oprogramowania, udostępnianych przez ACK Cyfronet AGH.



Po raz pierwszy konferencja została zorganizowana w 2008 roku, a jej program obejmował liczne wystąpienia naukowców – prezentujące wyniki prac badawczych uzyskanych przy wykorzystaniu zasobów sprzętowych i oprogramowania oferowanego w Cyfronecie. Konferencję uświetniły wykłady zaproszonych gości – Norberta Attiga z Jülich Supercomputing Centre i Jaapa A. Kaandorpa z University of Amsterdam. Pierwsza edycja konferencji spotkała się z dużym uznaniem wśród jej uczestników. Tego typu wydarzenie było bardzo oczekiwane i potrzebne.

Obecna tematyka konferencji KU KDM koncentruje się na obliczeniach i symulacjach dużej skali, nowych algorytmach w informatyce, narzędziach i technikach w systemach dużej mocy obliczeniowej, zastosowaniach informatyki w dydaktyce oraz bazach danych. Jednakże jej głównym celem jest możliwość prezentacji osiągnięć naukowych szerokiemu gronu badaczy z różnych dziedzin. Jest to również okazja do przybliżenia użytkownikom samego Centrum oraz jego zasobów, w tym infrastruktury PLGrid.

Użytkownicy, którzy na co dzień prowadzą badania z wykorzystaniem zasobów Cyfronetu, uzyskują możliwość przedstawienia roli tych zasobów, typowych scenariuszy użycia i aspektów wydajnościowych w trakcie swoich prezentacji. Uczestniczący w konferencji przedstawiciele Cyfronetu, dzięki sposobności spotkania się z tymi naukowcami, prowadzenia rozmów i konsultacji, pozyskują wiedzę niezbędną do podjęcia działań zmierzających do dostosowania infrastruktury obliczeniowej do potrzeb wymagających środowisk naukowych. Atmosfera sprzyja również wymianie doświadczeń pomiędzy naukowcami reprezentującymi te same lub różne dyscypliny oraz zapoznaniu się z nowymi technologiami oraz domenowo-specyficznymi usługami wdrażanymi przez Centrum.

Do ważnych części konferencji należą spotkania z dostawcami sprzętu i oprogramowania dla Cyfronetu, a także dyskusja panelowa na temat efektywnego wykorzystania tych zasobów. Ta ostatnia cieszy się zawsze bardzo dużym zainteresowaniem użytkowników – naukowców, którzy mogą wtedy uzyskać informacje o nowościach w infrastrukturze obliczeniowej Centrum, a także bezpośrednio poinformować ekspertów z Cyfronetu o problemach, jakie napotkali w trakcie pracy z tą infrastrukturą.

W roku 2022, z powodu trwającej pandemii koronawirusa (COVID-19), konferencja została zorganizowana online.

Artykuły naukowe opracowane na podstawie najlepszych prezentacji konferencyjnych są publikowane w jednym z dwóch znanych czasopism naukowych z dziedziny IT: Computing and Informatics (CAI) (*http://www.cai.sk)* lub Computer Science (CSCI) (*www.csci.agh.edu.pl*).

https://www.cyfronet.pl/kukdm22

Sat4Envi – kosmiczne technologie, powszechne wykorzystanie

Dzięki udostępnieniu danych satelitarnych z programu Copernicus, projekt Sat4Envi otworzył nowy rozdział dla podejmowania optymalnych decyzji dotyczących planowania przestrzennego i szybkiego reagowania w przypadku wystąpienia zagrożeń środowiskowych.

Dane satelitarne na wyciągnięcie ręki

Pod adresem *dane.sat4envi.imgw.pl* dostępne jest kluczowe dokonanie projektu – System Obsługi Klienta, czyli obsługiwane z poziomu przeglądarki internetowej narzędzie dostępu do hurtowni danych satelitarnych Sat4Envi. Dzięki tej platformie użytkownik końcowy zyskuje możliwość wyszukiwania, oglądania, zamawiania i pobierania danych satelitarnych oraz ich produktów pochodnych. System służy przede wszystkim do monitorowania powierzchni ziemi, monitorowania atmosfery i pogody oraz rozpoznawania zagrożeń meteorologicznych. Dostęp do wszystkich danych jest całkowicie darmowy i nie zależy od afiliacji, czy wykonywanej profesji. W wersji podstawowej nie wymaga rejestracji. Po zalogowaniu użytkownik ma możliwość skorzystania z dodatkowych funkcjonalności, jak chociażby zapisywanie „koszyka" najczęściej wybieranych pomiarów, czy dostęp do metadanych, jakimi opisane są prezentowane obrazy satelitarne. Dzięki nałożeniu warstw WMS (Web Map Service), zawierających m.in. granice jednostek administracyjnych, narzędzie jest funkcjonalne dla jednostek administracji publicznej czy instytucji regionalnych. Samo działanie Systemu Obsługi Klienta jest intuicyjne i pozwala na szybkie dotarcie do pożądanych typów danych.

Projekt i partnerzy

System operacyjnego gromadzenia, udostępniania i promocji cyfrowej informacji satelitarnej o środowisku (Sat4Envi) miał na celu szerokie i wygodne udostępnienie satelitarnych danych programu Copernicus oraz danych pochodzących z innych satelitów środowiskowych i meteorologicznych.

Za sukcesem Sat4Envi stoi zaangażowanie wielu specjalistów z czterech ściśle współpracujących jednostek. Liderem był Instytut Meteorologii i Gospodarki Wodnej – Państwowy Instytut Badawczy, a partnerami w projekcie: Centrum Badań Kosmicznych Polskiej Akademii Nauk, Polska Agencja Kosmiczna oraz Akademickie Centrum Komputerowe CYFRONET AGH. IMGW-PIB odpowiadał przede wszystkim za rozbudowę stacji odbioru i przetwarzania danych z satelitów okołobiegunowych (stacja powstała w Krakowie) oraz pełnił rolę koordynatora. Wspólny wysiłek pozwolił sprawnie realizować pozostałe założenia projektu, którymi były:

- budowa nowoczesnego archiwum danych satelitarnych (IMGW-PIB, CYFRONET),
- centrum szkoleniowe nowych technologii satelitarnych (CBK, PAK, IMGW-PIB),
- centrum udostępniania informacji naukowej dla użytkowników (CYFRONET, IMGW-PIB, CBK).

Projekt, o łącznej wartości 17 903 900 zł, był współfinansowany ze środków Europejskiego Funduszu Rozwoju Regionalnego w ramach Programu Operacyjnego Polska Cyfrowa.

Technologiczne wyzwania

Budowa portalu z interfejsem dostępowym wymagała bardzo starannego i wieloaspektowego projektowania, uwzględniającego łatwość wyszukiwania danych czy szybkość odnajdywania określonych funkcji. Przyjmując za cel stworzenie produktu o wysokiej funkcjonalności również dla użytkowników „niezaawansowanych", zespół programistyczny rozpoczął działania od rozbudowanej oceny wymagań. Analizowano przypadki użycia w różnych grupach użytkowników docelowych i przeprowadzano wielokrotne weryfikacje z użyciem makiet graficznych i prototypowych implementacji. W pracach nad interfejsem specjaliści w zakresie user experience sięgnęli również do techniki wywiadów pogłębionych w grupach reprezentantów przyszłych użytkowników (jak IMGW, CBK, PAK, Pracownia Teledetekcji Morza PAN czy Pieniński Park Narodowy).

- *Rolą Cyfronetu w projekcie była realizacja zadań programistycznych związanych z budową i utrzymaniem Systemu Obsługi Klienta, ale też zapewnienie infrastruktury informatycznej niezbędnej do gromadzenia i udostępniania danych. Cyfronet jest operatorem połączenia internetowego z IMGW-PIB Kraków, a w ramach infrastruktury PLGrid udostępnia zasoby obliczeniowe niezbędne dla prawidłowego funkcjonowania Sat4Envi. Z kolei do przetwarzania danych i tworzenia informacji z surowych danych satelitarnych wykorzystywany jest zainstalowany w Cyfronecie superkomputer Prometheus* – podkreśla prof. Kazimierz Wiatr, Dyrektor ACK Cyfronet AGH.

Sat4Envi wspiera w sytuacjach kryzysowych

Analiza danych satelitarnych z jak najmniejszymi opóźnieniami w stosunku do ich pozyskania może mieć kluczowe znaczenie dla prognozowania, monitorowania oraz przeciwdziałania skutkom kataklizmów. W szczególności mówimy tu o monitorowaniu powodzi, ocenie skali pożarów, czy monitorowaniu szkód w uprawach rolnych powstałych w wyniku suszy lub przemarzania. Ze względu na ważność tych działań i konieczność szybkiego reagowania, projekt Sat4Envi uwzględnia wsparcie dla specjalnych grup użytkowników, takich jak Centrum Informacji Kryzysowej.

Więcej informacji na temat projektu znajduje się na stronie: *https://sat4envi.imgw.pl/*.

Prometheus i jego infrastruktura służą udostępnianiu danych z programu Copernicus. Źródło: Sat4Envi

EPOS - multidyscyplinarna platforma wsparcia badań w zakresie nauk o Ziemi

Efektywne monitorowanie procesów zachodzących we wnętrzu Ziemi oraz ich skutków odczuwalnych na powierzchni wymaga wykorzystania różnorakich narzędzi i metodologii. EPOS – European Plate Observing System (System Obserwacji Płyty Europejskiej) ma w tym zakresie dostarczyć badaczom kompleksowych rozwiązań, przekraczających zarówno granice państw, jak i pojedynczych dyscyplin naukowych.

W październiku 2018 roku EPOS otrzymał status ERIC: European Research Infrastructure Consortium (Europejskiego Konsorcjum Infrastruktury Badawczej). W ten sposób stworzone zostały ramy prawne i organizacyjne pozwalające na występowanie na arenie międzynarodowej już nie jako projekt współrealizowany przez wielu partnerów, a jako odrębny podmiot z główną siedzibą w Rzymie.

Ideą stojącą za EPOS jest lepsze poznanie funkcjonowania Ziemi jako złożonego systemu, w którym z jednej strony zdarzenia naturalne (jak erupcje wulkanów, powodzie czy trzęsienia ziemi) wpływają na społeczeństwo i gospodarkę, a z drugiej, działalność człowieka zmienia środowisko. Z tego względu do efektywnych badań potrzebne jest podejście wieloaspektowe. Poprzez integrację europejskich infrastruktur badawczych w ponadnarodowy system udostępniający dane, ich produkty, oprogramowanie oraz usługi, EPOS opracowuje dostęp do możliwości, które do tej pory występowały wyłącznie w formie rozproszonej. Inicjatywa skupia się na dziesięciu głównych, powiązanych ze sobą węzłach tematycznych (Thematic Core Services – TCS): Dane sejsmologiczne, Obserwatoria przy uskokach tektonicznych, Dane i produkty GNSS (Global Navigation Satellite Systems), Obserwacje wulkanów, Dane satelitarne, Badania geomagnetyzmu, Zagrożenia antropogeniczne, Laboratoria Wielkoskalowe, Stanowiska badań geoenergetycznych związane z energią niskowęglową, Tsunami.

Źródło: EPOS - European Plate Observing System

Obliczeniowe nauki o Ziemi

Udostępnienie infrastruktury badawczej jako platformy internetowej wymaga ścisłej współpracy specjalistów w zakresie technologii informatycznych z przedstawicielami nauk o Ziemi. Synergiczne działania są realizowane na wielu poziomach – począwszy od prowadzonych przez naukowców pomiarów i eksperymentów generujących wielką ilość surowych danych, które są składowane i przetwarzane w systemach informatycznych – między innymi w ramach infrastruktury Cyfronetu. Współpraca jest też bardzo istotna przy budowie narzędzi służących do wizualizacji danych, sterowanych z poziomu przeglądarki internetowej za pomocą interfejsów graficznych (GUI). Docelowo, EPOS ma połączyć środowiska badawcze w ramach jednej, multidyscyplinarnej platformy, dostarczyć efektywnych i bezpiecznych narzędzi dostępowych i analitycznych oraz zainicjować powstawanie nowych usług w oparciu o dotychczasowe osiągnięcia. Co ważne, dane geofizyczne i geologiczne, podobnie jak zaawansowane oprogramowanie do ich wizualizacji i analizy, będą dostępne w sposób otwarty i bezpłatny. Dzięki temu EPOS wspiera interdyscyplinarne badania dotyczące przyczyn i skutków procesów zachodzących

wewnątrz i na powierzchni Ziemi. Ma to istotne znaczenie w świetle monitorowania zagrożeń środowiskowych, takich jak: trzęsienia ziemi, powodzie, osuwiska czy wybuchy wulkanów, jak również zagrożeń antropogenicznych związanych m.in. z działalnością kopalni. Poprzez kompleksową analizę zjawisk już zaobserwowanych, można lepiej przygotować się na zdarzenia, które jeszcze nie miały miejsca, starając się jak najbardziej ograniczyć ich negatywne skutki.

Obecnie EPOS wszedł w fazę operacyjną infrastruktury budowanej w ramach projektów EPOS PP i EPOS-IP. Platforma badawcza jest dostępna pod adresem: *https://www.ics-c.epos-eu.org/,* a portal integrujący informacje o działaniach EPOS pod adresem: *https://www.epos-eu.org/*.

Działania Cyfronetu na rzecz EPOS

ACK Cyfronet AGH uczestniczy w kolejnych projektach budujących EPOS, wspierając inicjatywę zarówno zasobami sprzętowymi, jak i wiedzą oraz umiejętnościami specjalistów. Równolegle prowadzone działania obejmują kilka sfer, jedną z których jest opracowanie platformy cyfrowej EPISODES dla węzła tematycznego Zagrożeń Antropogenicznych (TCS AH).

Platforma udostępniona pod adresem: *https://tcs.ah-epos.eu/* stanowi narzędzie do analizy sejsmiczności antropogenicznej i związanych z nią zagrożeń oraz do oceny potencjalnego wpływu eksploatacji geozasobów na środowisko. Dzięki danym pochodzącym ze stacji sejsmicznych oraz udostępnionym informacjom przemysłowym, platforma pozwala łatwiej analizować takie procesy jak: zalewanie sztucznych zbiorników, wydobycie surowców, gazów łupkowych czy wód podziemnych. Oprócz działań związanych z udostępnianiem i wizualizacją danych, specjaliści z Cyfronetu dbają o utrzymanie elementów integracji pomiędzy węzłem tematycznym TCS AH a portalem nadrzędnym EPOS ISC-C, który umożliwia dostęp do efektów działań we wszystkich węzłach tematycznych EPOS. Dodatkowo, Cyfronet zapewnia wirtualną przestrzeń roboczą do uruchamiania aplikacji od różnych dostawców oraz porządkowania danych. Obecnie platforma TCS AH liczy ponad 1500 zarejestrowanych użytkowników.

Z kolei w ramach projektu EPOS SP (EPOS Sustainability Phase) Cyfronet jest zaangażowany w działania mające na celu pozyskanie nowych użytkowników infrastruktury EPOS, w tym partnerów przemysłowych.

Równolegle trwają prace związane z integracją kolejnych pakietów oprogramowania wykorzystywanego w badaniach naukowych.

Dodatkowo, uczestnicząc w projektach krajowych, Cyfronet działa na rzecz rozbudowy polskiej Infrastruktury Badawczej EPOS-PL. Najnowsze prace, skupione w projekcie EPOS PL+, dotyczą opracowania oprogramowania wspierającego planowanie i prowadzenie badań z użyciem metod sztucznej inteligencji na danych IB EPOS-PL. Platforma EPOS-AI będzie umożliwiała łatwe korzystanie z danych zintegrowanych w programie, dobór metod uczenia maszynowego oraz uruchamianie zasobochłonnych etapów przetwarzania na zasobach obliczeniowych polskich Centrów Komputerów Dużej Mocy.

Sano: Centrum Zindywidualizowanej Medycyny Obliczeniowej – Międzynarodowa Fundacja Badawcza

W latach 2019-2026 Akademickie Centrum Komputerowe Cyfronet AGH wraz z pięcioma partnerami realizuje wyjątkowy projekt w ramach programu *Teaming for Excellence* Horyzontu 2020 Unii Europejskiej, programu Międzynarodowych Agend Badawczych prowadzonego przez Fundację na rzecz Nauki Polskiej oraz przy wsparciu finansowym Ministerstwa Nauki i Szkolnictwa Wyższego. W wyniku tego projektu w Krakowie powstało **Sano** – Centrum Zindywidualizowanej Medycyny Obliczeniowej.

Misja **Sano** obejmuje:

- rozwój nowych metod obliczeniowych, algorytmów, modeli i technologii związanych z medycyną spersonalizowaną,
- wprowadzenie do codziennej praktyki opieki zdrowotnej nowych rozwiązań diagnostycznych i terapeutycznych opartych na symulacjach komputerowych,
- stymulowanie powstawania i rozwoju przedsiębiorstw tworzących technologie umożliwiające wprowadzanie nowych metod diagnostycznych i terapeutycznych,
- wkład w nowe programy kształcenia dla potrzeb nowoczesnej, zindywidualizowanej medycyny.

Centrum Sano (*https://sano.science/*) zlokalizowane jest w Krakowie, gdzie kształci się licznych wysokiej klasy specjalistów w zakresie nauk medycznych i informatyki. Co ważne, krakowskie szpitale badawcze są wysoko oceniane w środowisku naukowym i systematycznie rośnie w tym mieście liczba firm w sektorze *life science*.

Rozwój **Centrum Sano** przyczyni się bezpośrednio do ulepszenia poziomu nauki w naszym regionie, poprzez nowe kontakty naukowe oraz możliwość kształcenia doktorantów na światowym poziomie, a ponadto wzmocni transfer wiedzy i technologii do gospodarki sprzyjając powstawaniu nowych firm działających w obszarze zaawansowanych technologii. Oddziaływanie Centrum będzie miało charakter ponadregionalny, czy wręcz międzynarodowy, przyczyniając się do rozwoju nowoczesnej medycyny, a w konsekwencji do wzrostu poziomu opieki zdrowotnej.

Program działania Centrum jest oparty m.in. na Polskiej Strategii Inteligentnej Specjalizacji i ma na celu rozwój współpracy nauki i biznesu w skali międzynarodowej. Kluczowymi wskaźnikami sukcesu są: liczba wysoko cytowanych publikacji naukowych i grantów uzyskanych przez Centrum, liczba rozwiązań opartych na modelach obliczeniowych wprowadzonych do praktyk klinicznych oraz liczba innowacyjnych rynkowych produktów i usług.

Centrum Zindywidualizowanej Medycyny Obliczeniowej Sano powstało dzięki międzynarodowej współpracy następujących instytucji: ACK Cyfronet AGH, Klaster LifeScience Kraków – Krajowy Klaster Kluczowy, University of Sheffield i Insigneo Institute, Forschungszentrum Jülich, Fraunhofer Institute for Systems and Innovation Research ISI oraz Narodowe Centrum Badań i Rozwoju.

EOSC – ponadnarodowa integracja zasobów naukowych

European Open Science Cloud (EOSC) jest inicjatywą zmierzającą do stworzenia wirtualnego środowiska wypełniającego założenia paradygmatu Otwartej Nauki. EOSC ma w łatwy i przejrzysty sposób udostępnić nie tylko dane naukowe, ale też zaawansowane narzędzia i zasoby do ich składowania, udostępniania, przetwarzania i zarządzania nimi.

W ramach EOSC tworzone są połączenia pomiędzy istniejącymi e-infrastrukturami badawczymi i naukowymi. Integracja następuje między innymi poprzez ujednolicenie zasad dostępu i autoryzacji dla badaczy z różnych państw. Dzięki osiągnięciom projektu EOSC-Hub stworzono platformę realizującą powyższe zadanie. EOSC-Portal stanowi miejsce styku pomiędzy dostawcami usług i zasobów naukowych oraz badaczami, którzy mogą skorzystać nie tylko z samych zasobów, ale też wsparcia technicznego i szkoleń. Zespół ACK Cyfronet AGH odegrał w tym zakresie kluczową rolę, zostając dostawcą portalu oraz rozwijając serwis Marketplace, z rozbudowanym katalogiem usług udostępnianych przez partnerów EOSC.

Dalszy rozwój EOSC

Kluczowymi założeniami dalszego rozwoju EOSC jest rozbudowa Portalu oraz integracja usług od większej ilości dostawców.

Realizowany w latach 2019-2021 projekt EOSC Enhance (*https://www.eosc-portal.eu/enhance*) miał na celu usprawnienie Portalu pod kątem wygody i szybkości użytkowania. Prowadzone przez naszych specjalistów prace dotyczyły między innymi zaawansowanej analityki zachowań użytkowników w celu stworzenia i wdrożenia najlepszych praktyk *user experience*. Jednocześnie udostępniono nowe funkcjonalności.

Do końca października 2022 roku trwał projekt EOSC Synergy (*https://www.eosc-synergy.eu/*), mający na celu wdrożenie standardów EOSC dla kolejnych 9-ciu krajowych e-infrastruktur. Cyfronet, oprócz koordynacji działań na poziomie narodowym, wspierał proces planistycznie, szukając nowych, skutecznych rozwiązań dla integracji w innych krajach.

Z kolei EOSC Future jest kontynuacją i rozszerzeniem dotychczas prowadzonych prac. Projekt ma podwyższyć jakość ekosystemu EOSC dla jeszcze lepszego wsparcia europejskiej nauki oraz zachęcić naukowców do korzystania z udostępnianych zasobów. Ten cel jest realizowany poprzez dalsze prace nad Portalem, integrację kolejnych infrastruktur i inicjatyw naukowych. Dodatkowo, EOSC Future ma zapewnić długotrwałą stabilność EOSC. W ramach projektu Cyfronet odpowiada za rozwój jednego z trzech głównych komponentów Portalu: części skierowanej do użytkownika, w szczególności za jego elementy wspierające funkcjonalności Marketplace.

Najnowsza z serii inicjatyw budujących EOSC, w których uczestniczy Cyfronet, to projekt FAIRCORE4EOSC, który zakłada rozbudowanie infrastruktury EOSC o komponenty zapewniające wsparcie dla paradygmatu FAIR. Cyfronet jest odpowiedzialny za wsparcie eksperckie w zakresie zarządzania danymi i spersonalizowanego wyszukiwania zasobów EOSC (w tym obiektów naukowych) przy użyciu metod AI (przede wszystkim Machine Learning) oraz wdrożenie rozwiązań wypracowanych w projekcie w Portalu użytkownika EOSC.

EuroHPC – europejska infrastruktura przetwarzania danych

Europejskie Wspólne Przedsięwzięcie w dziedzinie Obliczeń Wielkiej Skali (European High-Performance Computing Joint Undertaking – EuroHPC JU) zostało powołane w celu dynamicznego rozwoju istniejącej europejskiej infrastruktury komputerów dużej mocy i tym samym udostępnienia europejskim naukowcom mocy obliczeniowych porównywalnych z tymi dostępnymi w USA, Chinach i Japonii. EuroHPC JU zrzesza 33 państwa, w tym Polskę, która dołączyła w 2018 roku, oraz prywatnych członków. Głównym celem jest zakup i rozmieszczenie dwóch superkomputerów o mocy w skali eksaflopsów (EFlops = 10^{18} operacji zmiennoprzecinkowych na sekundę), które znajdą się w światowej TOP5 (pierwszej piątce listy TOP500).

Osiągnięcie eksaskali zaplanowano na rok 2023. Wcześniej jednak uruchamiane są kolejne superkomputery o mniejszej mocy, tworzące infrastrukturę, która pozwala na przeskalowanie technologii i oprogramowania. Do czasu opracowania niniejszej publikacji, EuroHPC JU dostarczyło 2 superkomputery przed-eksaskalowe: LUMI (Finlandia) oraz Leonardo (Włochy). Trzecia maszyna tej skali (Mare Nostrum 5) zostanie uruchomiona w Hiszpanii. Dodatkowo zainstalowano lub pomyślnie uruchomiono operacyjnie 5 superkomputerów petaskalowych: Discoverer (Bułgaria), Karolina (Czechy), MeluXina (Luksemburg), Vega (Słowenia) oraz Deucalion (Portugalia).

W czerwcu 2022 roku EuroHPC JU wytypowało pięć kolejnych miejsc, w których zostaną zainstalowane nowe superkomputery. Są to: Grecja, Irlandia, Niemcy, Polska i Węgry. Wśród nich, w Centrum Superkomputerowym Jülich, znajdzie się JUPITER – pierwszy europejski komputer o mocy obliczeniowej powyżej 1 Eksaflopsa (10^{18} operacji zmiennoprzecinkowych na sekundę), który w momencie pełnej instalacji stanie się najszybszą maszyną europejską. W Polsce już pod koniec 2023 roku powstanie system skali mid-range o mocy kilkukrotnie większej niż najszybszy obecnie superkomputer w Polsce (Athena). Operatorem nowego superkomputera będzie Akademickie Centrum Komputerowe CYFRONET AGH w Krakowie, a dostęp dla polskich naukowców zostanie zapewniony w ramach infrastruktury PLGrid.

Więcej o EuroHPC JU na stronie: *https://eurohpc-ju.europa.eu/*.

Superkomputer Karolina, Czechy. Fot. EuroHPC JU

Superkomputer Vega, Słowenia. Fot. Atos

EuroHPC PL – polski element EuroHPC JU

Na początku 2021 roku rozpoczęła się realizacja projektu „Narodowa Infrastruktura Superkomputerowa dla EuroHPC – EuroHPC PL", którego celem jest zapewnienie nowoczesnej infrastruktury, w szczególności zasobów superkomputerów, specjalizowanych akceleratorów i usług laboratoriów na potrzeby środowiska akademickiego, podmiotów gospodarki i społeczeństwa w Polsce.

Projekt jest realizowany przez Konsorcjum w składzie:

- Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie – Akademickie Centrum Komputerowe Cyfronet – Lider Konsorcjum,
- Instytut Chemii Bioorganicznej Polskiej Akademii Nauk – Poznańskie Centrum Superkomputerowo-Sieciowe,
- Politechnika Gdańska – Centrum Informatyczne Trójmiejskiej Akademickiej Sieci Komputerowej,
- Politechnika Wrocławska – Wrocławskie Centrum Sieciowo-Superkomputerowe,
- Narodowe Centrum Badań Jądrowych,
- Instytut Informatyki Teoretycznej i Stosowanej Polskiej Akademii Nauk,
- Centrum Fizyki Teoretycznej Polskiej Akademii Nauk.

EuroHPC PL ma ułatwić i poszerzyć dostęp do wielkoskalowego przetwarzania danych zarówno dla naukowców, jak i dla przedsiębiorców i w ten sposób zwiększyć konkurencyjność polskich podmiotów na tle Europy i świata. Podstawowym elementem budowanej infrastruktury jest moc obliczeniowa dostarczana przez superkomputery. Ich zastosowanie jest składnikiem niezbędnym do prowadzenia badań naukowych w wielu dziedzinach i znacząco przyspiesza oraz obniża koszty procesów badawczych. Jest to szczególnie widoczne na początkowym etapie procesu badawczego, w którym zbiór możliwych do rozważenia kierunków badań jest tak duży, że ich klasyczne sprawdzenie (np. w laboratoriach) nie byłoby możliwe zarówno ze względu na czasochłonność jak i koszty. Zasoby superkomputerowe zostaną wsparte architekturą algorytmów kwantowych i akceleratorów neuromorficznych. Dzięki temu znacznie zwiększą się możliwości analizy i wizualizacji danych. Dodatkowo projekt zakłada szeroki zakres wsparcia użytkownika i szkolenia. Zadania w ramach projektu realizowane są w ramach czterech laboratoriów.

Głównym celem Laboratorium modelowania i równoległego przetwarzania danych w środowisku przedeksaskalowym jest zaprojektowanie i wdrożenie zaawansowanych systemów obliczeniowych dużej skali. Laboratorium zastosowań obliczeń hybrydowych adresuje potrzeby związane z zastosowaniami i wdrożeniem do praktycznego użycia komputerów kwantowych. Z kolei zadania Laboratorium zastosowania superkomputerów w medycynie skupiają się nad praktycznym zastosowaniem dużych mocy obliczeniowych w aplikacjach medycznych. Laboratorium efektywności energetycznej i obliczeniowej oprogramowania HPC pracuje nad rozwojem technik mających na celu redukcję kosztów operacyjnych maszyny, takich jak zużycie energii i zapotrzebowanie na chłodzenie, a także nad optymalizacją oprogramowania działającego w środowisku superkomputerowym.

Więcej informacji na temat EuroHPC PL znajduje się na stronie: *https://eurohpc.pl*.

LUMI

Konsorcjum LUMI

LUMI – nr 1 w Europie, nr 3 na świecie

W czerwcu 2022, w czasie targów ISC High Performance w Hamburgu ogłoszono kolejną listę TOP500 superkomputerów o najwyższej mocy obliczeniowej świata. Zainstalowany w centrum danych CSC w Finlandii superkomputer LUMI zajął na niej trzecie miejsce, jednocześnie obejmując prowadzenie w Europie. Ten niewątpliwy sukces w środowisku HPC (High-Performance Computing) jest efektem współpracy 10 państw, w tym Polski, które zawiązały konsorcjum działające pod skrzydłami większej inicjatywy, jaką jest EuroHPC Joint Undertaking. Co ważne, dzięki działaniom koordynowanym przez ACK Cyfronet AGH i MEiN, Polacy uzyskali już dostęp do zasobów LUMI.

LUMI – światowa czołówka mocy i wydajności energetycznej

Wspólnie z pozostałymi systemami petaskalowymi oraz przed-eksaskalowymi, w których budowę zaangażowane jest EuroHPC JU, LUMI stanowi element pan-europejskiej infrastruktury, do której zostaną włączone w najbliższych latach również w pełni eksaskalowe komputery. Uzyskanie miejsca w TOP3 superkomputerów to wielki sukces dla państw europejskich, ponieważ pokazuje możliwość konkurowania z systemami z USA, Chin czy Japonii. Warto podkreślić, że do momentu przeprowadzenia testów High Performance Linpack (HPL), w których LUMI wykazało realną moc obliczeniową 151,9 PFlops, zainstalowano zaledwie część partycji GPU. Po pełnej instalacji, wydajność wzrośnie do 375 PFlops, a teoretyczna wydajność nawet do 550 PFlops. Takie osiągi stały się możliwe dzięki architekturze sprzętowej LUMI:

- partycja GPU jest złożona z 2560 węzłów, z których każdy zawiera 64-rdzeniowy procesor AMD Trento i cztery karty AMD MI250X,
- pojedyncza karta MI250X może dostarczyć 42,2 TFlop/s wydajności w testach High Performance Linpack,
- każdy węzeł GPU zawiera cztery karty połączeń sieciowych 200 Gb/s,
- dodatkiem do partycji GPU jest partycja CPU, wykorzystująca 64-rdzeniowe procesory ogólnego przeznaczenia AMD EPYC™ trzeciej generacji.

Jednak przy projektowaniu i budowie infrastruktury LUMI położono mocny nacisk nie tylko na moc obliczeniową, ale też na efektywne wykorzystanie energii. Z tego względu superkomputer LUMI zajął również trzecie miejsce na liście Green500 – superkomputerów o najwyższej wydajności energetycznej, liczonej jako stosunek mocy obliczeniowej do pobieranej energii elektrycznej. Dodatkowo, co warto podkreślić, zasilanie LUMI pochodzi w 100% ze źródeł odnawialnych (elektrownie wodne), a ciepło odpadowe odprowadzane z maszyny jest wykorzystywane do ogrzewania okolicznych budynków. W ten sposób LUMI jest w tej chwili flagowym europejskim przykładem wprowadzania idei zrównoważonego rozwoju do budowy systemów superkomputerowych.

Więcej o superkomputerze LUMI na stronie: *https://lumi-supercomputer.eu*.

Przedstawiciele Cyfronetu w czasie inauguracji LUMI. Od lewej: Mariusz Sterzel, kierownik Laboratorium Obliczeń Kwantowych i Marek Magryś, Zastępca Dyrektora ACK Cyfronet AGH ds. Komputerów Dużej Mocy

Dostęp do LUMI dla polskich naukowców

ACK Cyfronet AGH koordynuje prace związane z udostępnieniem polskim naukowcom zasobów oferowanych przez LUMI. Celem jest zapewnienie wydajnego i przyjaznego dla użytkownika środowiska pracy z superkomputerem, na bazie wypracowanych w ramach infrastruktury PLGrid dobrych praktyk.

W okresie od jesieni 2021 roku do wiosny 2022 roku Cyfronet zorganizował dwa konkursy na pilotażowe granty obliczeniowe na superkomputerze LUMI dla polskich naukowców. Projekty fazy pilotażowej miały na celu przetestowanie, kolejno partycji CPU (z procesorami ogólnego przeznaczenia) oraz fazy GPU (z procesorami graficznymi). Polska, jako członek konsorcjum, zgłosiła łącznie 5 projektów o potencjalnie dużym znaczeniu, wymagających użycia dużej ilości rdzeni obliczeniowych oraz potężnych zasobów dyskowych. Dostęp jest realizowany za pośrednictwem Portalu PLGrid (*https://portal.plgrid.pl)*, a dodatkowe informacje znajdują się na stronie *https://cyfronet.pl/lumi*.

Cyfronet liderem przedsięwzięć wpisanych na Polską Mapę Infrastruktury Badawczej

Wśród strategicznych infrastruktur wpisanych w styczniu 2020 r. na Polską Mapę Infrastruktury Badawczej znajdują się dwa przedsięwzięcia zgłoszone przez ACK Cyfronet AGH jako inicjatora i koordynatora konsorcjum PLGrid: *Narodowa Infrastruktura Superkomputerowa dla EuroHPC* oraz *Narodowa Infrastruktura Chmurowa PLGrid dla EOSC*.*

Celem programu **Narodowa Infrastruktura Superkomputerowa dla EuroHPC** jest budowa infrastruktury obliczeniowej dla badań naukowych dotyczących rozwiązań odpowiadających na aktualne i przyszłe potrzeby polskiego społeczeństwa, środowiska naukowego i gospodarki.

Infrastruktura będzie opierać się o nowoczesne systemy superkomputerowe umożliwiające realizację zarówno tradycyjnych zadań symulacyjnych, jak i analizę danych z wykorzystaniem metod sztucznej inteligencji. Produkcyjne systemy obliczeniowe budowane w ramach projektu będą należeć do czołówki światowych superkomputerów. Poza infrastrukturą obliczeniową projekt zapewni też dostęp do specjalistycznych szkoleń oraz eksperckiego wsparcia technicznego dla użytkowników z nauki i gospodarki, a także niezbędne procedury alokacji i rozliczania wykorzystania zasobów.

Projekt wpisuje się bezpośrednio w ramy międzynarodowej inicjatywy EuroHPC – European High-Performance Computing opublikowanej w komunikacie Komisji Europejskiej nr COM/2018/08 final – 2018/03 (NLE). EuroHPC jest przedsięwzięciem mającym na celu utworzenie europejskiego systemu komputerów dużej mocy klasy eksaskalowej, unikatowych na skalę światową, zbudowanych w oparciu o technologie rozwijane w Europie. Przedsięwzięcie będzie realizowane przez Konsorcjum PLGrid.

W efekcie realizacji przedsięwzięcia Narodowa Infrastruktura Superkomputerowa dla EuroHPC powstanie infrastruktura do prowadzenia badań na potrzeby nauki, gospodarki i społeczeństwa czerpiąca z najnowocześniejszych technologii w zakresie HPC wypracowanych w ramach międzynarodowej kooperacji EuroHPC. Infrastruktura będzie oferowała usługi w zakresie masywnie równoległych symulacji komputerowych, wysoce wydajnego przetwarzania zbiorów danych, wykorzystania metod sztucznej inteligencji, oprogramowania i narzędzi wysokiej produktywności, w tym wizualizacji danych, oraz wsparcia i szkolenia użytkowników. Poza głównymi systemami obliczeniowymi w skład infrastruktury wejdą także mniejsze systemy testowe i badawcze, służące weryfikacji nowych technologii procesorowych, akceleratorowych, pamięciowych i sieciowych w kontekście wykorzystania ich do budowy systemów produkcyjnych, a także prowadzeniu prac badawczych i rozwojowych w zakresie efektywnych infrastruktur HPC.

Usługi będą oferowane za pośrednictwem infrastruktury PLGrid, która integruje większość dostępnych w kraju zasobów obliczeniowych, co ułatwi proces alokacji zasobów i wsparcia użytkowników. Zrealizowana zostanie także integracja z europejskimi systemami preeskaskalowymi i petaskalowymi udostępnionymi w ramach programu Euro-HPC, w szczególności z maszyną budowaną przez konsorcjum LUMI, którego Polska jest członkiem.

Narodowa Infrastruktura Chmurowa PLGrid dla EOSC to program wykorzystania zasobów chmurowych do badań naukowych odpowiadających na aktualne i przyszłe potrzeby polskiego społeczeństwa, środowiska naukowego i gospodarki. Zakres tych badań obejmuje m.in. dane, infrastruktury i platformy przetwarzania danych oraz efektywne algorytmy i dedykowane aplikacje.

Podstawą programu są wymagania społeczeństwa, gospodarki i polskich badaczy, w szczególności współpracujących w ramach międzynarodowych zespołów naukowych. Grupy te wymagają zaawansowanych środowisk integracji rozproszonych zasobów: oprogramowania, infrastruktur i dedykowanych usług. Wymagania te mogą być spełnione jedynie przez zaawansowane technologie IT, połączone z zasobami obliczeniowymi, pamięci masowych i danych. Współdziałanie tych wszystkich elementów w ramach elastycznego ekosystemu umożliwiają technologie chmurowe.

Narodowa Infrastruktura Chmurowa PLGrid dla EOSC jest częścią ekosystemu Europejskiej Chmury dla Otwartej Nauki (EOSC, Deklaracja z dnia 26.10.2017). Polska rozwija obecnie dwa kluczowe komponenty tego federacyjnego, globalnie dostępnego i multidyscyplinarnego środowiska: Onedata – system jednolitego udostępniania i zarządzania danymi oraz Portal EOSC. Jako część europejskiego ekosystemu Narodowa Infrastruktura Chmurowa PLGrid będzie oferować zaufane i otwarte środowiska dla użytkowników w całym cyklu życia danych. Pozwoli to na publikację, wyszukiwanie, wykorzystywanie i wielokrotne użycie zgromadzonych danych, narzędzi, oprogramowania i innych wyników prac przez naukowców, gospodarkę i całe społeczeństwo.

Badania zaplanowane w ramach Narodowej Infrastruktury Chmurowej PLGrid dla EOSC pozwolą na opracowanie, walidację i w konsekwencji udostępnienie usług (ogólnych i dedykowanych), a tym samym wykorzystanie nowoczesnych technologii i efektywnych technik zarządzania, przetwarzania i ponownego wykorzystania danych przez środowiska naukowe, podmioty gospodarki i społeczeństwo.

W tym celu Infrastruktura będzie dostarczać:

- technologie rozwiązań dla środowisk rozproszonych, w tym chmurowych, obejmujące zarządzanie usługami w środowisku rozproszonym, automatyzację złożonych procesów, integrację platform i infrastruktur badawczych,
- oparte o krajowe technologie bezpieczne udostępnianie danych i zarządzanie nimi, zgodne ze standardami EOSC,
- badania specyficznych potrzeb użytkowników w zakresie wielkoskalowego przetwarzania danych w środowisku rozproszonym, w tym: przetwarzanie „blisko danych", wykorzystanie paradygmatu „data lakes" wraz z nowymi modelami analizy, skalowalne zasoby w środowisku rozproszonym,
- weryfikacje rozwiązań przygotowywanych na potrzeby społeczeństwa, nauki i gospodarki w zaawansowanych aplikacjach,
- katalog ogólnych serwisów i usług chmurowych dla gospodarki, nauki i społeczeństwa będących wynikiem prac badawczych i rozwojowych we współpracy z EOSC.

Pierwszy etap tych prac jest obecnie realizowany w ramach zasobów infrastruktury PLGrid.

**Materiały z broszury MNiSW pt. „Polska Mapa Infrastruktury Badawczej".*

Głównym celem projektu jest dostarczenie otwartego i w pełni operacyjnego ekosystemu FAIR EOSC. Projekt ma za zadanie rozbudowanie infrastruktury EOSC o komponenty zapewniające wsparcie dla paradygmatu FAIR (Findability, Accessibility, Interoperability, Reusability).

Celem projektu jest zdefiniowanie oraz opracowanie nowego i innowacyjnego programu europejskich studiów magisterskich, skupiających się na HPC. W założeniu program wyposaży studentów w kompetencje i wiedzę dotyczącą HPC, wymaganą przez środowisko akademickie i przemysł.

W ramach projektu PROTEUS-RS przewiduje się wykonanie modeli oraz szeregu symulacji numerycznych w oparciu o architektury komputerowe wysokiej wydajności na potrzeby projektowania procesu wytwarzania metalowych elementów długich.

Poprzez rozwój Portalu EOSC projekt ma doprowadzić do rozszerzenia ekosystemu EOSC, zintegrować istniejące infrastruktury i inicjatywy naukowe oraz zaangażować nowe dziedzinowe i narodowe społeczności naukowe.

Projekt zakłada dostarczenie transparentnego i zunifikowanego dostępu do danych z obserwacji Ziemi, głównie satelitarnych, w celu wspierania produkcji żywności oraz przemysłu rolno-spożywczego.

Celem projektu jest utworzenie w Krakowie ośrodka medycyny obliczeniowej. Centrum będzie głównym motorem europejskiego postępu w tym szybko rozwijającym się sektorze, opracowując zaawansowane metody inżynieryjne do zapobiegania, diagnozowania i leczenia chorób oraz zaspokajając ogólnoświatową potrzebę radykalnie usprawnionych systemów opieki zdrowotnej.

Celem projektu jest rozbudowa infrastruktury obliczeniowej umożliwiającej współpracę w obszarze problemów o dużej intensywności łączących duże wolumeny danych przekraczających granice jednego centrum danych.

Celem projektu EPOS-SP jest opracowanie i wdrożenie założeń pozwalających na zapewnienie trwałości infrastruktury EPOS wytworzonej w poprzednich projektach (EPOS-PP oraz EPOS-IP).

Projekt EOSC-Synergy wprowadził standardy EOSC dla infrastruktur krajowych w dziewięciu krajach Unii Europejskiej. Stało się to poprzez zharmonizowanie polityk oraz rozszerzenie dostępu do infrastruktur badawczych, danych naukowych i serwisów dziedzinowych.

Głównym zadaniem projektu PRACE-6IP była implementacja nowych rozwiązań oraz utrzymanie operacyjności środowiska PRACE w obszarze europejskich infrastruktur obliczeniowych HPC.

W ramach projektu stworzony zostanie system wspierania decyzji klinicznych (Clinical Decision Support System – CDSS) dla leczenia nowotworów (nauroblastoma, glejak) występujących u dzieci.

Głównym celem projektu EuroCC było stworzenie europejskiego systemu wsparcia technologicznego opartego o ściśle powiązane ze sobą Narodowe Centra Kompetencji w poszczególnych krajach europejskich, co pozwoliło na odniesienie korzyści środowisku akademickiemu, sektorowi przedsiębiorstw (zwłaszcza MŚP) oraz administracji publicznej z dostępnej wiedzy specjalistycznej, doświadczenia i zasobów EuroHPC.

Projekt EOSC Enhance miał na celu zbudowanie ulepszonej, bardziej zintegrowanej wersji Portalu EOSC, co umożliwiło poprawę i rozszerzenie rozwiązań ułatwiających odszukiwanie europejskich usług naukowych oraz zbiorów danych otwartej nauki.

Celem projektu było przygotowanie uruchomienia produkcyjnej infrastruktury dla otwartej nauki w Europie i praktyczne zastosowanie rozwiązań wypracowanych w ramach projektu EOSC-Pilot do realnego środowiska wielkiej skali rozproszonego na większość krajów europejskich.

Celem projektu jest budowa specjalistycznej infrastruktury ogólnego przeznaczenia na potrzeby obliczeń wielkoskalowych umożliwiającej podejmowanie wyzwań badawczych w kluczowych z punktu widzenia polskiego społeczeństwa, środowiska naukowego i gospodarki obszarach. Projekt jest polskim etapem rozwoju programu EuroHPC.

Zadaniem projektu jest opracowanie i dostarczenie produkcyjnych usług przechowywania, dostępu oraz zabezpieczania danych i zarządzania metadanymi, a także integracji rozwiązań dla przetwarzania dużych i złożonych wolumenów danych na bazie rozproszonej e-infrastruktury.

Projekt zakłada budowę unikalnych laboratoriów badawczych w oparciu o krajową sieć światłowodową PIONIER. Głównym celem projektu jest zbudowanie i udostępnienie platform dla jednostek naukowych, przedsiębiorców i innych podmiotów zainteresowanych prowadzeniem badań naukowych oraz prac rozwojowych w oparciu o nową, ogólnopolską infrastrukturę badawczą.

Celem bezpośrednim projektu jest budowa specjalistycznej e-infrastruktury przetwarzania danych, umożliwiającej optymalne wykorzystanie usług specjalistycznych oraz nowej generacji służących do stymulowania nowych obszarów zastosowań w nauce, gospodarce, edukacji oraz życiu społecznym.

W ramach projektu zostaną zwiększone funkcjonalności infrastruktury badawczej EPOS-PL. Powstanie nowe Centrum Infrastruktury Badawczej, nowy poligon pomiarowy oraz WNiP: „Platforma IT do badań metodami Sztucznej Inteligencji (EPOS-AI)".

Zadaniem projektu jest wytworzenie usług infrastruktury obliczeniowej oraz składowania danych na potrzeby projektu PRACE w ramach 6-ciu dedykowanych laboratoriów: 1) L. przetwarzania HPC oraz chmurowego, 2) L. dostępu do infrastruktury przetwarzania, 3) L. zarządzania i monitorowania usług, 4) L. usług zarządzania danymi, 5) L. zarządzania rozproszonymi danymi oraz transparentnego dostępu do danych, 6) L. bezpieczeństwa infrastruktury.

Głównym zadaniem projektu było udostępnienie pochodzących z satelitów Sentinel sieci Copernicus danych satelitarnych. W ramach projektu powstała infrastruktura do automatycznego pobierania danych bezpośrednio z satelitów, ich bezpiecznego przechowywania oraz udostępniania na potrzeby nauki, administracji i szkoleń.

Zadaniem projektu była budowa krajowej infrastruktury badawczej z zakresu nauk o Ziemi oraz jej integracja z międzynarodowymi bazami danych, serwisami i usługami realizowanymi w ramach europejskiego programu EPOS (European Plate Observing System).

Celem projektu był rozwój wyspecjalizowanego centrum kompetencji technologicznych w zakresie rozproszonych infrastruktur obliczeniowych, ze szczególnym uwzględnieniem technologii gridowych, obliczeń w chmurze oraz infrastruktur wspierających obliczenia na wielkich zbiorach danych. W rezultacie, użytkownikom zostały zaoferowane wielkie moce obliczeniowe i duże zasoby pamięci. Infrastruktura PLGrid została wzbogacona o nowe usługi bazowe i usługi dla użytkowników końcowych.

W trakcie projektu miała miejsce integracja wybranych usług dostępnych w sieci PIONIER i opracowano nowe usługi, w zakresie zwiększania niezawodności i bezpieczeństwa.

Celem projektu była rozbudowa 21 środowiskowych sieci teleinformatycznych nauki dla zapewnienia instytucjom naukowym dostępu do nowoczesnej oraz bezpiecznej infrastruktury sieciowej i umożliwiającej łączność z jednostkami naukowymi całego świata, wspierając polskie grupy naukowców.

Głównym zadaniem projektu było stworzenie i uruchomienie pięciu usług działających w oparciu o sieć PIONIER: wideokonferencji, eduroam, usług kampusowych, powszechnej archiwizacji oraz naukowej interaktywnej telewizji HD.

ACK Cyfronet AGH na targach ISC High Performance w Hamburgu

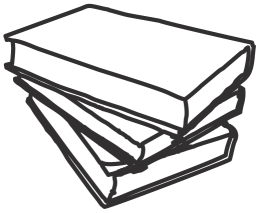
Cyfronet wystąpił jako wystawca w czasie jednego najważniejszych wydarzeń dla środowiska dostawców i użytkowników rozwiązań związanych z obliczeniami wielkiej skali.

Targi ISC High Performance (29.05.2022 – 02.06.2022) skupiły 137 wystawców i około 3000 uczestników z całego świata. Wśród wystawców znalazły się zarówno firmy – giganci technologiczni, jak Hewlett Packard Enterprise, Intel, czy NVIDIA, jak i czołowe inicjatywy i centra superkomputerowe, w tym m.in. EuroHPC JU, LUMI, Barcelona Supercomputing Center. Stoisko ACK Cyfronet AGH było jednym z nielicznych polskich akcentów w czasie targów. Główne aktywności Cyfronetu objęły: podtrzymywanie relacji z dotychczasowymi partnerami, nawiązywanie nowych kontaktów z potencjalnymi dostawcami sprzętu i oprogramowania oraz ocenę nowych rozwiązań technologicznych możliwych do wdrożenia w ramach zarządzanej infrastruktury.

Udział Cyfronetu w ISC był również związany z działaniami promocyjnymi w ramach projektów: Narodowe Centra Kompetencji EuroHPC – EuroCC oraz Narodowa Infrastruktura Superkomputerowa dla EuroHPC – EuroHPC PL.

Młodzi naukowcy w Cyfronecie

Organizowany corocznie konkurs na najlepszą pracę doktorską zrealizowaną w oparciu o zasoby obliczeniowe ACK Cyfronet AGH jest jedną z ważniejszych tradycji w naszym Centrum. Ocenie konkursowej poddawana jest wartość naukowa zgłoszonej pracy doktorskiej, możliwość jej praktycznego zastosowania oraz zakres wykorzystania zasobów obliczeniowych i pamięci składowania danych w Cyfronecie. Na przestrzeni ostatnich lat Konkurs stał się ważnym narzędziem promującym badania prowadzone przez młodych naukowców, a do kolejnych edycji Konkursu został zgłoszony szereg prac opisujących ciekawe zagadnienia naukowe z takich dziedzin jak chemia, biofizyka, fizyka czy informatyka. Młodzi badacze łączyli w swoich badaniach wyniki bezpośrednich pomiarów w laboratorium i obliczenia na komputerach dużej mocy. W swojej pracy korzystali z różnych narzędzi uruchamianych na szerokim wachlarzu architektur obliczeniowych oferowanych przez Cyfronet.



Prezentacja najważniejszych tez prac doktorskich Laureatów Konkursu miała miejsce w trakcie Dnia Otwartego Cyfronetu. Z przyjemnością prezentujemy fragmenty kilku wybranych wywiadów z uczestnikami Konkursu, które znajdziecie Państwo na kolejnych stronach.

Zapraszamy serdecznie do uczestnictwa w kolejnej edycji Konkursu!

http://www.cyfronet.pl/konkurs

Laureaci jednej z poprzednich edycji

Dr Dawid Dułak

Rozmowa z autorem pracy:
„Mechanizm transformacji amyloidowej bazujący na zmianie rozkładu hydrofobowości"

Jak rozpoczęła się Pańska przygoda z nauką? Czy może się Pan z nami podzielić swoją historią?

W moim wypadku prowadzenie badań doktoranckich wynikło z pasji do tematów biofizycznych. Kierunek moich zainteresowań zaczął się kształtować na przedmiocie „Bioinformatyka" na studiach magisterskich (Informatyka na Uniwersytecie Jagiellońskim). Dalej rozwijałem swoją wiedzę medyczną poprzez studia inżynierskie (Fizyka Medyczna na Akademii Górniczo-Hutniczej), gdzie pisałem pracę inżynierską związaną tematycznie z fałdowaniem białek. Wstępne zrozumienie tematu i moje zainteresowanie badaniami na tym etapie przerodziło się w regularną współpracę z Zakładem Bioinformatyki i Telemedycyny Uniwersytetu Jagiellońskiego.

Nie pozostało to bez wpływu na kolejne decyzje…

Od dziesięciu lat współpracowałem z Zakładem Bioinformatyki i Telemedycyny Uniwersytetu Jagiellońskiego, badając temat fałdowania się białek w kontekście modelu stworzonego w ramach pracy wyżej wymienionego Zakładu. Dalsza droga była niejako naturalną konsekwencją podjętych wcześniej wyborów. Chciałem prowadzić badania w obszarze, w którym już od lat się poruszałem. Kwestią otwartą pozostawał wybór, który fragment lub jaki element badań przyjąć jako oś przewodnią rozprawy doktorskiej.

Fałdowanie się białek jest niezwykle fascynującym procesem, w którym ogromna liczba aminokwasów łącząc się ze sobą przyjmuje zawsze jedną i tę samą formę (o ile aminokwasy są ułożone w tej samej kolejności). Choć organizm wykonuje tę operację bardzo szybko i bardzo często, my nie jesteśmy go w stanie odtworzyć. Badając ten problem, Zakład Bioinformatyki i Telemedycyny Uniwersytetu Jagiellońskiego pod przewodnictwem prof. Ireny Roterman-Koniecznej stworzył informatyczny model fałdowania się białka. Moim celem stało się usprawnienie tego modelu poprzez zrozumienie przyczyny, dla której niektóre białka fałdują się niepoprawnie (tzw. amyloidy, które przyczyniają się m.in. do rozwoju choroby Alzheimera).

Jak wspomina Pan pracę z zespołem naukowym?

Byłem zaangażowany w prace najbardziej od strony informatycznej, a kontakty z zespołem pozwalały mi szybko i w przystępny sposób zrozumieć niuanse świata biologiczno-chemiczno-fizycznego. Najbardziej cieszyła mnie wymiana spostrzeżeń, rezultatów badań i wzajemnej pomocy przy pracy badawczej.

Czy w trakcie prac wystarczający był własny sprzęt komputerowy?

Do tego typu obliczeń, tj. symulacji fałdowania wykorzystuje się modelowanie komputerowe, które jest bardzo wymagające obliczeniowo ze względu na ilość elementów (w tym wypadku atomów) i sił działających pomiędzy nimi. Należało najpierw odnaleźć odpowiednie oprogramowanie. Korzystałem z pakietów wchodzących w skład programu GROMACS, a także z autorskiego programu napisanego przez Panią Małgorzatę Gadzała. O ile pierwszy z nich jest dosyć popularny i można w łatwy sposób dotrzeć do informacji na jego temat, to działanie drugiego pozwolę sobie krótko opisać. Umożliwia on symulowanie środowiska wodnego jako pole i jego wpływ na fałdowanie białek. Modelowanie wymagało naprzemiennego działania jednej i drugiej aplikacji w różnej ilości powtórzeń, więc wykonywałem ok. 500 symulacji dla jednego białka. Podczas kilkuletniej

pracy badawczej poddałem symulacji fałdowania kilkadziesiąt sekwencji aminokwasów (niektóre z nich kilkukrotnie) tworzących natywnie konkretne białka. Do pracy doktorskiej wybrałem pięć z nich w związku z przyjmowaniem przez nie formy amyloidowej. Głównymi komputerami wykorzystywanymi przeze mnie na etapie prowadzenia obliczeń (symulacji komputerowych) były superkomputery Zeus, a później Prometheus. Zwykły komputer osobisty – nawet wyposażony w najwyższej klasy procesor, kartę graficzną i pamięć – wykonywałby pojedynczą symulację kilka dni, tygodni, a nawet miesięcy, więc jest to wielka rzecz, że takie maszyny są w Krakowie i możemy je wykorzystywać do tak ważnych – w mojej ocenie – badań naukowych.

Białko TAU czyli amyloid tworzący złogi (tj. wiele łańcuchów połączonych ze sobą) w przestrzeni międzykomórkowej (choroba Alzheimera)

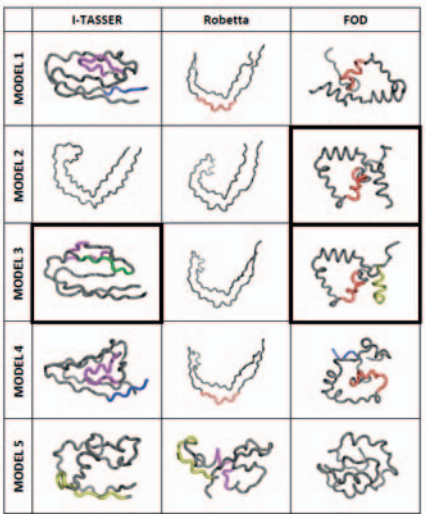
Współpraca z ACK Cyfronet AGH trwa już dobrych kilka lat, a istotna dla badań aplikacja właśnie tam powstała. Tylko dzięki superkomputerom – które są kluczowe w symulacjach dynamiki molekularnej – byłem w stanie przeprowadzić opisane wyżej badania. Zespół ACK Cyfronet AGH zawsze był dostępny i pomocny przy wszelkiego rodzaju wyzwaniach, więc liczę na dalszą, owocną współpracę i możliwość korzystania z zasobów ośrodka.

Cieszymy się, że mogliśmy pomóc, a i w przyszłości zapraszamy do kontaktu z Cyfronetem! Bazując na własnych doświadczeniach, jaką radę może Pan dać młodym naukowcom?

Może zabrzmi to bajkowo, ale w mojej ocenie człowiek rozpoczynający karierę naukową powinien dać ponieść się marzeniom i wyobraźni, bo te rodzą pasję. A potem cóż… Trzeba wytrwale i cierpliwie realizować kolejne postawione sobie cele.

Czy planuje Pan kontynuowanie swojego rozwoju w obranym kierunku?

Jak już wspomniałem, temat ten jest dla mnie niezwykle interesujący. Pomimo silnie rozwijającego się obszaru sztucznej inteligencji, która jest w stanie w dużym stopniu przewidzieć strukturę białka bazując na jego sekwencji aminokwasów, nie jest ona w stanie odpowiedzieć na pytanie jak ten mechanizm dokładnie wygląda. Dlatego jeszcze długa droga przed nami, aby w pełni zrozumieć naturalny proces, który za fałdowaniem białek stoi. Niewątpliwie będę kontynuował badania w tym obszarze, więc kto wie, może przyjdzie czas także na wyższy stopień „wtajemniczenia".



5 najlepszych struktur (modeli) pojedynczych łańcuchów, bazujących na sekwencji białka TAU, wygenerowanych przy użyciu trzech różnych narzędzi (I-TASSER, ROBETTA i DRIPPY). Pogrubiono te, które potencjalnie nie powinny tworzyć złogów. Struktury (modele) wygenerowane przy użyciu narzędzia DRIPPY uzyskano z wykorzystaniem oprogramowania w ramach infrastruktury PLGrid w ACK Cyfronet AGH

Dziękujemy za rozmowę i życzymy spełnienia planów naukowych.

Dr inż. Michał Koziarski

Rozmowa z autorem pracy:
„Imbalanced data preprocessing techniques utilizing local data characteristics"

Czy może Pan w syntetyczny sposób scharakteryzować tematykę Pańskiej pracy?

Tematyką mojej pracy była klasyfikacja danych niezbalansowanych, a konkretnie opracowanie nowych metod przetwarzania wstępnego tego typu danych, które pomogłyby zredukować negatywny wpływ niezbalansowania na jakość klasyfikacji. O niezbalansowaniu danych możemy mówić wtedy, kiedy jedna z klas, które chcielibyśmy rozpoznawać wśród naszych danych, jest słabiej reprezentowana niż inne klasy – w rzeczywistych danych niezbalansowanie jest bardzo powszechne, a pytanie dotyczy raczej tego, jak silny jest jego wpływ, a nie czy w ogóle występuje.

Większość standardowych algorytmów klasyfikacji jest słabo przystosowana do operowania na danych niezbalansowanych i ma tendencję do preferowania klas silniej reprezentowanych kosztem jakości predykcji na klasach rzadziej reprezentowanych – co jest przeciwieństwem tego, co zwykle chcielibyśmy osiągnąć. Jednym z podejść do odwrócenia tego trendu jest zastosowanie metod wstępnego przetwarzania danych, to jest algorytmów, które manipulują danymi treningowymi: albo poprzez redukcję liczby obserwacji z klas silniej reprezentowanych, albo poprzez tworzenie syntetycznych obserwacji z klas rzadziej reprezentowanych – i właśnie tworzeniem nowych metod tego typu zajmowałem się w mojej pracy.

Proszę powiedzieć, jak zrodziło się zainteresowanie uczeniem maszynowym?

Uczeniem maszynowym (i szerzej sztuczną inteligencją) zacząłem się interesować na pierwszych latach moich studiów, zasadniczo od pierwszego zetknięcia z tą tematyką. Z mojego punktu widzenia był to problem, którego rozwiązanie umożliwiłoby rozwiązanie za jednym zamachem wszystkich innych problemów (tj. posiadając dostatecznie silne algorytmy sztucznej inteligencji bylibyśmy w stanie rozwiązać wszystkie pozostałe problemy).

Rozwój sztucznej inteligencji jest postrzegany jako niezwykle ważny trend w nauce. Czy spodziewa się Pan spektakularnego przełomu, czy też będziemy mieli raczej do czynienia z ewolucją liczoną w dekadach?

Spektakularny przełom już się zasadniczo dokonał, po prostu nie tyle w AGI (ang. *Artificial General Intelligence* – generalnej sztucznej inteligencji), czyli w uproszczeniu w tym, jak przedstawiana jest sztuczna inteligencja w filmach i literaturze, ale w zastosowaniach uczenia maszynowego do rozwiązywania konkretnych problemów praktycznych. A te są wszędzie wokół nas: poczynając od zawartości naszych telefonów, przez różnego rodzaju diagnozowanie medyczne i przyspieszanie procesu wynajdywania lekarstw, a na tak fundamentalnych tematach jak predykcja struktur białkowych kończąc – żeby podać tylko kilka przykładów.

Już na obecnym etapie Pańskie algorytmy wspomagają diagnostykę onkologiczną – prosiłbym o kilka zdań na temat tego typu współpracy.

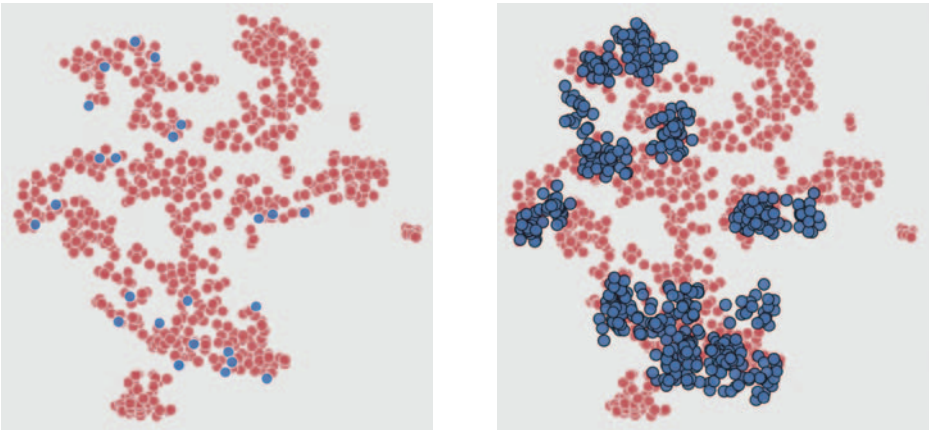
W części aplikacyjnej mojej pracy skupiłem się właśnie na automatycznej detekcji nowotworów na zdjęciach histopatologicznych, czyli mikroskopowych obrazach tkanki. Problem jest szczególnie istotny ze względu na fakt, że w Polsce brakuje nam aktualnie doświadczonych lekarzy histopatologów, więc jakiekolwiek przyspieszenie ich pracy miałoby duże znaczenie praktyczne.

Zadanie to sprowadza się do zaklasyfikowania regionów tkanki jako zdrowych lub zmienionych nowotworowo, a w tym drugim przypadku – dodatkowo do oceny stopnia złośliwości nowotworu. Jest to problem mocno niezbalansowany, bo tkanka

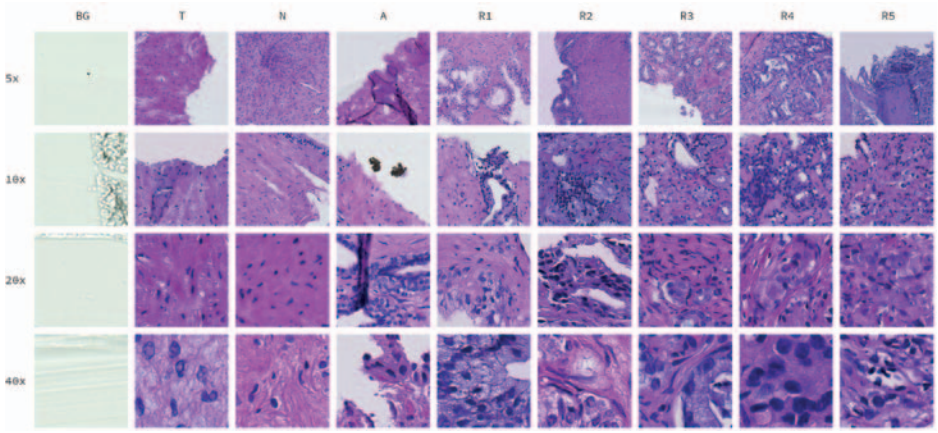
zdrowa jest zdecydowanie dominująca, a zwłaszcza niektóre stopnie złośliwości występują stosunkowo rzadko. Poza tym, od strony uczenia maszynowego zagadnienie jest ciekawe, bo zbiory danych jak i pojedyncze skany histopatologiczne są bardzo duże. Poza tym, mamy do czynienia z dużą niepewnością na poziomie etykiet, a cały proces klasyfikacji wymaga zastosowania sieci konwolucyjnych. Miałem więc możliwość zastosowania niektórych z opracowanych w mojej pracy algorytmów w finalnie opracowanym rozwiązaniu.

Czy Cyfronet spełnił Pańskie oczekiwania w zakresie dostępności oraz jakości zasobów obliczeniowych?

Bez dostępu do zasobów Cyfronetu moja praca zwyczajnie by nie powstała, a na pewno nie w swojej aktualnej formie – udostępnione zasoby umożliwiły mi opracowanie części bardziej złożonych obliczeniowo metod, a w przypadku tych mniej wymagających – przeprowadzenie dokładnych i obszernych prac eksperymentalnych.



Po lewej: przykład niezbalansowanego zbioru danych, z kolorem oznaczającym klasę obserwacji. Tradycyjne algorytmy klasyfikacji nie radzą sobie z silnie niezbalansowanymi danymi. Po prawej: zbiór po dodaniu syntetycznych obserwacji, wygenerowanych przy pomocy jednego z opracowanych algorytmów



Przykład fragmentów zdjęć histopatologicznych z różnych klas, klasyfikacja których była przedmiotem części aplikacyjnej doktoratu. Tego typu dane medyczne są często silnie niezbalansowane, ze znacząco mniejszą liczbą zdjęć zawierających zmiany nowotworowe

Dr Mercedes Kukułka

Rozmowa z autorką pracy:
„Theoretical investigations on absorption and luminescence properties of selected organic species for potential applications in optoelectronics"

Jak przybliżyłaby Pani swoje badania osobie niezwiązanej z dziedziną?

Podstawowym celem moich badań była identyfikacja oraz zrozumienie, w oparciu o wyniki obliczeń kwantowo-chemicznych, związku pomiędzy szeroko rozumianą strukturą (geometria molekularna, struktura elektronowa) a właściwościami fotofizycznymi w wybranych układach organicznych o potencjalnym zastosowaniu w optoelektronice. Upraszczając, moje badania przyczyniły się do wskazania elementów struktury kilku grup związków chemicznych, które mają wpływ na to, że dany układ absorbuje bądź emituje promieniowanie.

Czy współczesna fotofizyka i fotochemia znajduje już zastosowanie w tak zwanym życiu codziennym, czy jesteśmy tu na etapie badań podstawowych?

Oczywiście, że znajduje! Wiadomo, że od odkrycia do wdrożenia mija sporo czasu, jednak w ciągu ostatnich lat jesteśmy świadkami bardzo dynamicznego rozwoju w zakresie nowych technologii. Spójrzmy chociaż na rozwój technologii wykorzystywanych przy produkcji monitorów ekranowych. Od odkrycia ciekłego kryształu (w 1888 r.) do skonstruowania pierwszego wyświetlacza ciekłokrystalicznego (ang. *Liquid-Crystal Display*, *LCD*) upłynął prawie wiek, a do momentu wyprodukowania na skalę przemysłową ekranu o przekątnej większej niż 10″ – kolejne trzy dekady. Historia monitorów opartych na technologii OLED (ang. *Organic Light-Emitting Diode*) jest już znacznie krótsza. Zjawisko elektroluminescencji organicznych związków chemicznych zostało zaobserwowane po raz pierwszy w latach 50. XX w., pierwsze wydajne urządzenie zostało skonstruowane w 1987 r., a wyświetlacze OLED produkowane na skalę komercyjną to, podobnie jak w przypadku monitorów LCD, przełom XX i XXI w. Obecnie na rynku spotkamy także wyświetlacze QLED (ang. *Quantum dot Light-Emitting Diode*), wykorzystujące technologię kropek kwantowych. W tym przypadku idea narodziła się w latach 90. XX w., natomiast pierwsze telewizory QLED pojawiły się w sprzedaży już w 2013 r.

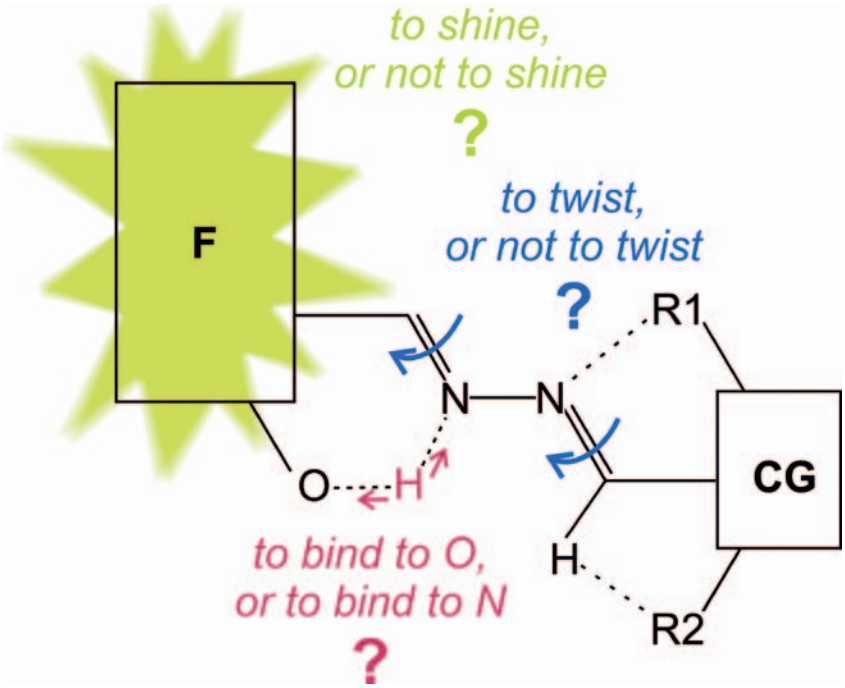
Konkurencyjny, rozwijający się globalny przemysł XXI w. ciągle poszukuje nowych materiałów, które będą oferować lepsze właściwości, czy które będzie cechować większa wydajność. Coraz częściej ośrodki naukowo-badawcze nawiązują bliską współpracę z przemysłem, co jest częściowo wspierane przez rozwiązania systemowe. Przedsiębiorstwa prowadzące działalność badawczo-rozwojową mogą bowiem korzystać z pewnych ulg podatkowych, natomiast naukowcom łatwiej jest zdobyć finansowanie na badania, jeśli mają one charakter aplikacyjny.

Nie bez znaczenia dla szybkiego rozwoju w badaniach nad nowymi materiałami jest też fakt upowszechnienia się badań teoretycznych. Analiza rezultatów badań *in silico* stanowi nie tylko podstawę do interpretacji danych eksperymentalnych, ale również pozwala na dogłębne zrozumienie relacji pomiędzy strukturą molekularną a własnościami układu, co jest kluczowe dla opracowywania nowych materiałów o pożądanych właściwościach.

Jakie dalsze wyzwania rysują się na horyzoncie i czy dostęp do komputerów dużej mocy zmienia coś w tym obrazie?

Wynik oddziaływania światła ze związkami chemicznymi nie zawsze jest łatwy do przewidzenia, a badanie stanów elektronowo wzbudzonych wciąż stanowi duże wyzwanie dla metod teoretycznych. W przypadku obliczeń własności zależ-

nych od tych stanów często konieczne jest zastosowanie bardziej zaawansowanych metod, które z kolei wymagają znacznie większych zasobów obliczeniowych, a przez to są ograniczone do układów o niewielkich rozmiarach. Dodatkowo, w poszukiwaniach zależności między strukturą związków chemicznych a ich aktywnością optyczną i własnościami fotochemicznymi coraz częściej wykorzystuje się metody uczenia maszynowego. W obydwu przypadkach dostęp do komputerów dużej mocy jest kluczowy.



Schematyczne przedstawienie zależności między strukturą molekularną a własnościami fotofizycznymi grupy ligandów typu zasad Schiffa. F – fluorofor, CG – grupa kontrolująca, R1 oraz R2 – podstawniki przyłączone do CG

Czy Cyfronet zapewnił wystarczające wsparcie w zakresie dostępności oraz jakości zasobów obliczeniowych?

Bez wsparcia Cyfronetu realizacja moich badań byłaby praktycznie niemożliwa. Cyfronet zapewnił mi zarówno dostęp do oprogramowania naukowego, jak i niezbędnych mocy obliczeniowych. Nie sposób tu nie wspomnieć o prężnie działającym dziale pomocy technicznej – zawsze mogłam liczyć na wsparcie i pomoc pracowników Centrum, a czas reakcji na zgłaszane problemy był ekspresowy.

Czy ma Pani jakieś rady lub spostrzeżenia dla studentów zainteresowanych podjęciem studiów doktoranckich?

Studia doktoranckie to bardzo angażujący czasowo etap w karierze młodego naukowca, wpływający w istotny sposób również na jego życie osobiste. Bez pasji i zainteresowania tematyką badań ukończenie go jest niezwykle trudne, o ile w ogóle możliwe. Należy też pamiętać, że doktorat to bieg długodystansowy, a nie sprint. Wyniki zazwyczaj nie przychodzą łatwo, nie każdy eksperyment się uda – trzeba uzbroić się w cierpliwość i być wytrwałym. Doktorat to nie tylko badania nad konkretnym tematem, to ogólny rozwój, zapoznanie ze środowiskiem pracy naukowej, nabycie większej samodzielności, umiejętności radzenia sobie z wyzwaniami, przyjmowanie porażek, realizacja celów mimo przeciwności losu. Takie doświadczenie ma ogromną wartość nie tylko w dalszej pracy jako naukowiec, ale ogólnie – w życiu codziennym.

Dr Ewa Machalska

Rozmowa z autorką pracy:
„Badania wzmocnienia sygnału ramanowskiej aktywności optycznej układów molekularnych"

Co w największym stopniu skłoniło Panią do realizacji badań w oparciu o rejestrację i analizę widm ROA?

Ramanowską aktywnością optyczną (ROA), jak też elektronowym dichroizmem kołowym (ECD) zainteresowałam się w trakcie pracy magisterskiej, w ramach której skupiłam się na eksperymentalnym i teoretycznym wyjaśnieniu procesu agregacji chiralnych karotenoidów. Następnie, chęć kontynuacji rozpoczętych badań i możliwość praktycznego zastosowania spektroskopii ROA w wielu dziedzinach nauki, w tym w obszarach biomedycznych, skłoniła mnie do podjęcia studiów doktoranckich na Wydziale Chemii Uniwersytetu Jagiellońskiego oraz dalszego rozwiązywania problemów badawczych, jakie niesie ze sobą, nie pozwalając popaść w rutynę, metoda ROA. Biorąc pod uwagę potencjalnie silny wpływ zmiany chiralności na pożądane właściwości biologiczne i lecznicze cząsteczek, wyjątkowe znaczenie niesie dla mnie aspekt praktycznego wykorzystania technik chiralooptycznych, w tym w przemyśle farmaceutycznym.

W jakim zakresie ułatwieniem, a w jakim utrudnieniem jest fakt, że spektroskopia rozpraszania ramanowskiego jest obszarem o stosunkowo niskim poziomie eksploracji?

W mojej opinii, na przełomie ostatnich lat spektroskopia rozpraszania ramanowskiego, mimo iż oparta jest na stosunkowo słabym efekcie, wzmocniła swój poziom eksploatacji w wielu obszarach nauki. Wyraźny aplikacyjny charakter spektroskopii ramanowskiej widoczny jest w badaniach biomedycznych, np. w analizie białek, cukrów, lipidów, a także składników krwi, w tym erytrocytów oraz leukocytów. Z pewnymi niedogodnościami, jakie bez wątpienia niesie ze sobą spektroskopia ramanowska, grono naukowców nauczyło się walczyć poprzez opracowanie nowych strategii wzmacniających uzyskiwany sygnał ramanowski.

Co w Pani badaniach stanowiło największe wyzwanie i w jaki sposób sobie Pani z nim poradziła?

Niewątpliwie najbardziej czasochłonne, ale też niezwykle intrygujące wyzwanie stanowiły dla mnie badania nad szeregiem pochodnych witaminy B_{12}, na przykładzie których udało się odkryć i wyjaśnić nieznany wcześniej w literaturze efekt ECD-Raman, czyli interferencję zjawiska elektronowego dichroizmu kołowego oraz rozpraszania ramanowskiego, który zdominował naturalne rezonansowe widmo ROA kobalamin. Po wielu miesiącach pracy w laboratorium, generalnie polegających na rejestracji wzmocnionych rezonansowo widm ROA analogów kobalamin, wspólnie z promotor mojej rozprawy doktorskiej Panią Prof. dr hab. Małgorzatą Barańską jak też Panem dr Grzegorzem Zającem udało się nam opracować kilka eksperymentalnych oraz teoretycznych strategii eliminacji efektu ECD-Raman na rezonansowych widmach ROA. Co więcej, skrupulatne badania pochodnych witaminy B_{12} doprowadziły nas do wniosku, iż uzyskanie wiarygodnego rezonansowego widma ROA barwnych substancji jest możliwe, ale tylko wtedy, gdy pod uwagę weźmie się efekt ECD-Raman. W przeciwnym wypadku, nieuwzględnienie potencjalnego wystąpienia zjawiska ECD-Raman podczas pomiarów rezonansowych widm ROA może prowadzić do błędnych interpretacji chiralooptycznych widm i wniosków dotyczących struktury molekularnej nie tylko kobalamin, ale również wielu innych chiralnych związków.

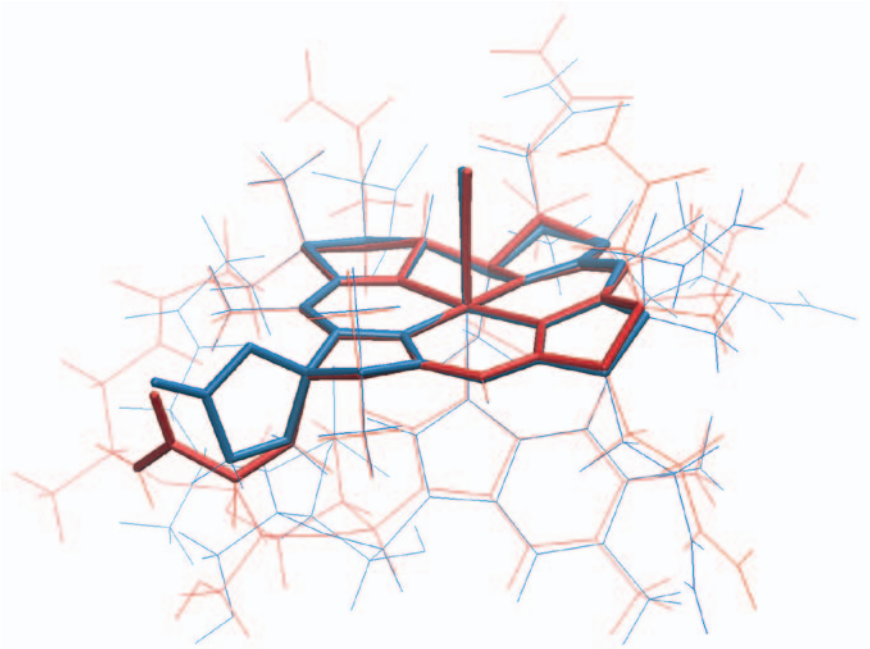
W jaki sposób skorzystała Pani z zasobów udostępnianych przez Cyfronet? W czym były najbardziej pomocne?

W badaniach, które stanowiły podstawę mojej rozprawy doktorskiej, poprzez infrastrukturę PLGrid wykorzystałam zasoby obliczeniowe superkomputera Prometheus. W tym miejscu chciałabym zaznaczyć, że efektywne obliczenia kwantowo-chemiczne nie byłyby możliwe bez owocnej współpracy z Panem dr Grzegorzem Zającem z Jagiellońskiego Centrum Rozwoju Leków w Krakowie.

Bez wątpienia zasoby obliczeniowe udostępnione przez Cyfronet były najbardziej pomocne na etapie wyjaśnienia wzmocnionych rezonansowo widm ROA witaminy B_{12} i jej pochodnych. Wykonane obliczenia, z jednej strony, pozwoliły na uzyskanie widm ROA, których profil pozostawał w bardzo dobrej zgodności z wynikami eksperymentalnymi, a z drugiej symulacje widm były niepodważalnym potwierdzeniem dwuznakowości eksperymentalnych widm ROA.

Jaką poradę mogliby od Pani otrzymać młodzi naukowcy, którzy dopiero rozpoczynają pracę nad doktoratem lub planują ją rozpocząć?

Myślę, że najlepszą poradą, jaką mogę z całą pewnością dać wszystkim kandydatom rozpoczynającym studia doktoranckie, to żeby nie bali się oni nowych wyzwań, jakie stawia przed nimi nauka. Na pozór interdyscyplinarność podejmowanych badań może, z jednej strony, być niezwykle trudnym oraz przerażającym zadaniem dla każdego młodego naukowca, ale z drugiej, po pewnym czasie, pozwoli im stworzyć świeże, odmienne od dotychczasowego podejście do problemów badawczych. Jest to możliwe do osiągnięcia przede wszystkim wtedy, gdy Doktoranci będą podejmować się tematów, które ich fascynują, cieszą, a zarazem są potężnym fundamentem do kreowania nowych pomysłów, które mogą wykorzystać w przyszłości.



Porównanie zoptymalizowanych struktur najbardziej stabilnych konformerów witaminy B_{12} ((CN)Cbl, kolor niebieski) oraz jej pochodnej ((CN)13-epi-Cbl(e-lakton), kolor czerwony). Pierścień korynowy i łańcuchy boczne c zostały pogrubione

Dr inż. Bartłomiej Rzeszotarski

Rozmowa z autorem pracy:
„Transport ładunku i spinu w nanoukładach na bazie silicenu"

Silicen, ze względu na swoją budowę i niektóre właściwości, jest popularnie porównywany do grafenu. Co sprawiło, że w badaniach skupił się Pan akurat na nanoukładach na bazie silicenu? Na jakich polach jego zastosowanie może przeważać nad zastosowaniem grafenu?

Silicen posiada wbudowane oddziaływania spin-orbita, które są wystarczająco silne, by móc kontrolować spin elektronu lub uzyskać w strukturze stan topologicznego izolatora. Wynika to ze złamania symetrii w geometrii ułożenia atomów, gdzie wyróżnia się dwie podsieci – górną i dolną. Takie ułożenie atomów krzemu w silicenie jest naturalne, tzn. w takiej konfiguracji otrzymujemy minimum energii układu ze względu na wiązania (konfiguracja planarna jest niestabilna). Natomiast w prostym grafenie jednowarstwowym nie zachodzi złamanie tej symetrii i oddziaływania spin-orbita są pomijalnie słabe. By wywołać w grafenie silne oddziaływania spin-orbita, trzeba złamać symetrię ręcznie, np. poprzez domieszkowanie, zwijanie w nanorurki, bądź układanie w stos kilku skręconych względem siebie monowarstw grafenu. Prostota układu silicenu zaważyła o jego wyborze, a ponadto technologia krzemowa jest już bardzo dobrze znana od lat, dlatego widzieliśmy ogromny potencjał tego materiału zarówno w kontekście możliwości przeprowadzenia eksperymentów, jak i ewentualnych zastosowań w spintronice jako elementów budowy komputerów kwantowych. Naszą uwagę zwrócił działający tranzystor oparty na silicenie, dla którego opis eksperymentu został opublikowany w 2015 roku w czasopiśmie Nature – zapoczątkowało to naszą chęć do dalszych badań nad tym materiałem.

Jakie, według Pana, dalsze wyzwania stoją przed naukowcami badającymi silicen?

Stabilizacja silicenu jest bardzo trudna – jest to materiał, który bardzo szybko się utlenia. Dodatkowo podłoże, na którym osadza się silicen, nie powinno wpływać na jego właściwości, czyli dwuwymiarowego gazu elektronowego (przykładowo, metaliczne podłoże sprawia, że silicen przestaje być półprzewodnikiem). Aktualnie trwają prace związane ze znalezieniem odpowiedniego materiału podłoża, który nie zmodyfikuje znacząco struktury pasmowej, a jednocześnie ustabilizuje strukturę, by mogła pracować w warunkach narzuconych przez eksperyment. Choć w obecnym trendzie przeważają prace eksperymentalne nad skręconymi dwuwarstwami grafenu, to nie wyklucza to pojawienia się nowych prac eksperymentalnych dla silicenu, ale na drodze do wdrożeń praktycznych stoi jeszcze wiele wyzwań.

Jaki aspekt pracy naukowej daje Panu największą satysfakcję?

Największą satysfakcję daje mi zrozumienie istoty badanego zjawiska do takiego poziomu, że mogę opowiedzieć o nim w sposób prosty i zrozumiały dla każdego, nawet niezwiązanego z nauką.

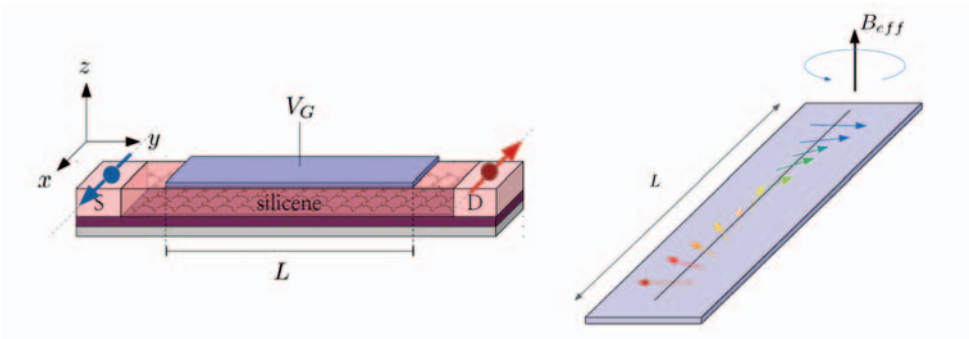
Po wielu godzinach pracy wyrabia się swego rodzaju intuicja, która podsuwa pewne obrazy dla analogii i rozumowania, które mogę przekazać innym. Wtedy jako naukowiec czuję się spełniony, ponieważ przekazuję wiedzę dalej.

W jaki sposób zasoby Cyfronetu przysłużyły się realizacji części Pana badań związanej z obliczeniami i ich analizą?

Wszystkie obliczenia były prowadzone z rozdzielczością atomową, co generuje ogromny koszt numeryczny. Gdyby nie zasoby Cyfronetu, rozwiązanie ujętych w rozprawie układów dla przestrzeni badanych parametrów w stosownym czasie (trwania doktoratu) byłoby nieosiągalne. Dlatego w części numerycznej cała zasługa leży w infrastrukturze superkomputerów Zeus i Prometheus znajdujących się w ACK Cyfronet AGH, za której udostępnienie ogromnie dziękuję.

Co mógłby Pan doradzić osobom, które dopiero rozpoczynają studia doktoranckie? Na co te osoby powinny zwrócić największą uwagę?

W pracy naukowej przez 90% czasu szukamy rozwiązań, czegoś nie rozumiemy, sprawdzamy czy nie ma błędów lub ich po prostu szukamy. Jest to okres „bólu", który trzeba wytrzymać. W końcu pokonujemy tę przeszkodę i otrzymujemy wyniki. I pozostałe 10% czasu przeznaczamy na pisanie publikacji, analizę wyników i ich obronę przed recenzentami. Praca naukowa nie jest prosta, jest często usiana frustracją i sytuacjami, w których nasza wiedza i umiejętności są podważane. Jeśli miałbym udzielić porady osobom zaczynającym studia doktoranckie, to przede wszystkim poleciłbym znaleźć radość między tymi 90. a 10. procentami powyżej opisanymi. I nie poddawać się, rozmawiać z innymi, szukać pomysłów. Nie jesteśmy sami na tej trudnej drodze.



Schemat układu inwertera spinowego wykonanego na bazie silicenu. Do obrotu spinu wykorzystano zjawisko precesji w efektywnym polu magnetycznym B_{eff}, pochodzącym z wbudowanych oddziaływań spin-orbita w silicenie, którym można sterować przy pomocy potencjału bramki V_G

Dr Zuzanna Wojdyła

Rozmowa z autorką pracy:
„Mechanizmy reakcji katalizowanych przez dioksygenazy zależne od 2-oksoglutaranu – badania metodami biochemii obliczeniowej"

Jak zaczęła się Pani przygoda z nauką?

Z chemią obliczeniową zetknęłam się po raz pierwszy na wykładach prowadzonych przez prof. Ewę Brocławik na UJ. Spodobały mi się zagadki, które można rozwiązywać w oparciu o obliczenia kwantowo-chemiczne. Gdy dowiedziałam się, że chemię obliczeniową można połączyć z badaniem enzymów, wiedziałam, że jest to kierunek, w którym chcę pójść. I tak ukończyłam studia magisterskie, a potem doktoranckie.

Jak widać – bardzo szybko ukierunkowały się Pani zainteresowania. Którymi enzymami postanowiła się Pani zajmować?

Enzymy, którymi się zajmowałam – dioksygenazy zależne od 2-oksoglutaranu, mają duży, ale jeszcze niewykorzystany potencjał przemysłowy. Są one zdolne do rozcinania i modyfikacji wiązań C-H. Reakcje te zachodzą z dużą selektywnością (powstaje niewiele produktów ubocznych). Dzięki temu mogłyby być wykorzystywane w produkcji złożonych związków chemicznych, w przypadku których szczególnie istotna jest czystość końcowego produktu, np. w przemyśle farmaceutycznym.

Wraz z zespołem wybraliśmy kilku przedstawicieli tej rodziny enzymów. Chcieliśmy zrozumieć, w jaki sposób kontrolują one przebieg reakcji i co sprawia, że różne dioksygenazy tej grupy, mimo wykorzystywania takiego samego centrum aktywnego, katalizują różne reakcje (hydroksylacji, odwodornienia lub cyklizacji). Wiedza ta znalazłaby zastosowanie na przykład w inżynierii enzymów do katalizowania pożądanych przemian – i taki był nasz dalekosiężny cel. Badania prowadzone w zespole łączą metody eksperymentalne i obliczeniowe, a nasza współpraca nadal się rozwija. Obliczenia pozwalają na uzupełnienie i interpretowanie wyników eksperymentów, a doświadczenia pozwalają na konfrontację wyników uzyskanych *in silico z rzeczywistością. Taki sposób prowadzenia badań bardzo mi się podoba.*

Jaki aspekt chemii obliczeniowej jest dla Pani najbardziej interesujący?

Moim ulubionym aspektem chemii obliczeniowej jest możliwość szczegółowego prześledzenia katalizowanej reakcji – od kompleksu enzym-substrat, do jej końcowych produktów. Interesująca jest możliwość testowania takich kanałów reakcji, które nie są obserwowane w doświadczeniach. Jest to ciekawe uzupełnienie badań eksperymentalnych, szczególnie że częstą „strategią" enzymów jest nie tylko obniżenie bariery energetycznej wybranej reakcji, ale również utrudnienie alternatywnych przemian. Dzięki obliczeniom można pokazać, w jaki sposób te przemiany są hamowane i wskazać czynniki promujące tę „właściwą" reakcję.

Co w trakcie prowadzonych badań było największym wyzwaniem?

Dla każdego badanego enzymu najciekawszym, ale i najbardziej wymagającym etapem było wyłuskanie efektu, który był tym decydującym dla zajścia obserwowanej reakcji. Enzymy mogą wpływać na reakcję na różne sposoby. Ściśle determinują położenie substratu lub specyficzne oddziaływania. W zależności od analizowanego kompleksu enzym-substrat obserwuje się dominację odmiennych efektów, a te z kolei mają wpływ na powstawanie różnych produktów. Czasem identyfikacja tego kluczowego czynnika wymagała sporo wyobraźni i pracy.

W jaki sposób prowadzone były prace?

Modelowanie reakcji katalizowanych przez enzymy opierało się przede wszystkim na wynikach eksperymentów – jako punkt wyjścia potrzebna była struktura kompleksu enzym-substrat. Na późniejszych etapach zestawiałam obraz reakcji uzyskany w wyniku obliczeń z rezultatami doświadczalnymi. Dane eksperymentalne mogą oczywiście pochodzić z literatury, ale ciekawsza jest dla mnie możliwość porównania rezultatów uzyskanych *in silico* z wynikami prowadzonych równolegle przez zespół badań laboratoryjnych. Żeby choć częściowo uchwycić wieloaspektowość katalizy enzymatycznej, potrzeba wielu narzędzi, np. pakiet Amber do symulacji dynamiki molekularnej (uwzględnienie ruchliwości cząsteczki enzymu i związanego substratu), Gaussian (śledzenie przebiegu reakcji), NCIPLOT (wizualizacja oddziaływań w układzie). Korzystałam też z pisanych zespołowo (mini)programów.

Dla obliczeniowego badania enzymów duże moce obliczeniowe są nieodzowne, dlatego korzystałam z możliwości, jakie dawały superkomputery Zeus i Prometheus. Dzięki współpracy z Cyfronetem mogłam się skupić na (bio)chemicznych aspektach pracy, a w razie problemów technicznych mogłam liczyć na ekspertów obsługujących Helpdesk. To pozwoliło mi efektywnie wykorzystać czas, więc był to na pewno dobry wybór.

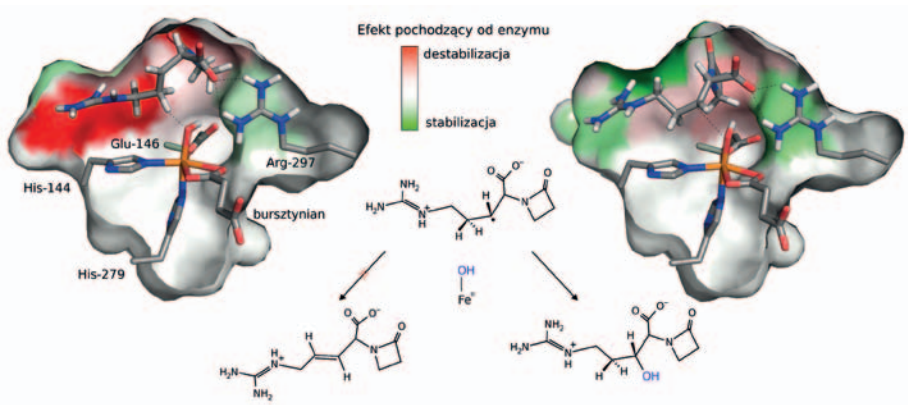
Co może zrobić młody naukowiec, gdy napotka jakieś przeszkody?

Myślę, że dobrą odpowiedzią będzie cytat z Thomasa Edisona: *„Nie poniosłem porażki. Znalazłem dziesięć tysięcy rozwiązań, które nie działają”.* To jest jeden z wielu cytatów, którym dzieli się Gaussian na koniec obliczeń i dość dobrze oddaje podejście, które pomogło mi w prowadzeniu badań.

Czy chciałaby Pani podzielić się z nami swoimi planami na przyszłość?

Planuję nowy projekt badawczy, w ramach którego chciałabym podejść do reakcji katalizowanych przez metaloenzymy nieco bardziej systematycznie – określić, w jaki sposób zmiany w otoczeniu (kieszeni wiążącej) i w samym centrum aktywnym wpływają na przebieg reakcji. To będzie wymagało zmiany perspektywy i na pewno będzie rozwijające.

Dziękuję za rozmowę i życzę spełnienia planów zawodowych.



Wizualizacja wpływu kieszeni wiążącej enzymu na selektywność katalizowanej reakcji – otoczenie zwiększa barierę energetyczną związaną z utworzeniem wiązania podwójnego (z lewej), obniża barierę dla hydroksylacji (z prawej)

Dr inż. Dominik Żurek

Rozmowa z autorem pracy:
„Akceleracja obliczeń algorytmów uczenia maszynowego oraz wybranych populacyjnych algorytmów inteligencji obliczeniowej ze zredukowaną precyzją danych poprzez implementację w układach GPGPU"

Jak przybliżyłby Pan swoje badania osobie niezwiązanej z dziedziną?

Celem badań było opracowanie i zanalizowanie skutecznych i efektywnych metod przyśpieszania obliczeń nadających się do przetwarzania równoległego, związanych z algorytmami sztucznej inteligencji oraz algorytmami ewolucyjnymi. Łatwo zauważyć, że pojawiły się tu trzy aspekty, tj. akceleracja obliczeń, sztuczna inteligencja oraz algorytmy ewolucyjne, które postaram się pokrótce przybliżyć.

Pojęciu sztucznej inteligencji (ang. *Artificial Intelligence – AI*) przypisuje się kilka znaczeń, najczęściej jednak termin ten definiowany jest jako uczące się maszyny lub systemy informatyczne, które przetwarzają informację w oparciu o reguły ludzkiego rozumowania. Sztuczna inteligencja zajmuje się zagadnieniami, które nie są efektywnie algorytmizowane w oparciu o modelowanie wiedzy, więc do ich rozwiązania należy wprowadzić algorytmy posiadające znamiona inteligencji. Przez określenie „inteligencja" rozumie się w tym przypadku zdolność do samodzielnego przystosowania się do zmiennych warunków. Przez proces uczenia się systemu rozumie się dokonanie autonomicznej zmiany w systemie, zachodzącej na podstawie zdobytych doświadczeń i prowadzącej do poprawy jakości jego działania. Przy takiej definicji zakłada się, że istnieje możliwość oceny jakości podejmowanych decyzji, czyli umiejętność odróżnienia zmian korzystnych od niekorzystnych. Sztuczna inteligencja potrafi przyswajać wiedzę poprzez wyodrębnianie wzorców z surowych danych. Wprowadzając do algorytmu elementy ludzkiej inteligencji możliwe jest wytrenowanie go tak, by potrafił rozpoznawać obrazy, rozumiał język naturalny czy też był zdolny do logicznego rozumowania.

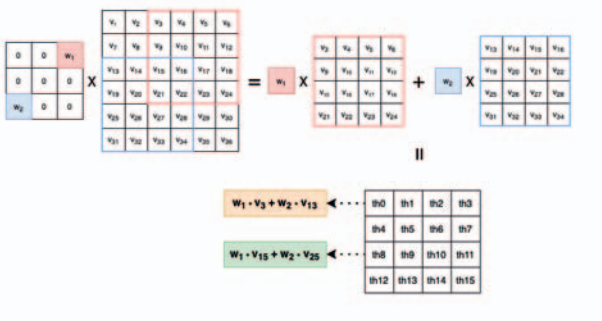
Jedną z poddziedzin sztucznej inteligencji jest tzw. inteligencja obliczeniowa (ang. *Computational Intelligence – CI*), która polega głównie na zdolności przystosowania się systemu do zmieniającego się środowiska. Kładzie ona duży nacisk na ulepszanie i rozwój aplikacji w świecie rzeczywistym. Cechą odróżniającą algorytmy tego typu od innych algorytmów sztucznej inteligencji jest brak korzystania ze zdefiniowanego modelu, lecz podejmowanie prób zbudowania go samodzielnie na podstawie dostarczonych zbiorów uczących. Algorytmy inteligencji obliczeniowej obejmują algorytmy wywodzące się ze sztucznej inteligencji związane z inteligentnym przetwarzaniem danych, w których istotną grupę stanowią algorytmy ewolucyjne. Ich inspiracją są procesy naturalnej ewolucji organizmów. Przenosząc inspiracje biologiczne na algorytmy ewolucyjne, populacja osobników jest odzwierciedleniem liczby rozwiązań (osobnik = rozwiązanie). Ewolucja rozwiązań algorytmu jest odpowiednikiem zmian występujących w populacji, za co odpowiedzialne są operatory mutacji (losowa modyfikacja) oraz rekombinacji (wymiana materiału genetycznego), co sprowadza się do wyszukiwania kolejnych rozwiązań.

Wspomniane grupy algorytmów do prawidłowego działania potrzebują ogromnych zasobów mocy obliczeniowej, stąd też rodzi się potrzeba opracowania metod, dzięki którym możliwe będzie szybsze ich wykonanie, czyli „akceleracja obliczeń".

Co zadecydowało o tym, że poświęcił Pan rozprawę doktorską zagadnieniu akceleracji w układach GPGPU?

Układy GPGPU specjalizują się w obliczeniach wysoce równoległych, poprzez przeprowadzanie obliczeń zgodnie z ar-

chitekturą SIMD (ang. *single instruction stream multiple data*). Jak zostało wyżej wspomniane, w moich badaniach skupiłem się na algorytmach uczenia maszynowego oraz algorytmach ewolucyjnych, ze względu na ich szerokie spektrum zastosowań. Obie te grupy algorytmów wykazują tendencję do równoległości, stąd wybór kart graficznych jako akceleratora sprzętowego wydawał się najbardziej naturalny.



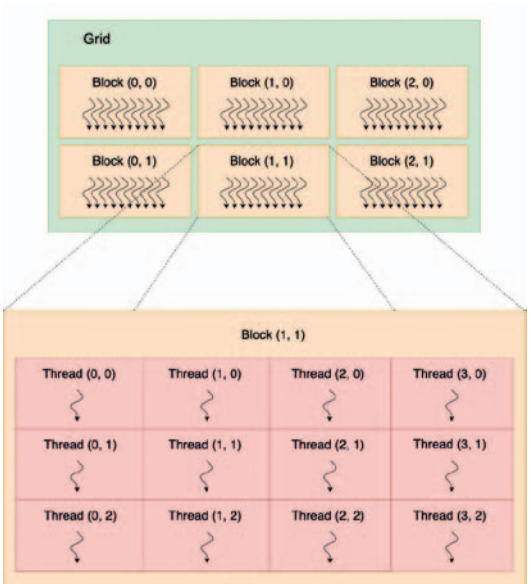
Obliczanie konwolucji na układach GPGPU przy użyciu operacji na macierzach rzadkich

Spróbujmy spojrzeć w przyszłość: jak badania mogą się przełożyć na wdrożenia praktyczne? Czy jest dziedzina szczególnie oczekująca na przełom w tej materii?

W moim przekonaniu zdecydowanie taką dziedziną jest neuroewolucja, którą obecnie się zajmuję, dzięki wiedzy i doświadczeniu zdobytym podczas pracy nad doktoratem. Neurowolucja zajmuje się rozwojem sieci neuronowej poprzez użycie algorytmów ewolucyjnych (czyli dotyka obu grup algorytmów, które były przedmiotem moich badań). Rozwój ten może odbywać się na dwóch płaszczyznach, tj. na poziomie poszukiwania optymalnej architektury sieci neuronowej, bądź za jego przyczyną może odbywać się sam proces uczenia sieci (mutacja wag). W obu przypadkach wymagana jest ogromna moc obliczeniowa. W związku z tym metody potrafiące przetwarzać w sposób bardziej efektywny algorytmy sztucznej inteligencji oraz algorytmy ewolucyjne na pewno znajdą przełożenie na wdrożenia praktyczne w tej materii.

Jak w kontekście badań w dziedzinie akceleracji i obróbki danych ocenia Pan rolę Cyfronetu?

W swoich badaniach użyłem bardzo wydajnej i nowoczesnej karty graficznej, tj. Nvidia Tesla V100-SXM2-32GB. Dzięki niej dla niektórych z podjętych problemów możliwe stało się osiągnięcie ogromnych przyśpieszeń względem starszych kart oraz względem wielordzeniowych procesorów ogólnego przeznaczenia. Do wszystkich tych akceleratorów miałem dostęp dzięki Cyfronetowi, więc z całą pewnością mogę powiedzieć, że bez Cyfronetu nie byłoby możliwe uzyskanie tak imponujących wyników.



Schemat grupowania wątków na karcie graficznej

Czy miałby Pan jakieś rady dla kolegów i koleżanek rozważających podjęcie studiów doktoranckich?

Na początku swojej przygody naukowej trafiłem do zespołu „Akceleracji obliczeń" w Cyfronecie, gdzie poznałem wybitnych specjalistów, którzy okazali się również wspaniałymi ludźmi. Dzięki temu obudziła się we mnie pasja oraz głęboka chęć bliższego zaznajomienia się z tematyką, którą podjąłem w rozprawie doktorskiej i którą zajmuję się do dziś. Dlatego też młodym naukowcom rekomenduję nawiązanie współpracy z zespołem naukowym zajmującym się interesującą ich tematyką. Dzięki temu staną się częścią grupy biorącej udział w skomplikowanych projektach badawczych, a co za tym idzie, będą mogli szybciej zdobywać cenne doświadczenie i ubogacać swoją wiedzę.

Nagrody Narodowego Centrum Nauki dla użytkowników zasobów obliczeniowych ACK Cyfronet AGH

NARODOWE CENTRUM NAUKI

Nagroda Narodowego Centrum Nauki to wyróżnienie dla młodych uczonych za znaczące osiągnięcia naukowe, dokonane w ramach badań podstawowych prowadzonych w polskiej jednostce naukowej, udokumentowane publikacjami afiliowanymi w takiej jednostce. Nagrodę ustanowiła Rada Narodowego Centrum Nauki w lutym 2013 r. Wyróżnienie przyznawane jest w trzech obszarach badawczych: 1) nauki humanistyczne, społeczne i o sztuce, 2) nauki o życiu, 3) nauki ścisłe i techniczne. Wielokrotnie laureaci tych nagród podkreślają rolę zasobów obliczeniowych Cyfronetu w osiągnięciu wyników badań.

Nagroda Narodowego Centrum Nauki dla dr hab. Sebastiana Glatta z Małopolskiego Centrum Biotechnologii UJ*

W dniu 6 października 2021 laureatem Nagrody w obszarze nauk o życiu został dr hab. Sebastian Glatt. Nagroda została przyznana za następujące osiągnięcia: ustalenie struktury przestrzennej i mechanizmu działania kompleksu białkowego Elongator; opisanie mechanizmu modyfikacji tRNA przez kompleks elongatora; strukturalna i funkcjonalna charakterystyka eukariotycznych kompleksów modyfikujących tRNA; określenie struktury i funkcji składników kompleksu elongatora oraz powiązanie ich mutacji z powstawaniem zaburzeń neurorozwojowych i neurodegeneracyjnych.

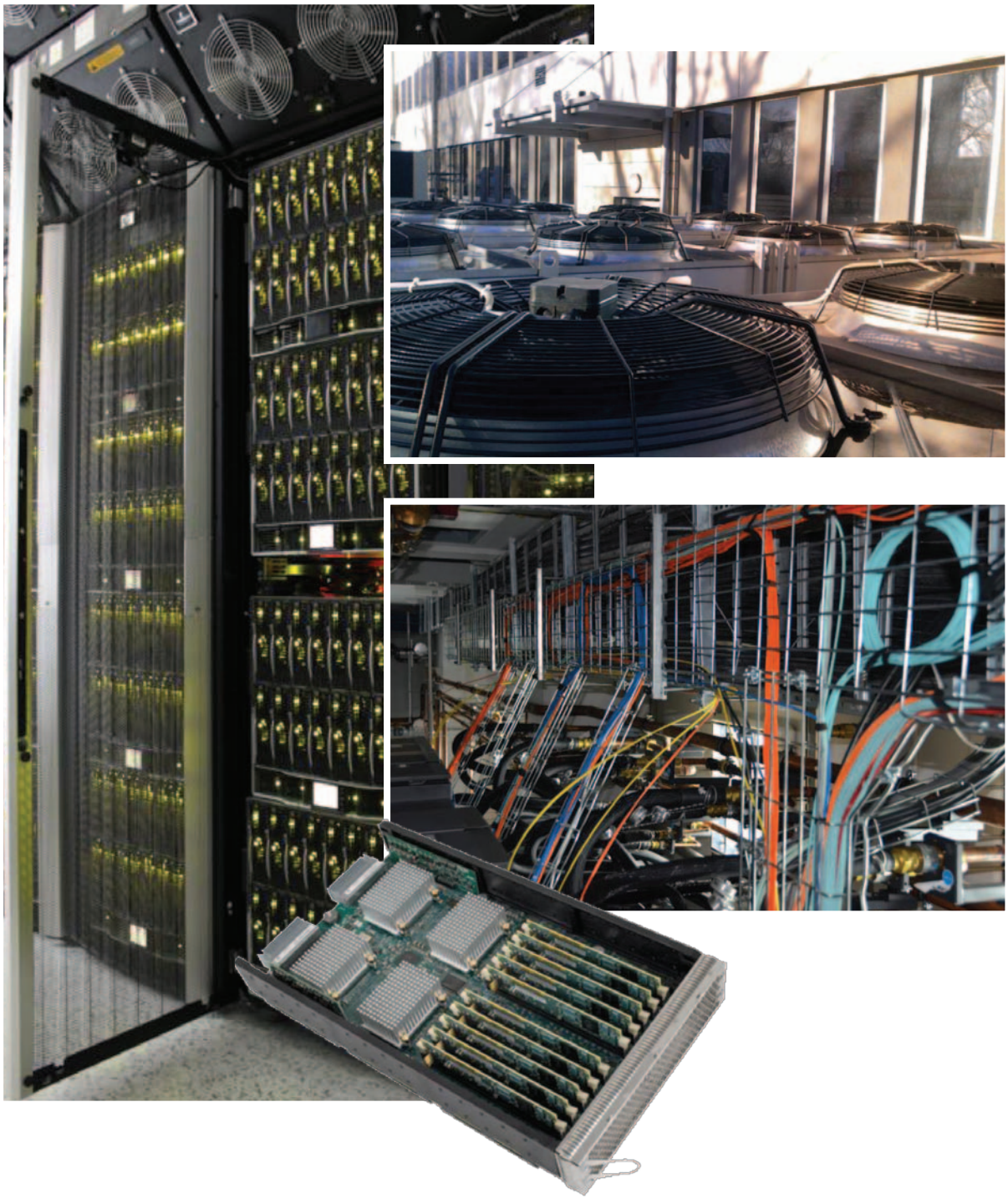
Dr hab. Sebastian Glatt ukończył studia i obronił doktorat na Uniwersytecie Wiedeńskim. Prowadził badania w Europejskim Laboratorium Biologii Molekularnej w Heidelbergu. Od 2015 pracuje w MCB UJ, kieruje także zespołem laboratoryjnym Cryo-EM przy Krajowym Centrum Kriomikroskopii Elektronowej. Zainicjował bardzo owocną współpracę z naukowcami z innych ośrodków w Polsce, Niemczech, Francji, Belgii, Szwajcarii i Australii.

– *W mojej grupie badawczej specjalizujemy się w badaniu trójwymiarowego kształtu białek i kwasów nukleinowych za pomocą mikroskopii krioelektronowej i krystalografii. Wyróżnienia, takie jak Nagroda NCN, otrzymuje się indywidualnie, ale pracuje na nie cały zespół* – mówi dr hab. Sebastian Glatt.

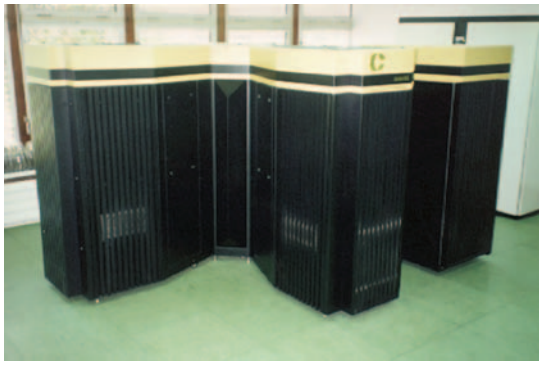
Dr hab. Sebastian Glatt wspierał swoje badania obliczeniami na udostępnianym przez Cyfronet superkomputerze Prometheus, w ramach grantu obliczeniowego *plgribosomes2021*: „Eukaryotic Ribosomes".

Serdecznie gratulujemy Laureatowi i życzymy dalszych sukcesów!

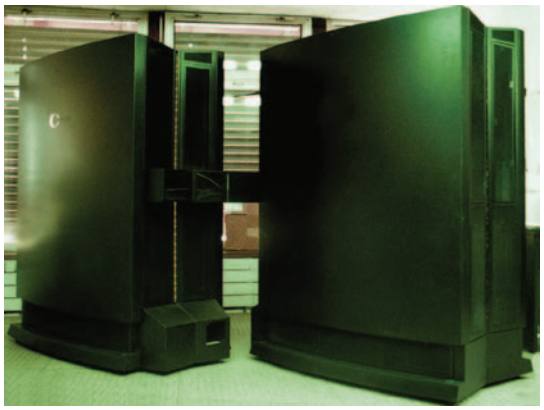
**Materiały prasowe za https://www.ncn.gov.pl/*



CDC CYBER 72



Convex C3840



Exemplar SPP1600/XA

1973 Powstanie CYFRONETu

1975 Instalacja komputera CDC CYBER 72 w Centrum

1990 Instalacja pierwszego w Krakowie węzła sieci EARN / BITNET (komputer IBM 4381)

1991 Instalacja Convex 120 – pierwszego komputera wektorowego w Europie Środkowo-Wschodniej.

Pierwsze internetowe połączenie międzymiastowe z Warszawą.

Początek budowy Miejskiej Sieci Komputerowej w Krakowie

1994 Uruchomienie połączenia internetowego z Warszawą o prędkości 2 Mb/s

1996 Komputer Exemplar SPP1600/XA umieszczony na prestiżowej liście TOP500.

Instalacja pierwszej automatycznej biblioteki taśmowej ATL 2640

1997 Uruchomienie podsieci komunikacyjnej ATM w Miejskiej Sieci Komputerowej.

Przyłączenie Centrum do krajowej sieci POL-34

1998 Instalacja komputera SGI Origin2000

2000 Zwiększenie przepustowości połączenia sieciowego Centrum do 155 Mbps

2002 Instalacja komputera RackSaver PC w ramach projektu CrossGrid

2003 Instalacja pierwszego w Polsce komputera HP Integrity SuperDome

2005 Instalacja macierzy dyskowej HP Storage Works XP12000.

Zwiększenie przepustowości połączenia sieciowego Centrum do 622 Mbps

2006 Instalacja macierzy dyskowej HP Storage Works EVA 8000 i superkomputera SGI ALTIX 3700 (Baribal) o teoretycznej mocy obliczeniowej 0,8 TFlops

2007 Podpisanie umowy Konsorcjum PL-Grid dotyczącej ogólnopolskiej infrastruktury gridowej PL-Grid.

Instalacja superkomputera SGI ALTIX 4700 z modułem akceleracyjnym SGI RASC.

Instalacja serwerów IBM BladeCenter HS21 (6,2 TFlops).

Instalacja macierzy dyskowej HP EVA 8100

2008 Rozbudowa konfiguracji superkomputera SGI ALTIX 3700 (do 1,5 TFlops).

Rozpoczęcie eksploatacji łącz sieci szkieletowej w technologii 10 Gbps Ethernet.

Połączenie MSK w kierunku Warszawy i Bielska-Białej łączami sieci PIONIER, każde o przepływności 2x10 Gbps.

Instalacja superkomputera Zeus (HP Cluster Platform 3000 BL), posiadającego 2 048 rdzeni

2009 Rozpoczęcie projektu „Infrastruktura Informatycznego Wspomagania Nauki w Europejskiej Przestrzeni Badawczej – PLGrid"

2010 Po rozbudowie konfiguracji do 9 544 rdzeni Intel Xeon superkomputer Zeus zostaje wpisany na listę TOP500 na 161 pozycji

2011 Wdrożenie wydajnego sieciowego systemu współdzielenia plików dla infrastruktury obliczeniowej, bazującego na rozwiązaniach firmy Hitachi Data Systems.

Całkowita pojemność zasobów dyskowych przekroczyła 2 PB.

Rozbudowa konfiguracji superkomputera Zeus do 12 032 rdzeni Intel Xeon.

Zeus zajmuje 80 miejsce na liście TOP500

2012 Rozpoczęcie realizacji projektu PLGrid Plus – dziedzinowo zorientowane usługi i zasoby w infrastrukturze PL-Grid.

W kwietniu, ScaleMP, wiodący dostawca wirtualnych rozwiązań dla superkomputerów, ogłosił „Zeusa-vSMP" największym systemem v-SMP w Europie.

Zeus w pierwszej setce rankingu TOP500.

Połączenie MSK w kierunku Rzeszowa łączem sieci PIONIER o przepływności 2x10 Gbps

SGI Origin2000

SGI ALTIX 3700

HP Cluster Platform 3000 BL

Medal jubileuszowy

Hala Maszyn

Superkomputer Prometheus

2013 Po rozbudowie do 25 468 rdzeni, superkomputer Zeus osiągnął teoretyczną moc obliczeniową 374 TFlops.

Wybicie medalu z okazji jubileuszu 40-lecia ACK Cyfronet AGH

2014 Oddano do użytku nową Halę Maszyn.

Rozpoczęcie realizacji dwóch nowych projektów: PLGrid NG i PLGridCore.

Połączenie MSK w kierunku Katowic łączem sieci PIONIER o przepływności 2x10 Gbps

2015 Instalacja superkomputera Prometheus (41 472 rdzenie), który na liście TOP500 zajmuje 49 miejsce (edycja lipcowa), a następnie (po rozbudowie do 53 568 rdzeni) 38 miejsce (edycja listopadowa).

Po raz pierwszy w historii polskiej informatyki dwa superkomputery z jednego ośrodka (Zeus i Prometheus) znalazły się jednocześnie na prestiżowej liście TOP500.

Oddano do użytku nowe, zapasowe centrum danych.

Rozpoczęto realizację projektów INDIGO-DataCloud, EGI-Engage, EPOS-IP i PRACE-4IP

Łączność pomiędzy Centrami Komputerów Dużej Mocy w Polsce (Gdańsk, Kraków, Poznań, Warszawa i Wrocław) realizowana jest z przepustowością 2x100 Gbps

2016 Prometheus zajmuje 48 miejsce (edycja czerwcowa) i 59 miejsce (edycja listopadowa) na liście TOP500

2017 Prometheus zajmuje 71 miejsce (edycja czerwcowa) i 77 miejsce (edycja listopadowa) na liście TOP500.

Dalszy dynamiczny rozwój Centrum, w tym powstanie 6 nowych laboratoriów.

Rozpoczęto realizację projektów Sat4Envi, Gliomed, EPOS-PL oraz eXtreme DataCloud

2018 Prometheus (53 604 rdzenie, 2,4 PFlops) zajmuje 103 miejsce (edycja czerwcowa) i 131 miejsce (edycja listopadowa) na liście TOP500.

Rozpoczęcie realizacji projektów EOSC-Hub i PRIMAGE

2019 Cyfronet reprezentuje Polskę w konsorcjum LUMI, złożonym z ośmiu krajów, które będą wspólnie budować jeden z najszybszych europejskich superkomputerów.

Prometheus zajmuje 174 miejsce (edycja czerwcowa) i 241 miejsce (edycja listopadowa) na liście TOP500.

Prezentacja Cyfronetu na stoisku wystawowym w trakcie konferencji ISC'19.

Cyfronet udostępnia nowy system obliczeniowy dla badań z użyciem metod sztucznej inteligencji o mocy ponad 4 PFlops dla operacji tensorowych i 256 TFlops dla standardowych obliczeń.

Rozpoczęto realizację projektów PRACE-LAB, PRACE-6IP, SANO, EOSC-Synergy i EOSC Enhance

2020 Wśród strategicznych infrastruktur wpisanych w styczniu 2020 r. na Polską Mapę Infrastruktury Badawczej znajdują się dwa przedsięwzięcia zgłoszone przez ACK Cyfronet AGH jako inicjatora i koordynatora konsorcjum PLGrid: *Narodowa Infrastruktura Superkomputerowa dla EuroHPC* oraz *Narodowa Infrastruktura Chmurowa PLGrid dla EOSC*.

Prometheus (53 748 rdzeni, 2,7 PFlops) zajmuje 288 miejsce (edycja czerwcowa) i 324 miejsce (edycja listopadowa) na liście TOP500.

Superkomputer Prometheus wspomaga naukowców w walce z koronawirusem.

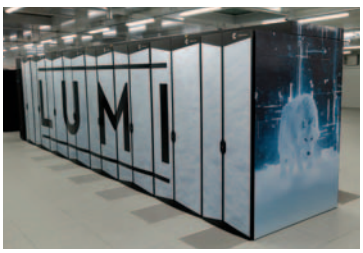
Rozpoczęto realizację projektów EPOS PL +, PRACE-LAB2, EPOS SP, PROTEUS-RS i EUROCC

2021 Uruchomiono superkomputer Ares o mocy obliczeniowej 4,0 PFlops.

Prometheus zajmuje 373 i 440 miejsce, a Ares 216 i 267 miejsce na liście TOP500.

Uruchomiono nową wersję strony internetowej ACK Cyfronet AGH.

Rozpoczęto realizację projektów EuroHPC PL, PIONIER-LAB, KMD3, AGH – PANDA3, EGI ACE, EOSC Future i FINDR



2022 Uruchomiono superkomputer Athena o mocy obliczeniowej 7,7 PFlops.

Po raz pierwszy w historii polskiej informatyki trzy superkomputery z jednego polskiego centrum obliczeniowego (Athena, Ares i Prometheus) na liście TOP500.

Prometheus (53 748 rdzeni, 2,7 PFlops) po raz 15-ty na liście TOP500: 475 miejsce.

Ares (37 824 rdzenie, 4,0 PFlops) po raz 3-ci na liście TOP500: 290 miejsce.

Athena (6 144 rdzeni, 7,7 PFlops) po raz 1-szy na liście TOP500: 105 miejsce.

Inauguracja superkomputera LUMI.

Rozpoczęto realizację projektów: EUMaster4HPC, FAIRCORE4EOSC, EuroScienceGateway, InterTwin, DT-GEO i Geo-INQUIRE.

Inauguracja obchodów 50-lecia ACK Cyfronet AGH

KAMPUS AGH

LEGENDA

1. Rektorat
2. Wydział Górnictwa i Geoinżynierii
3. Wydział Inżynierii Metali i Informatyki Przemysłowej
4. Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii Biomedycznej
5. Wydział Informatyki, Elektroniki i Telekomunikacji
6. Wydział Inżynierii Mechanicznej i Robotyki
7. Wydział Geologii, Geofizyki i Ochrony Środowiska
8. Wydział Geodezji Górniczej i Inżynierii Środowiska
9. Wydział Inżynierii Materiałowej i Ceramiki
10. Wydział Odlewnictwa
11. Wydział Metali Nieżelaznych
12. Wydział Wiertnictwa, Nafty i Gazu
13. Wydział Zarządzania
14. Wydział Energetyki i Paliw
15. Wydział Fizyki i Informatyki Stosowanej
16. Wydział Matematyki Stosowanej
17. Wydział Humanistyczny
18. Akademickie Centrum Materiałów i Nanotechnologii AGH
19. Centrum Energetyki AGH
20. Biblioteka Główna
21. Szkoła Ochrony i Inżynierii Środowiska im. Walerego Goetla
22. Studium Języków Obcych
23. Studium Wychowania Fizycznego i Sportu
24. Basen AGH
25. Centrum e-Learningu AGH
26. **Akademickie Centrum Komputerowe CYFRONET AGH**
27. Uczelniane Centrum Informatyki
28. Dział Nauczania
29. Centrum Studentów Zagranicznych
30. Centrum Dydaktyki AGH
31. Uczelniana Komisja Rekrutacyjna
32. Miasteczko Studenckie AGH
33. Uczelniana Rada Samorządu Studentów
34. Centrum Karier AGH
35. Centrum Transferu Technologii
36. Dział Współpracy z Administracja i Gospodarką
37. Krakowskie Centrum Innowacyjnych Technologii INNOAGH
38. Centrum Obsługi Projektów
39. Dział Współpracy z Zagranicą
40. Biuro ds. Osób Niepełnosprawnych
41. Muzeum AGH
42. Muzeum Geologiczne WGGiOŚ AGH
43. Wydawnictwa AGH
44. Akademickie Centrum Kultury Klub STUDIO
45. Klub Studencki Gwarek
46. Klub Studencki Zaścianek
47. Klub Studencki Karlik
48. Klub Studencki Filutek

ISBN 978-83-61433-41-5

Akademickie Centrum Komputerowe
CYFRONET AGH
ul. Nawojki 11

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

cyfronet@cyfronet.pl
www.cyfronet.pl