

Jak znaleźć igłę w stogu siana

Rola obliczeń numerycznych w eksperymentach fizyki wysokich energii

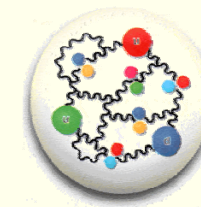
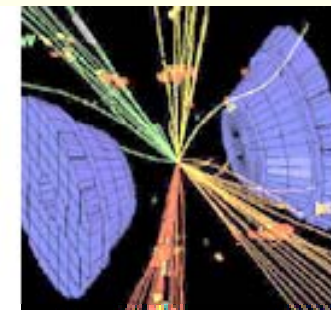
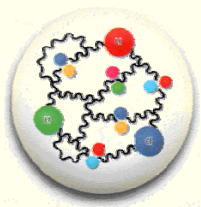
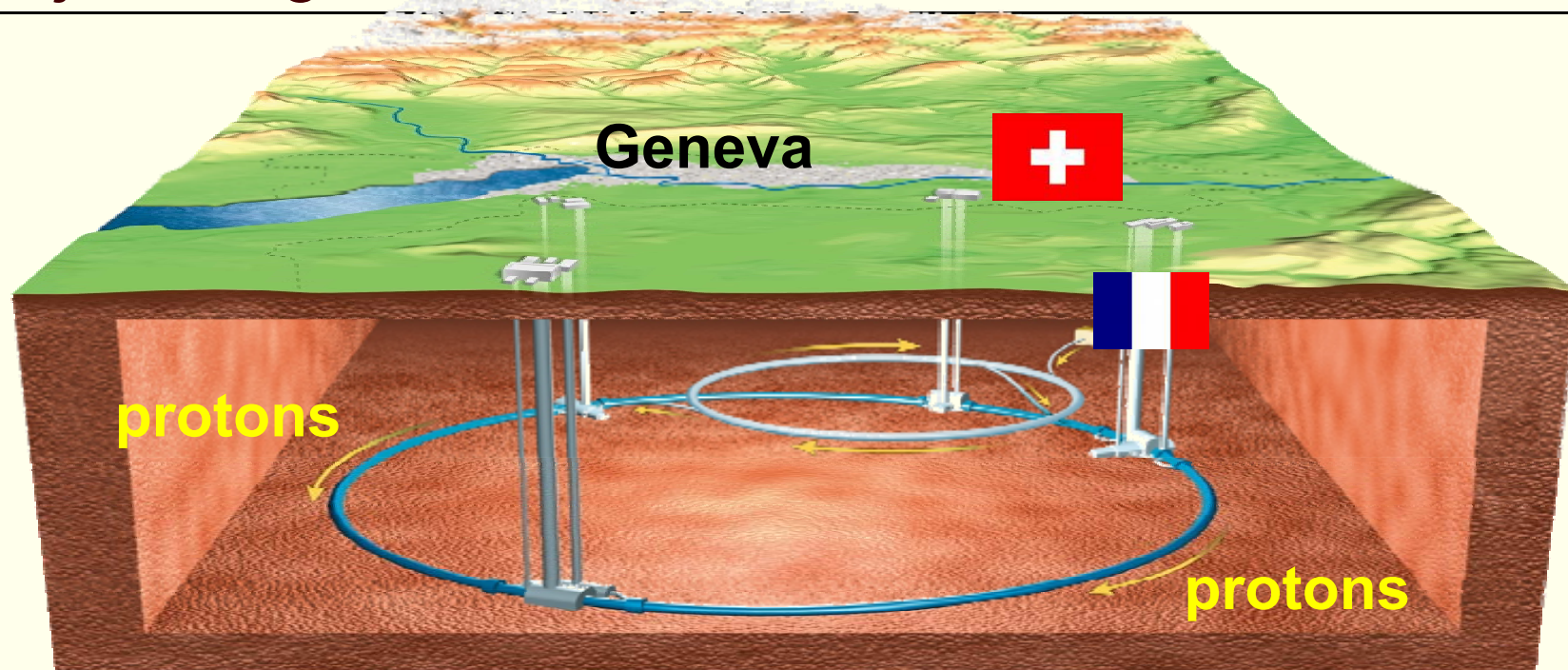
Mariusz Witek
IFJ PAN Kraków

Plan

- Co jest igłą a co stogiem siana?
- Projektowanie eksperymentu
- Zbieranie danych
- Opracowywanie danych



Co jest stogiem siana? - LHC



$$14 \text{ TeV} = 14 \cdot 10^{12} \text{ eV}$$

Co jest stogiem siana?

- Przekięcia pęków wiązek co 25 nanosekund (40 MHz)
- $10 * 10^6$ zderzeń w ciągu 1 sekundy widocznych w LHC-b!

Dane przepływające przez eksperyment

	Sekunda	Dzień	Rok
I. przypadków	10^7	10^{12}	10^{14}
Ilość danych	300 GB	30 PT	3 EB ($3*10^{18}$)
I. dysków (300 GB)	1	100 tys	10 mln

- Dla eksperymentów ATLAS i CMS jest znacznie gorzej
 - ATLAS 2PB/sek \equiv 3 mln CD/s
- Zapis wszystkich przypadków jest niemożliwy
- Na każdy przypadek należy jednak „zerknąć” i zdecydować czy jest interesujący.
 - Do „przeglądania” przypadków służy **system wyzwalania** czyli filtracji przypadków (tzw. **tryger**).



1 Megabyte (1MB)
A digital photo

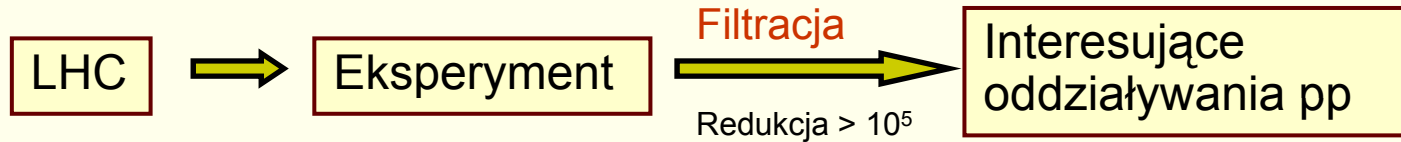
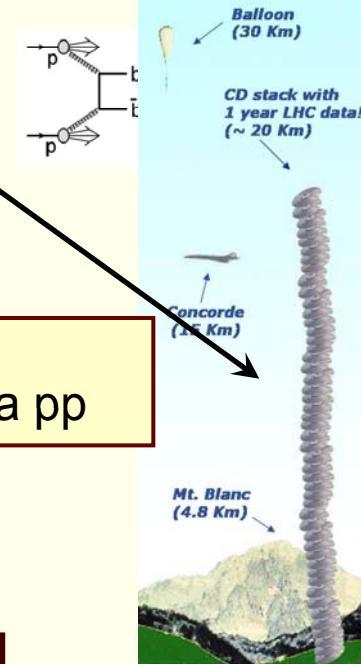
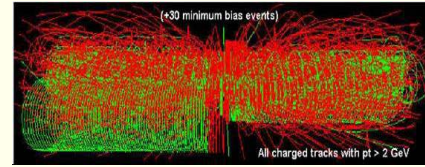
1 Gigabyte (1GB)
= 1000MB
A DVD movie

1 Terabyte (1TB)
= 1000GB
World annual book production

1 Petabyte (1PB)
= 1000TB
Annual production of one LHC experiment

1 Exabyte (1EB)
= 1000 PB
World annual information production

Co jest stogiem siana?



15 PetaBytes / rok przefiltrowanych danych z czterech eksperymentów na LHC



Co jest igłą ?

- Większość zderzeń to zwykłe, dobrze znane oddziaływania.
- Te interesujące które niosą informacje o nieznanym zjawiskach są bardzo rzadkie.
 - Typowo 1 przypadek na 100 mln!
- Jak „zauważyć” taki przypadek w aparaturze i zapisać go na taśmę gdy mamy tak duży strumień danych \equiv tak mało czasu na analize on-line?

Czego szukamy w eksperymencie LHC-b

Łamanie symetrii CP



Parzystość CP

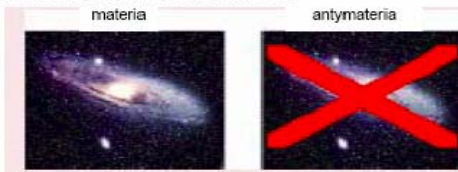
- jednoczesne wykonanie operacji sprzężenia ładunkowego C i inwersji przestrzennej P, tzn. operacji $q \rightarrow -q$ oraz $\vec{r} \rightarrow -\vec{r}$. CP przeprowadza cząstkę w jej antycząstkę o przeciwnym pędzie i skrętności
- niezachowanie parzystości P (1956) oraz parzystości C (1957) w oddziaływaniach słabych

Łamanie parzystości kombinowanej CP

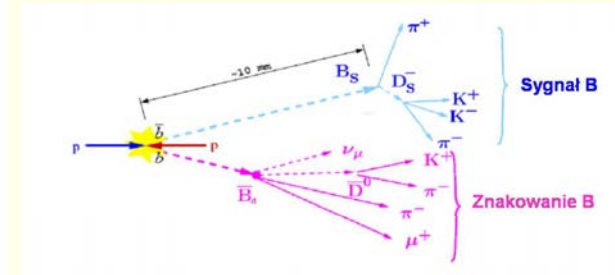
- łamanie symetrii CP w rozpadach długożyciowych mezonów K^0 (1964)

Asymetria materia-antymateria

- wszystkie obserwacje wskazują, że obecny wszechświat składa się głównie z materii
- hipoteza Sacharowa (1967)
 - jednym z 3 warunków powstania asymetrii materia-antymateria jest **łamanie CP**



Badanie łamania CP od 40 lat (zjawiska subtelne)
 LHCb: duże próbki danych, rozszerzenie obszaru badań



Łamanie CP w układzie mezonów pięknych (B)

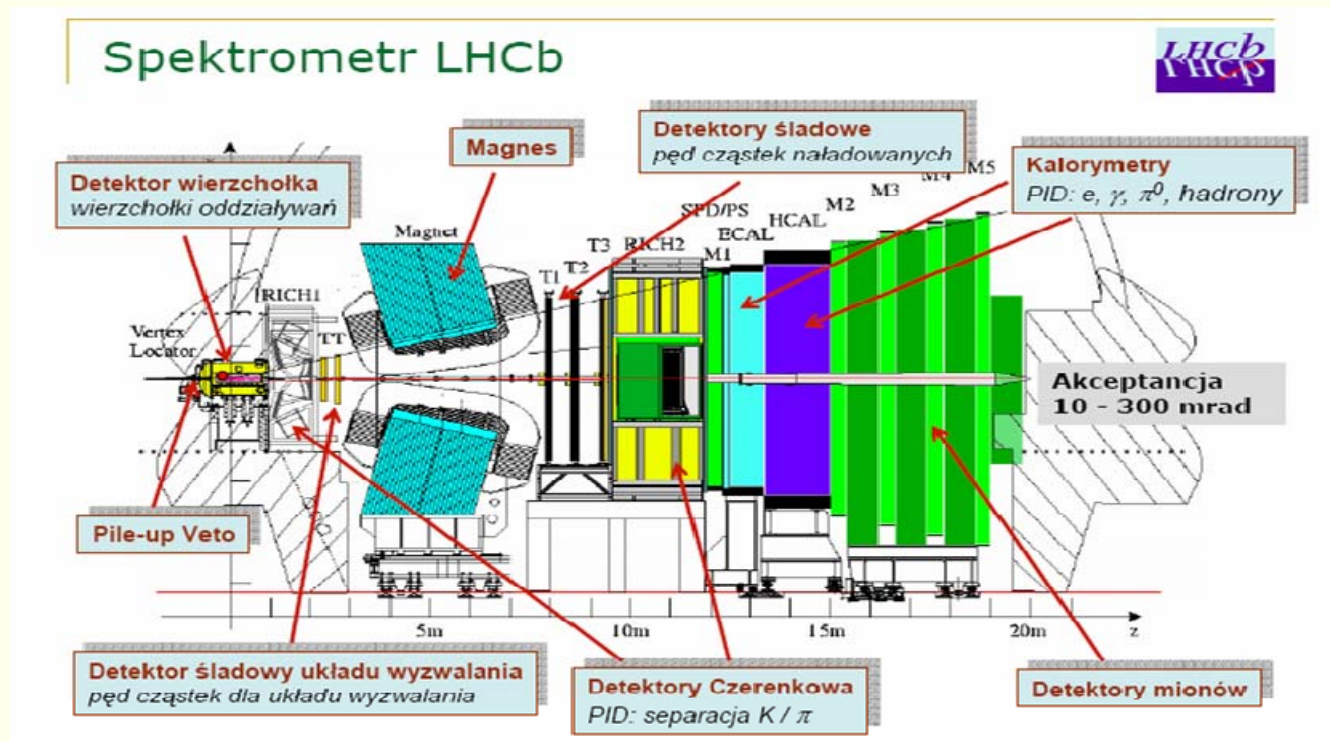
Zjawiska rzadkie $\sim 10^{-4} - 10^{-8}$



Pozostałe 3 eksperymenty na LHC mają swoje zestawy „igieł”

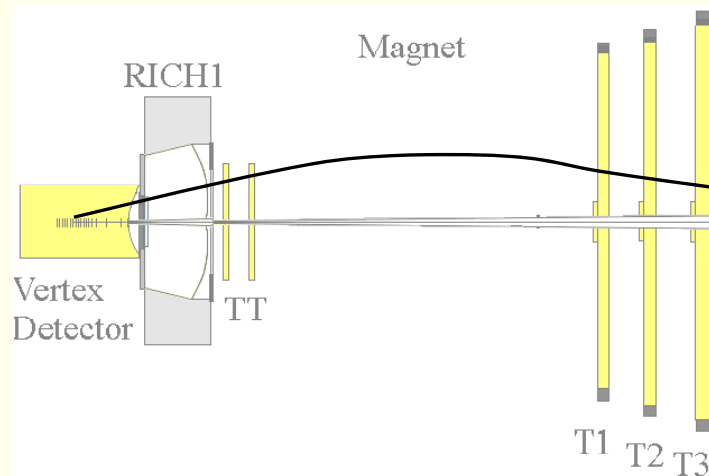
Projektowanie eksperymentu

- Eksperyment powinien charakteryzować się maksymalnym potencjałem badawczym przy zadanym budżecie (konieczna optymalizacja)
- Projektowanie jest procesem wieloletnim (~10 lat) w trakcie którego wprowadza się wiele poprawek spowodowanych np. rozwojem technologii.



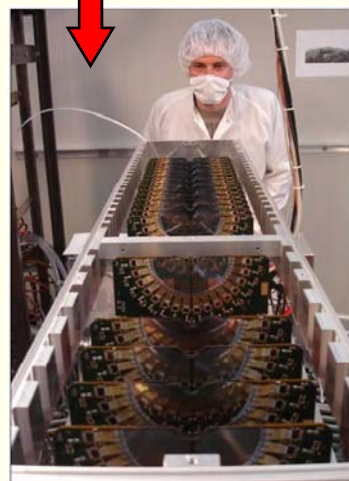
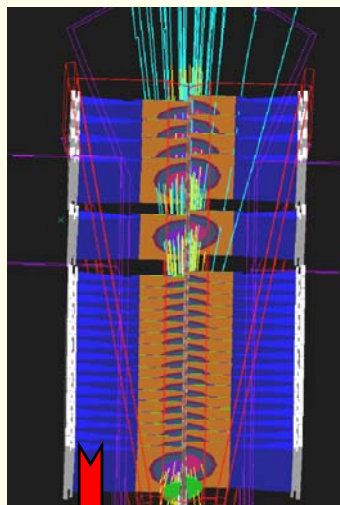
Typowy proces optymalizacji

- Projektujemy zestaw detektorów do pomiaru pędu
 - Najlepszy detektor to taki, który nic nie waży i ma 100 % wydajności
 - Symulacje komputerowe -> dobieramy pozycje i liczbę płaszczyzn pomiarowych, zakładamy jakie materiały będą użyte.
 - Sprawdzamy czy dokładność pomiaru pędu pozwoli na osiągnięcie dobrej jakości wyników analizy fizycznej.
 - Jeżeli nie to zmieniamy geometrię, szukamy nowych materiałów lub technik pomiarowych...
 - Praktyka -> stosujemy konkretne materiały
 - Budujemy prototyp itd..

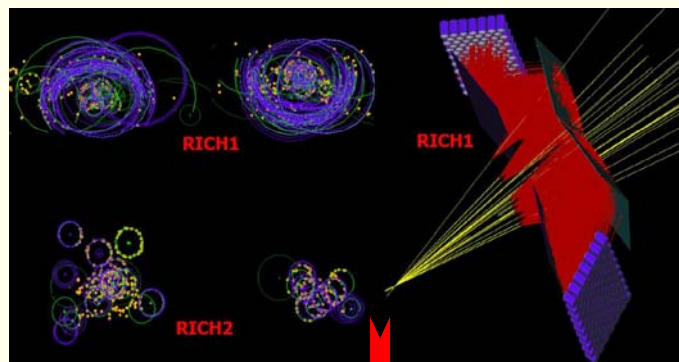


Modelowanie aparatury

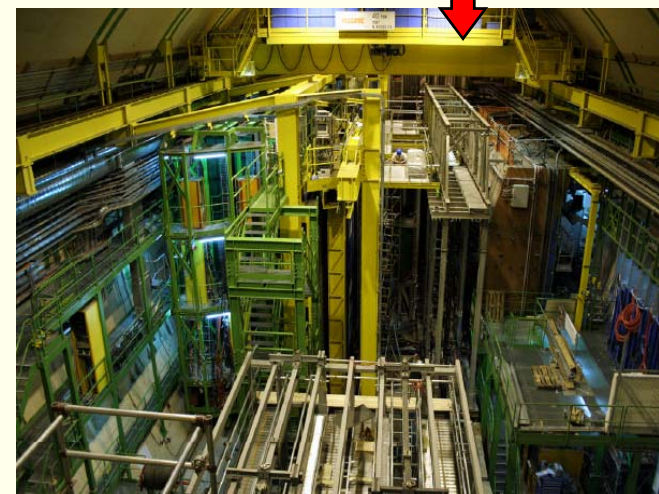
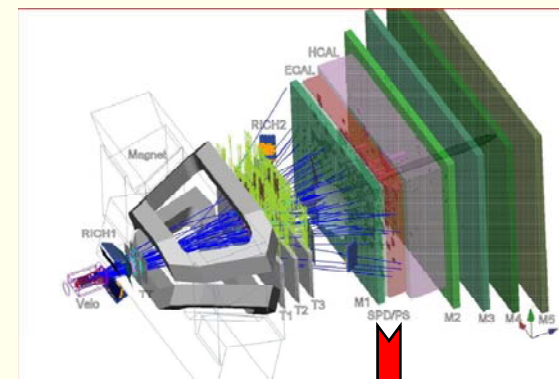
Detektor wierzchołka



Detektor RICH
Identyfikacja cząstek



Pełny spektrometr

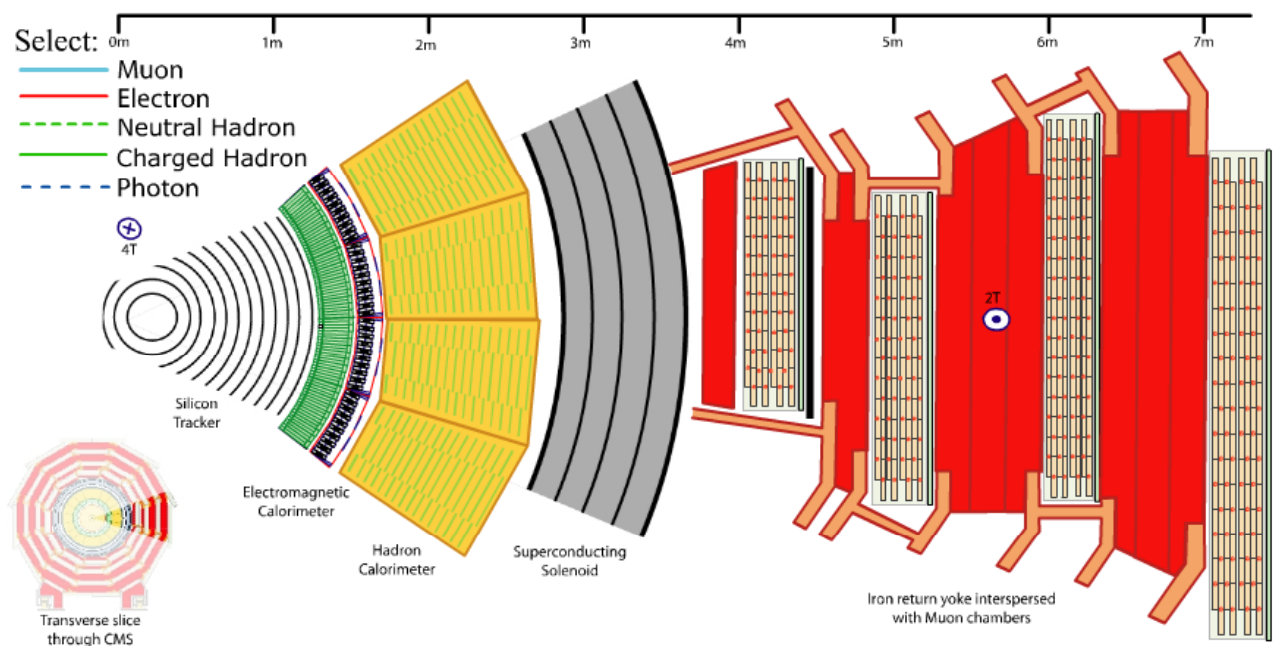


Symulacja zjawisk w detektorach

Transverse slice through CMS detector

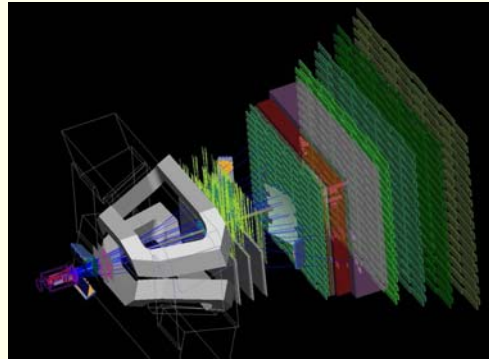
Click on a particle type to visualise that particle in CMS

Press "escape" to exit



Slajd ze strony [www eksperymentu CMS](http://www.experimentu CMS)

Symulacja i prawdziwy eksperyment



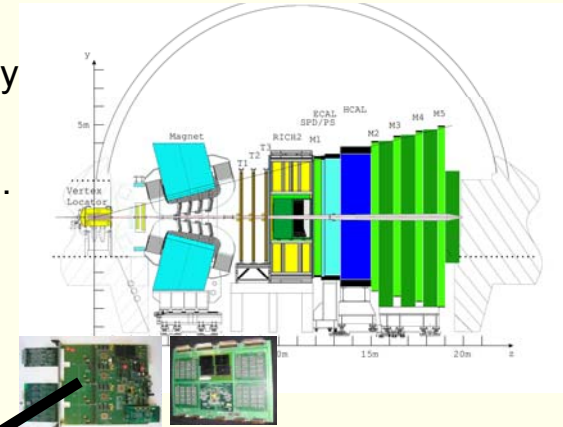
Symulacja

1. Generator przypadków (teoria)
2. Modelowanie odpowiedzi aparatury (Geant4)

Detektor „działa” wspaniale wewnątrz komputerów. Czy w takim razie potrzebujemy budować prawdziwy?

Tak. W prawdziwym eksperymencie pojawiają się zjawiska o których być może nawet nie pomyśleliśmy.

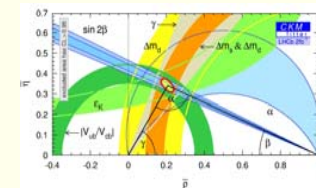
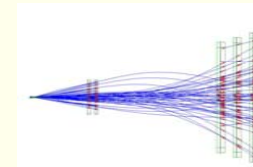
W efekcie modelowania otrzymujemy dane symulowane w formie takiej jak będą napływać z elektroniki odczytu aparatury eksperymentalnej.



Prawdziwy eksperyment

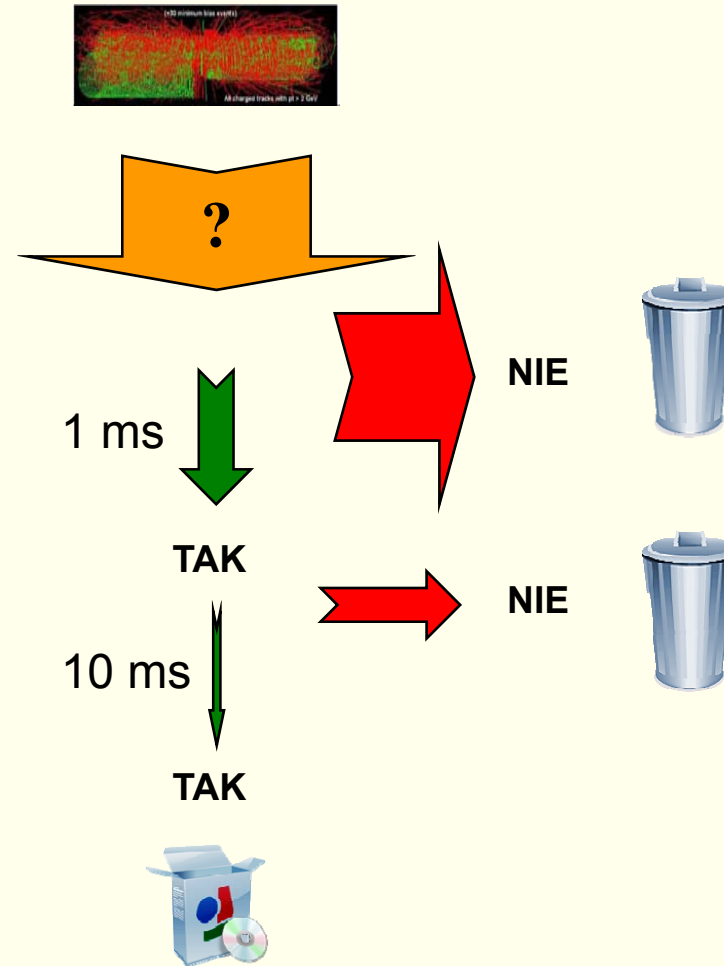
Rekonstrukcja

Analiza fizyczna

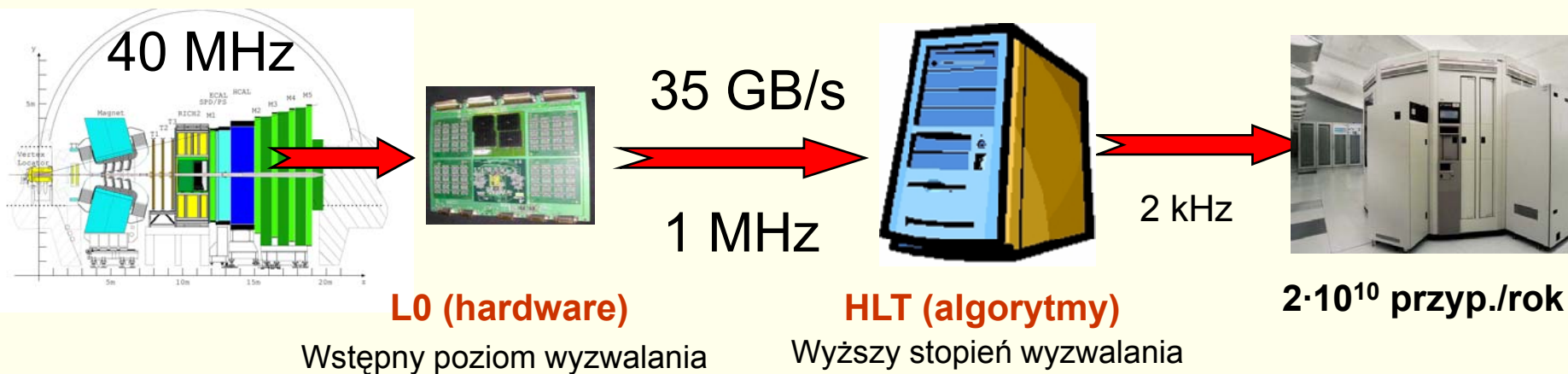


Zbieranie danych - system wyzwalania

- Zadaniem systemu wyzwalania jest przede wszystkim jak najszybsze odrzucanie zbędnych przypadków (a nie wybieranie tych interesujących).
- Wtedy stopniowo mamy do dyspozycji więcej czasu na dokładniejszą analizę przypadków zaakceptowanych i dokładniejsze odrzucanie.



Zbieranie danych

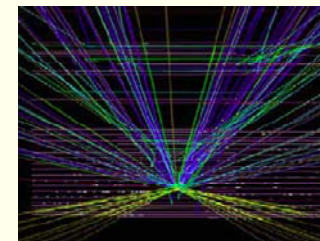


1 MHz → 1 μs/przypadek.
W ciągu 1 μs nie można
nawet zdekodować danych
Co zrobić ???



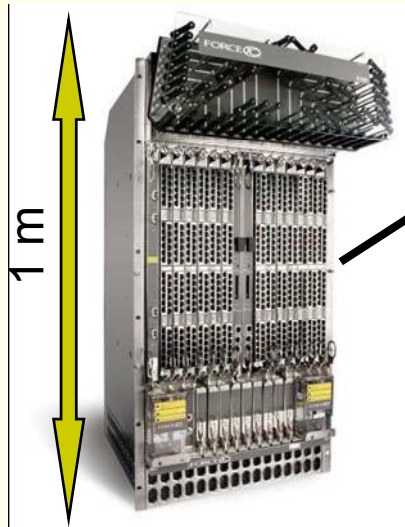
1000 procesorów → 1 ms/przypadek

Możliwa staje się
rekonstrukcja śladów w
detektorze wierzchołka w
rzucie $r\phi$

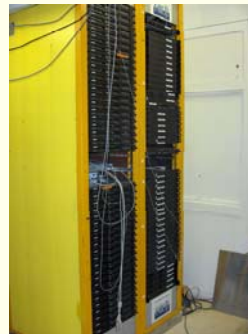
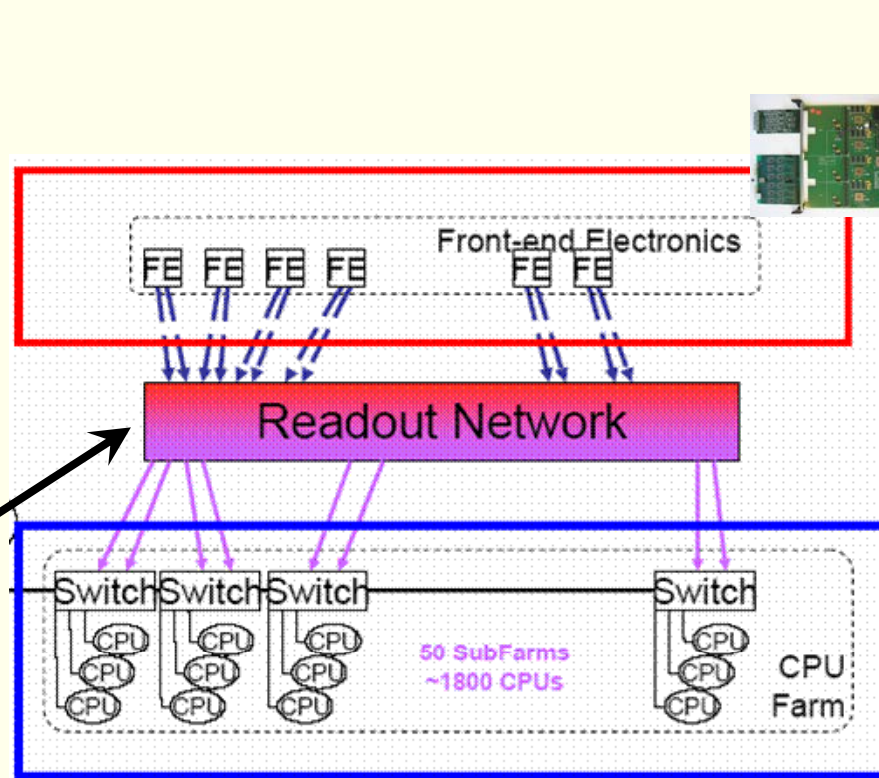


Farma procesorów on-line

1 MHz na wejściu
→ 30 GB/s



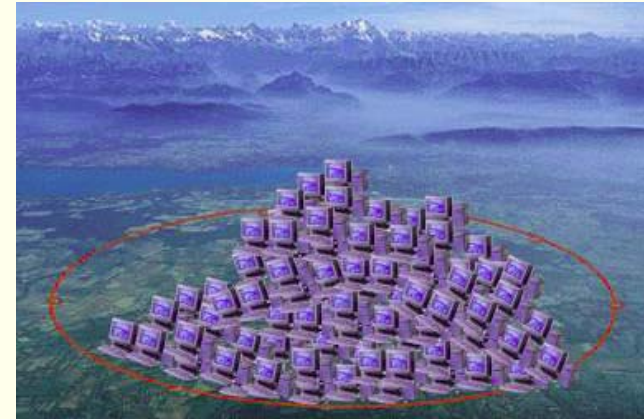
Force10 E1200, 1260 GbE ports
Przepustowość 50 GB/s



1000 CPU
2 kHz na wyjściu
Zapis 250 MB/s

Opracowywanie danych

- Analiza danych zapisanych na taśmy
 - Przypadki przefiltrowane
- niespotykane dotąd wyzwania
 - ~kilka PB danych na rok
 - 200 tys procesorów
 - 10 tys dysków
- Skoncentrowanie infrastruktury w jednym miejscu jest niesłychanie trudne (i niecelowe bo blokuje rozwój innych ośrodków).
 - Instalacja, zasilanie, chłodzenie, obsługa systemów operacyjnych i oprogramowania
- Konieczny model rozproszonych obliczeń czyli zespół współpracujących ze sobą farm komputerowych.
 - **Czy można jednak stworzyć coś bardziej uniwersalnego co byłoby przydatne także dla innych dziedzin?**



GRID – globalny komputer

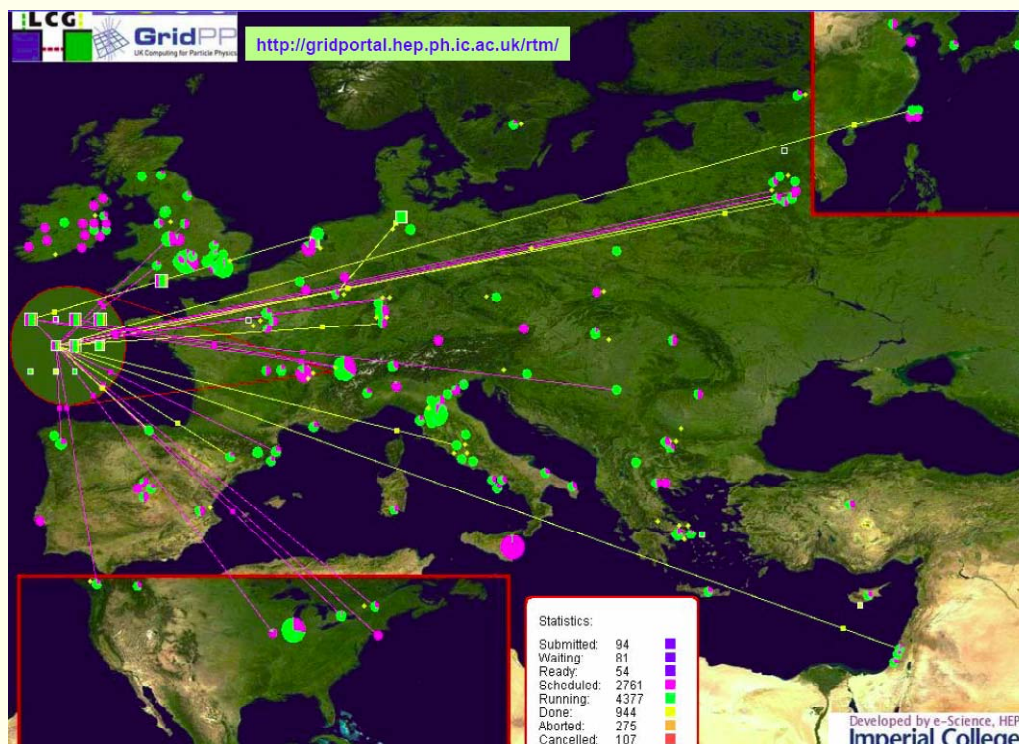
- WWW – przewrót w dostępie do informacji (wynaleziona w CERN).
 - Jednolity dostęp do informacji bez względu na miejsce na Ziemi.
- GRID – jednolity dostęp do zasobów obliczeniowych.
 - GRID widziany przez każdego użytkownika jako **jeden wielki komputer**.
 - Dobra platforma współpracy dla dużych projektów.
 - Wspólne narzędzia i dane, dostęp do mocy obliczeniowej.
 - Analogia z **siecią elektryczną** (także w przypadku opłat za pobraną moc obliczeniową)



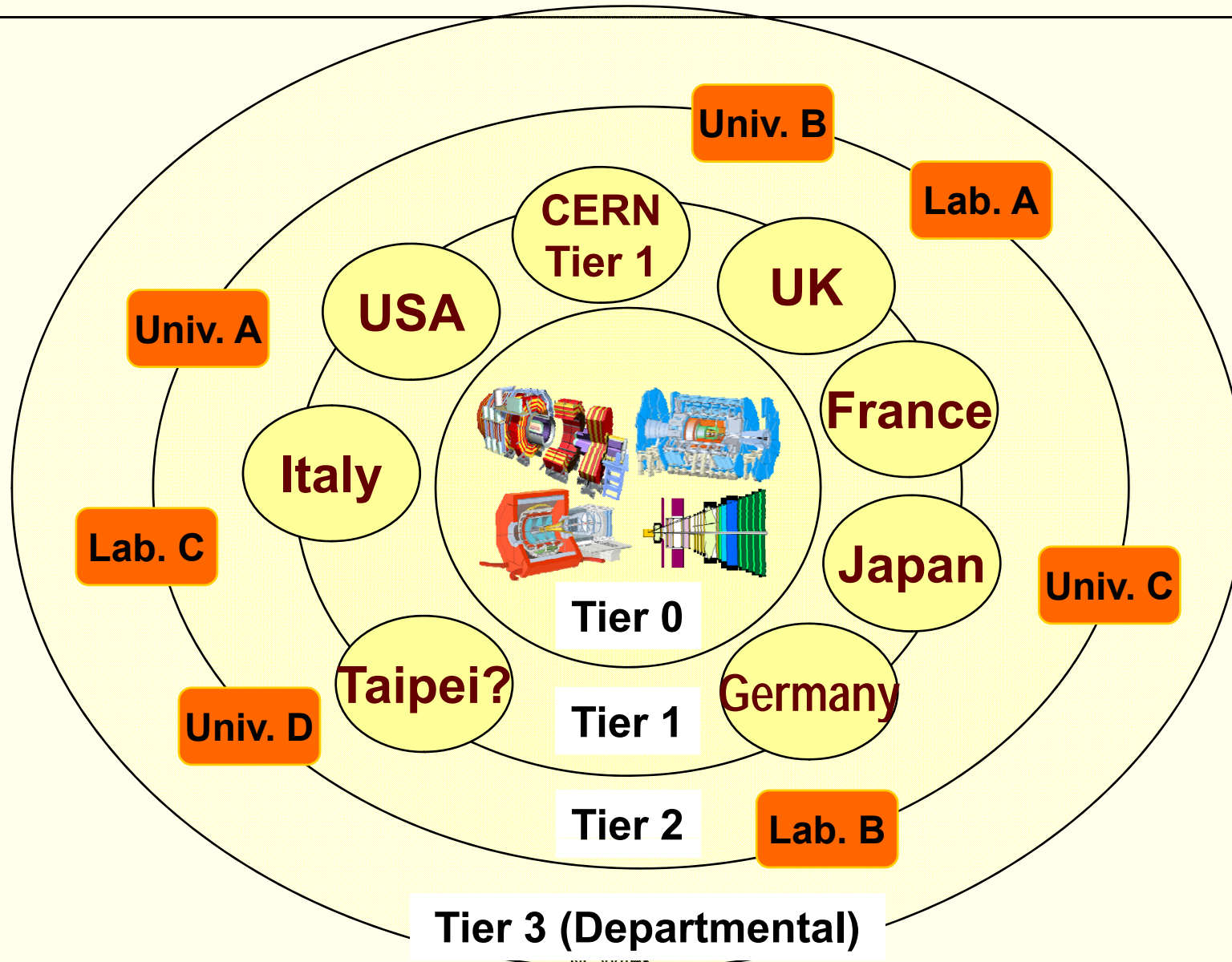
- Pierwszą udaną inicjatywą na dużą skalę były obliczenia w ramach projektu SETI (*Search for Extra-Terrestrial Intelligence*)
 - Każdy posiadacz PC mógł udostępnić moc obliczeniową np. w nocy i uruchomić program analizujący sygnały przychodzące z kosmosu.
 - Projekt zaoszczędził wiele milionów \$

Ogólnosiwiatowa siec GRID

- EGEE jest największym ogólnosiwiatowym projektem GRID
 - Obecnie ponad 40 tys procesorów rozproszonych po całym świecie
 - 100 tys zadań wykonywanych jednocześnie
 - 5 PB przestrzeni dyskowej
- Koordynatorem projektu jest CERN
- Fizyka cząstek jest głównym użytkownikiem (największe potrzeby)

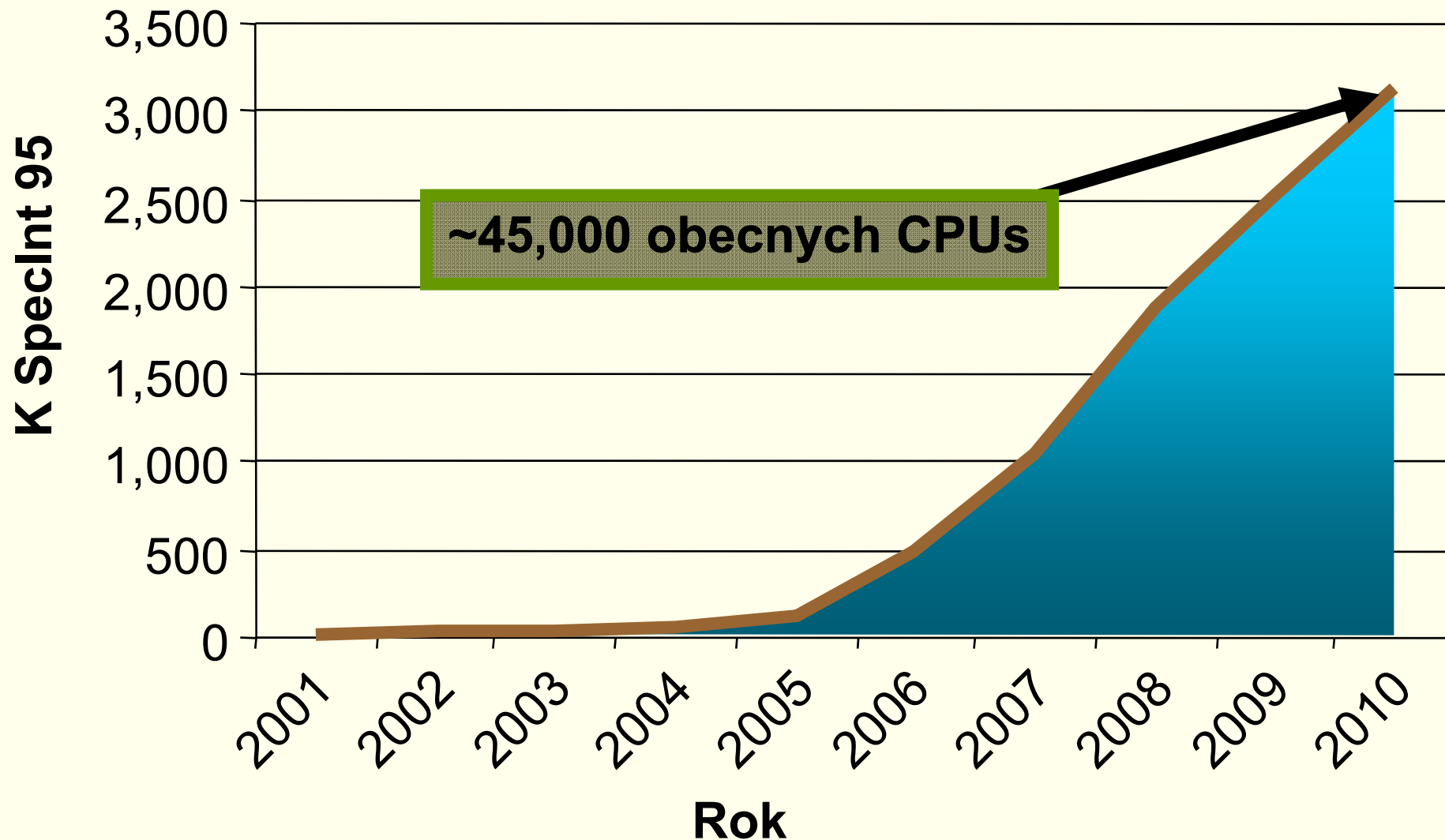


Struktura hierarchiczna (Multi Tier)



CERN Tier 0+1

Moc obliczeniowa



Technologie farm PC

- ❑ Początkowo CERN zdecydował zakupić komputery w formie zwykłych PC dostępnych w sklepach.
- ❑ W międzyczasie technologie RACK i BLADE osiągnęły konkurencyjne ceny.
- ❑ Obecnie w dużych ośrodkach wygrywają technologie specjalistyczne RACK i BLADE



Pierwsze klastry w CERN w standardowej technologii.



Hala w CERN przygotowana do montażu PC w szafach RACK

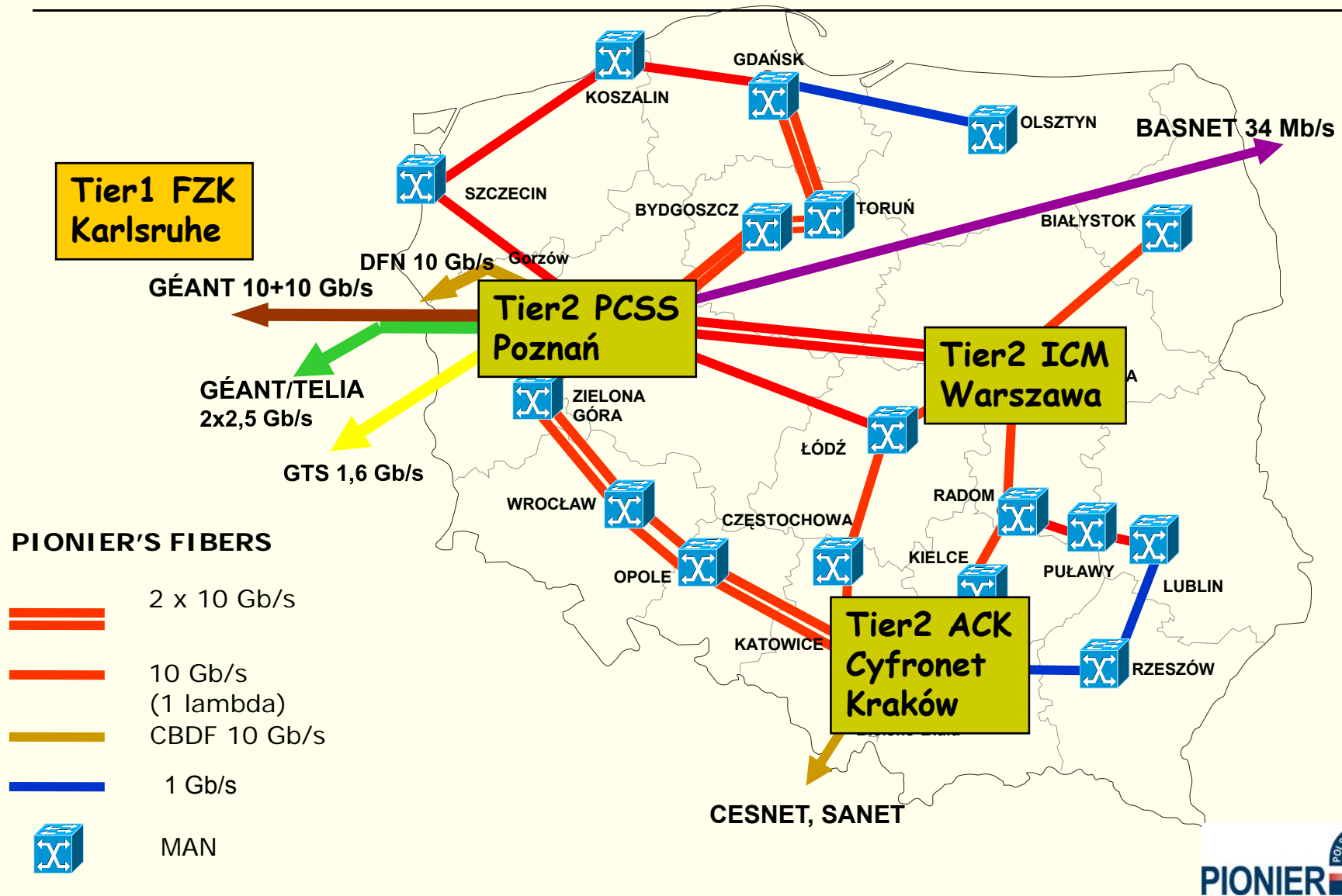


Hala w ośrodku obliczeniowym Tier-1 w Karlsruhe, Niemcy.



PC typu RACK i BLADE

Polska infrastruktura WLCG



Podsumowanie

- Eksperymenty na LHC dostarczają ogromnej ilości danych.
- Przypadki których szukamy zdarzają się bardzo rzadko, ~1 na 100 mln.
- Jedynie dzięki komputerom udaje się zaprojektować eksperyment, zebrać i analizować dane.
- CERN stymuluje rozwój nowych technologii informatycznych (WWW, GRID)