



*Polska Infrastruktura
Informatycznego Wspomagania Nauki
w Europejskiej Przestrzeni Badawczej*

GridSpace

Wirtualne Laboratorium w PL-Grid: GridSpace

**M. Bubak, B. Baliś, T. Bartyński, E. Ciepiela,
W. Funika, T. Gubała, D. Hareźlak,
M. Kasztelnik, J. Kocot, M. Malawski,
J. Meizner, P. Nowakowski, and K. Rycerz**

Akademickie Centrum Komputerowe Cyfronet AGH



- ◆ **PL-Grid**
- ◆ **Przykładowe obliczenia w PL-Grid**
- ◆ **Motywacja i cel wirtualnego laboratorium**
- ◆ **Przykładowy problem z życia naukowca wzięty**
- ◆ **Przykładowy problem – rozwiązany z pomocą GridSpace**
- ◆ **Przykładowy problem – demonstracja (film)**

PL-Grid: Infrastruktura - sprzęt

- ◆ ACK Cyfronet AGH (obecnie)
 - ◆ Dostępnych prawie **10000 rdzeni** (ok. **104 TFLOPs**)
 - ◆ Ponad **1500 TB pamięci masowych**
 - ◆ **Klaster Zeus** najmocniejszym komputerem w Polsce (**Top500 miejsce 84, Listopad 2010**, poprzednio miejsce 161)
- ◆ Projekt PL-Grid docelowo:
 - ◆ Moc obliczeniowa: ok. **215 TFLOPs**
 - ◆ Ponad **2500 TB pamięci masowych**
- ◆ Infrastruktura niezbędna do utrzymania zasobów w stanie produkcyjnym **7 dni w tygodniu**
 - ◆ Efektywna klimatyzacja zasobów
 - ◆ System zabezpieczeń na wypadek awarii zasilania

Podziękowania: Mariusz Sterzel, Cyfronet

PL-Grid: Infrastruktura - oprogramowanie

- ◆ Pakiety oprogramowania naukowego (komercyjne i darmowe), narzędzia typu kompilatory, biblioteki numeryczne, MPI, ... – wszystko dostępne w jeden zunifikowany sposób
- ◆ Zestaw zaawansowanych narzędzi do organizacji eksperymentów obliczeniowych w środowisku rozproszonym (InsilicoLab, GridSpace2, g-Eclipse, VINE Toolkit, Migrating Desktop)
- ◆ Pomoc w zrozumieniu zagadnień związanych z uruchamianiem aplikacji naukowych na rozległych zasobach obliczeniowych
- ◆ Wsparcie przy projektowaniu własnych aplikacji naukowych i ich wdrażaniu na infrastrukturze PL-Grid
- ◆ Systemy: Portal Użytkownika, Helpdesk, Szkolenia w Blackboard
- ◆ System zarządzania grantami obliczeniowymi (faza wdrożenia)
- ◆ Bazy danych użytkowników (faza wdrożenia)

Podziękowania: Mariusz Sterzel, Cyfronet

Obliczenia w PL-Grid: Chemia kwantowa

- ◆ Główne zadanie – obliczenia struktury elektronowej molekuł
 - ◆ wykorzystywane do:
 - badania oddziaływań pomiędzy cząsteczkami
 - modelowania reakcji i procesów chemicznych oraz katalitycznych
 - badania materiałów o znaczeniu przemysłowym i poszukiwania nowych o zadanych właściwościach, np. polimerów, materiałów magazynujących energię itp..
 - ◆ główne dziedziny zastosowania: chemia, fizyka i biologia

Przykład: obliczenia struktury elektronowej i drgań normalnych cząsteczki fullerenu C₆₀ z atomem miedzi

Obliczenia pakietem Turbomole. Dostępne również ADF, Gaussian i inne ogólnodostępne

Podziękowania: Mariusz Sterzel, Cyfronet

Obliczenia w PL-Grid: Biologia i nauki pokrewne

- ◆ Główny użytkownik infrastruktury, zastosowania:
 - ◆ przewidywanie struktury białek,
 - ◆ poszukiwanie nowych leków,
 - ◆ analiza krótkich sekwencji DNA,
 - ◆ badanie własności enzymatycznych białek,
 - ◆ modelowanie procesu zwijania (foldowania) białek
 - ◆ badania zjawisk elektrycznych w komórkach serca
 - ◆ przewidywania własności biopolimerów i ich zastosowania
 - ◆ opis aktywności komórek nerwowych mózgu i siatkówki oka
 - ◆ projektowanie i wdrażanie systemów pomiarowych na potrzeby neurobiologii
 - ◆ badania nad gruźlicą – własności inhibitorów primazy i ligazy DNA
 - ◆ ...

Podziękowania: Mariusz Sterzel, Cyfronet

Obliczenia w PL-Grid: Biologia i nauki pokrewne cd.

- ◆ Najaktywniejszy użytkownik – **badania aktywności antybiotyków przeciwgrzybiczych** (Politechnika Gdańska) – ponad **165 lat CPU**
- ◆ Zastosowanie nowoczesnego interfejsu sieciowego (Infiniband) wraz ze specjalnie przygotowanymi wersjami oprogramowania pozwoliło **przyspieszyć obliczenia ponad 5x**. Dla innych pakietów oprogramowania podobne usprawnienia są w trakcie wdrożenia

Podziękowania: Jacek Czub, Anna Neumann, PG

Obliczenia w PL-Grid: Fizyka Wysokich Energii

Wspieramy wszystkie eksperymenty Wielkiego Zderzacza Hadronów, projektu mającego na celu odkrycie „boskiej cząstki” – bozonu Higgs’a

Obliczenia w PL-Grid: Projektowanie układów scalonych

- ◆ Efektywne modelowanie poprawności działania chipu możliwe dzięki wykorzystaniu wielu rdzeni jednocześnie

Przykład:
wykorzystanie
pakietu
komercyjnego
Cadence
(licencja
własna
użytkowników)

Podziękowania: Michał Dwużnik, Krzysztof Świętek, AGH

Obliczenia w PL-Grid: Astronomia

- ◆ **Cherenkov Telescope Array (CTA)**
 - ◆ Inicjatywa mająca na celu budowę sieci nowej generacji naziemnych detektorów promieniowania γ pracujących w zakresie **10 GeV do 100 TeV**
 - ◆ Projekt został zakwalifikowany jako kluczowy dla nauki w Europie przez ESFRI
-
- ◆ Obecnie jest w końcowej części fazy rozruchowej
 - ◆ Dane zbierane przez dostępne przyrządy: CANGAROO, H.E.S.S. , MAGIC, MILAGRO oraz VERITAS są **przechowywane na infrastrukturze PL-Grid w ACK Cyfronet AGH** i na bieżąco analizowane przez polskich astronomów



Materiały za: G. Lamana, D. Torres,
CTA

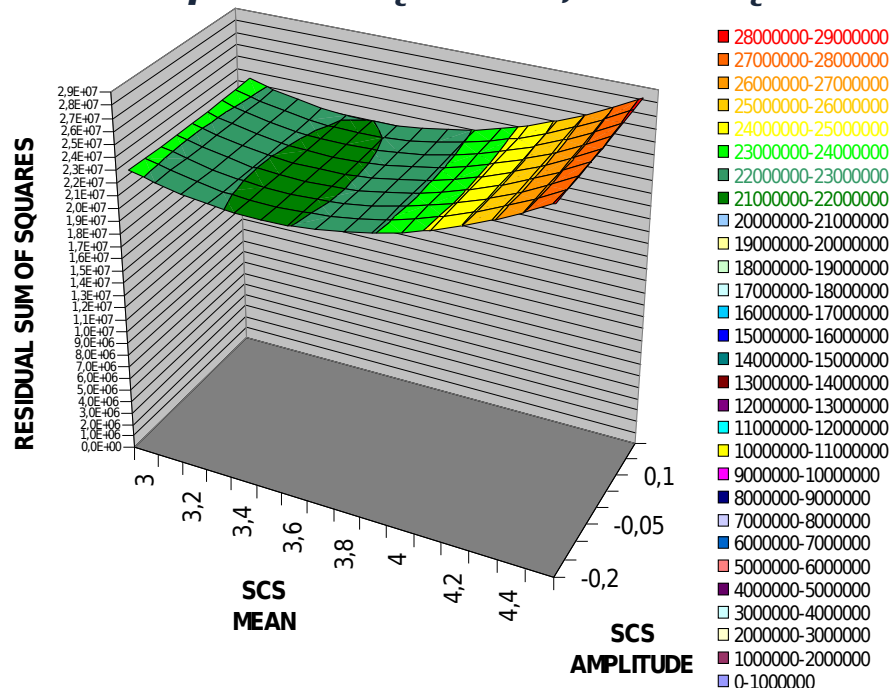
Obliczenia w PL-Grid: Fizyka – projekt nEDM

- ◆ Projekt mający na celu bardzo **dokładne pomiary momentu dipolowego neutronu** o rozdzielczości kilku 10^{-28} e.cm
- ◆ Ponad 12 partnerów
- ◆ Strategia: „ultra-zimne” neutrony przechowywane w próżni w temperaturze pokojowej
- ◆ Obecnie realizacja fazy I i II projektu – test komponentów i budowa aparatu pomiarowego
- ◆ **Organizacja wirtualna nEDM w PL-Grid umożliwia współpracę i wymianę danych pomiędzy partnerami z Polski i Europy**
- ◆ Wymagane zasoby dyskowe – ok. 20 TB

Obliczenia w PL-Grid: Badania sezonowych zmian zawartości komórek somatycznych (SCS) w mleku krów rasy polskiej holsztyńsko-fryzyjskiej

- ◆ SCC i SCS (logarytm z SCC) – powszechnie używane w hodowli bydła jako wskaźniki stanu zdrowotności wymienia
- ◆ Średnia wartość SCC [ml^{-1}]
 - ◆ 1. laktacja – do 145 tys.
 - ◆ 2. laktacja – do 200 tys.
 - ◆ 3. laktacja – do 240 tys.
 - ◆ **Ponad 500 tys.: oznaka choroby**
- ◆ Badania: analiza statystyczna SCS na podstawie ok. 12 mln danych dot. ok. 900 000 krów
- ◆ funkcja okresowa (stały okres 12 miesięcy, parametry: średnia, amplituda i przesunięcie fazy)

Wyniki (suma kwadratów błędu) dla przesunięcia o 1,5 miesiąca



Wniosek: Wartość średnia SCS (z wykresu) wynosi ok. 3,5; amplituda ok. 0,1.

Podziękowania: Ewa Ptak, Maciej Gierdziewicz, WHBZ UR, Kraków

Motywacje i cel wirtualnego laboratorium

- ◆ Umożliwić naukowcom **tworzenie, uruchamianie, zarządzanie, dzielenie się oraz udostępnianie programów** (nazywanych przez nas *eksperymentami*) złożonych z etapów pisanych w różnych językach (programowania jak i domenowych), **które realizują całościowo pewną metodę naukową** prowadzącą od danych wejściowych poprzez preprocessing, symulacje obliczeniowe, analizy, postprocessing aż po wizualizację i opracowanie rezultatów
- ◆ Ułatwić (e-)naukowcom przeprowadzającym w ramach swych badań obliczeń z wykorzystywaniem zasobów centrum komputerowego:
 - ◆ Klatry komputerowe (już dostępne)
 - ◆ Pakiety oprogramowania (już dostępne, lub instalowane na bieżąco na życzenie użytkowników)
 - ◆ Inne „aktywa”, np. licencje na używanie oprogramowania
- ◆ Zwiększyć efektywność pracy naukowców i zespołów naukowych:
 - ◆ Uwolnić użytkowników od powtarzania czynności, które mogą (bez żadnych negatywnych implikacji) zostać zautomatyzowane
 - ◆ Umożliwić bardziej interaktywny sposób wykorzystania klastrów, oprogramowania itp.
 - ◆ Sprawić by raz napisany program (fragment programu) mógł być z łatwością wykorzystany w innych programach i/lub przez inne osoby w innym kontekście
- ◆ Zaoferować ogólne rozwiązanie adresowane do szerokiego zakresu domen zastosowań
- ◆ Stworzyć jedno łatwo osiągalne i kompleksowe miejsce pracy - **portal internetowy**

Motto

- ◆ Sprawić, by aplikacje naukowe było tak łatwo dostępne jak strony WWW...
- ◆ by uruchamianie aplikacji naukowych było tak proste jak używanie stron WWW...
- ◆ by tworzenie aplikacji naukowych było tak proste jak tworzenie stron WWW...
- ◆ ...niezależnie od tego jak ogromna i złożona jest infrastruktura komputerowa je uruchamiająca.

Przykładowy problem z życia naukowca wzięty

- ◆ Zespół naukowy z Wydziału Chemii Uniwersytetu Jagiellońskiego
- ◆ Chciano zastosować metodę, którą znaleziono opublikowaną w wiodącym journalu, ale poza opisem brak programu, który by ją realizował
- ◆ Setki obrazów SEM do przetworzenia:
 - ◆ Identyfikacja porów w nanomateriałach
 - ◆ Odrzucenie szumów i nieistotnych porów, proste statystyki
 - ◆ Obliczenie stopnia regularności struktury (im większy tym lepszy)
 - ◆ Zapis wartości różnych miar regularności
 - ◆ Wizualizacja metryk na wykresach i diagramach
 - ◆ Analiza „na oko” i wybór podzbioru interesujących wartości
 - ◆ Bardziej dogłębna wizualizacja wybranego podzbioru
- ◆ Jeden komputer PC, brak odpowiedniego oprogramowania
- ◆ Zespół 3 naukowców, kilku studentów do pomocy
- ◆ Zwrócili się o pomoc do ACK Cyfronet

Przykładowy problem – rozwiązany przy pomocy GridSpace

- ◆ Wybór odpowiednich języków/narzędzi do odpowiednich zadań (języki *general-purpose* vs *domain-specific*) które już są dostępne na klastrze Zeus w Cyfronecie:
 - ◆ Identyfikacja porów w nanomateriałach – JNano
 - ◆ Odrzucenie szumów i nieistotnych porów, proste statystyki – Python
 - ◆ Obliczenie stopnia regularności struktury – Mathematica
 - ◆ Zapis wartości różnych miar regularności – Mathematica
 - ◆ Wizualizacja metryk na wykresach i diagramach – GnuPlot
 - ◆ Analiza „na oko” i wybór podzbioru interesujących wartości – interakcja z człowiekiem
 - ◆ Ekstrakcja wybranych wartości – Bash
 - ◆ Bardziej dogłębna wizualizacja wybranego podzbioru – GnuPlot
- ◆ Wejście na GridSpace2 Experiment Workbench (portal) i wybór Zeusa jako miejsce gdzie eksperyment zostanie uruchomiony (*Experiment Host*)
 - ◆ Pisanie, uruchamianie, udoskonalanie, uruchamianie itd. kolejnych fragmentów eksperymentu w sposób eksploracyjny aż do osiągnięcia zadowalającego efektu
 - ◆ Zapisywanie eksperymentu jako zwykłego pojedynczego pliku i udostępnianie go zespołowi
 - ◆ Eksperyment ma przypisany adres URL – wystarczy go przesłać, a adresat będzie mógł go uruchomić

- ◆ Wejście na stronę <https://gs2.plgrid.pl> - Experiment Workbench (Portal)
- ◆ Wejście na ui.cyfronet.pl – Experiment Host (maszyna UI w centrum komputerowym)
- ◆ Przeprowadzenie metody krok po kroku...
- ◆ Publikacja wyniku eksperymentu
- ◆ Opisanie eksperymentu
- ◆ Publikacja eksperymentu

- ◆ Przykładowe eksperymenty stale tworzone i dołączane do portfolio <http://chomik.cyfronet.pl/trac/gridspace2#ExampleExperiments>
- ◆ Prezentacje przykładowych obliczeń dziedzinowych
 - ◆ Protein Folding
 - ◆ Water Simulation
 - ◆ Parallel FASTA
 - ◆ Parsing chemistry results with CCLib
- ◆ Prezentacje funkcjonalności platformy
 - ◆ Web GUI Demo
 - ◆ Playing with Twitter
 - ◆ ...

Zapraszamy do współpracy!

- ◆ Stań się użytkownikiem PL-Grid
 - ◆ <https://portal.plgrid.pl>
- ◆ Spróbuj demo GridSpace2 (dostępne dla użytkowników PL-Grid)
 - ◆ <https://gs2.cyfronet.pl>
- ◆ Użyj GridSpace2 jako oficjalnej usługi PL-Grid (dostępne dla użytkowników PL-Grid)
 - ◆ <https://gs2.plgrid.pl>
- ◆ Zasięgnij pełnej informacji o technologii GridSpace
 - ◆ <http://dice.cyfronet.pl/gridspace>
- ◆ Dowiedz się więcej o naszym zespole, jego dokonaniach, aktualnych przedsięwzięciach i planów na przyszłość
 - ◆ <http://dice.cyfronet.pl>
- ◆ Oferujemy pomoc w nauce GridSpace i zastosowywania go do konkretnych problemów