



**HAL**  
open science

## Data challenges for ATLAS

F. Ohlsson-Malek

► **To cite this version:**

F. Ohlsson-Malek. Data challenges for ATLAS. Worskhop du Club des Utilisateurs Informatique du CEA (CUIC), Jun 2003, Saint Malo, France. 1-43 transparents. in2p3-00020354

**HAL Id: in2p3-00020354**

**<https://hal.in2p3.fr/in2p3-00020354>**

Submitted on 29 Jan 2004

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

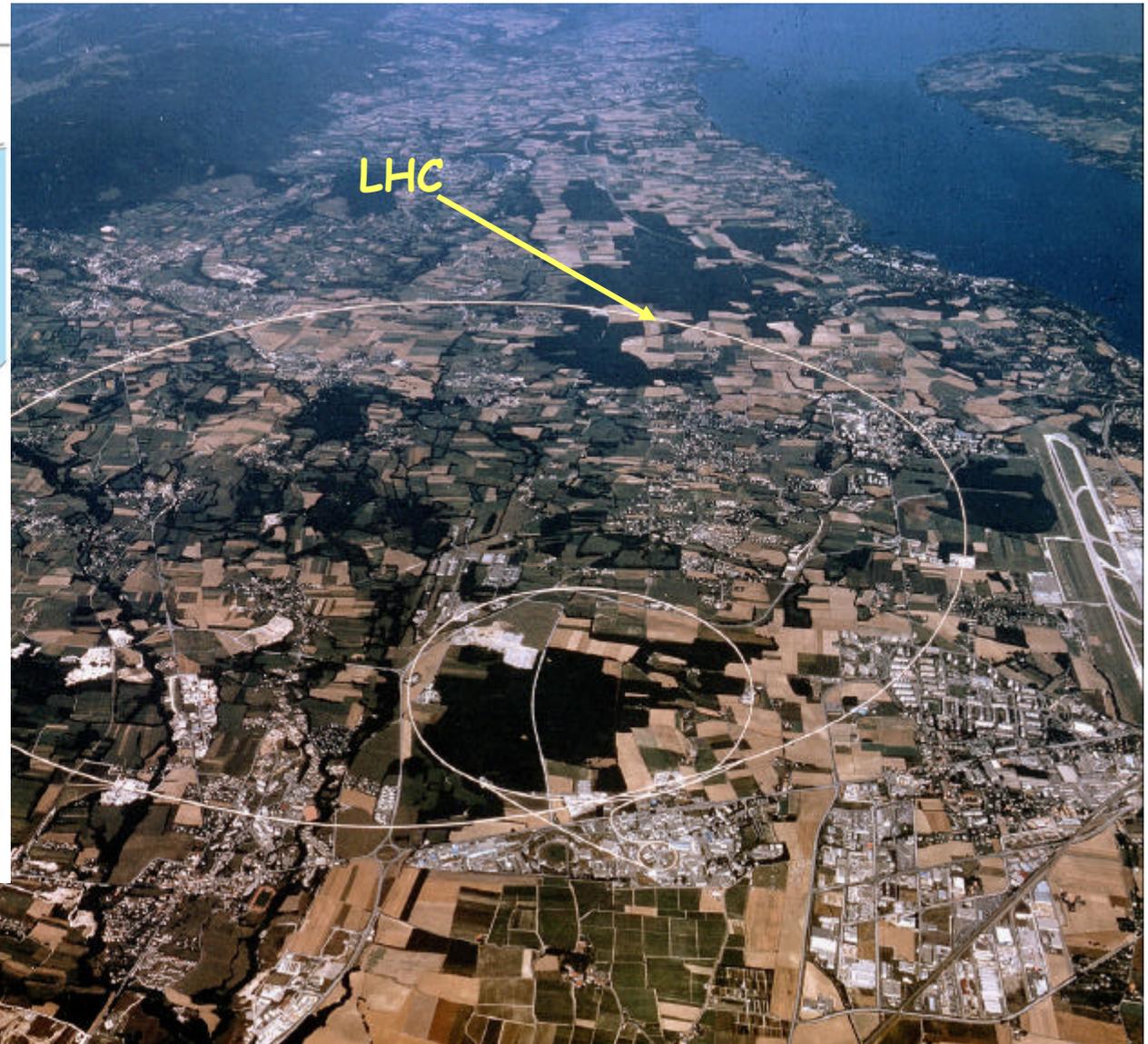
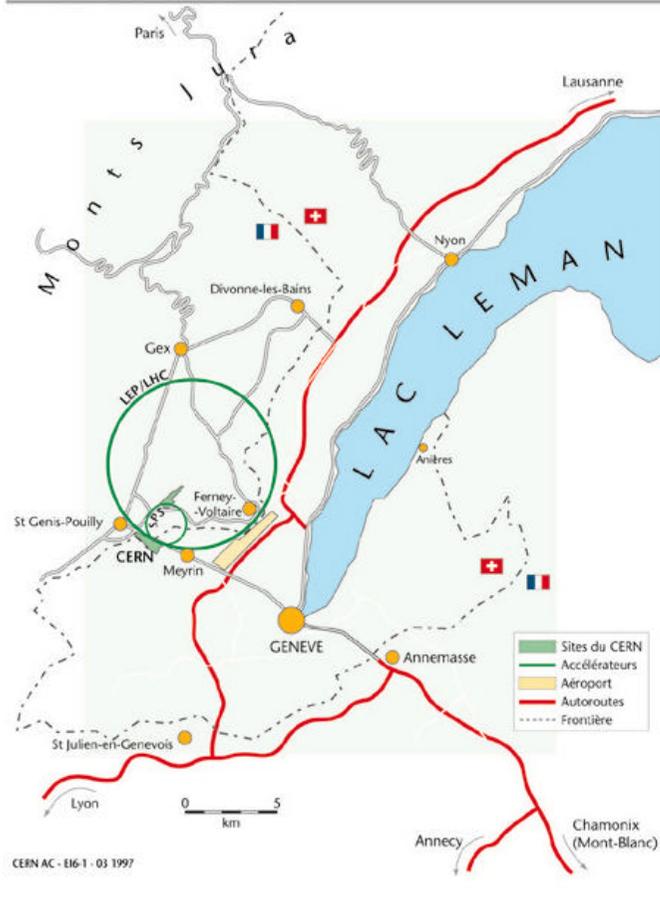
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Data Challenges pour le LHC

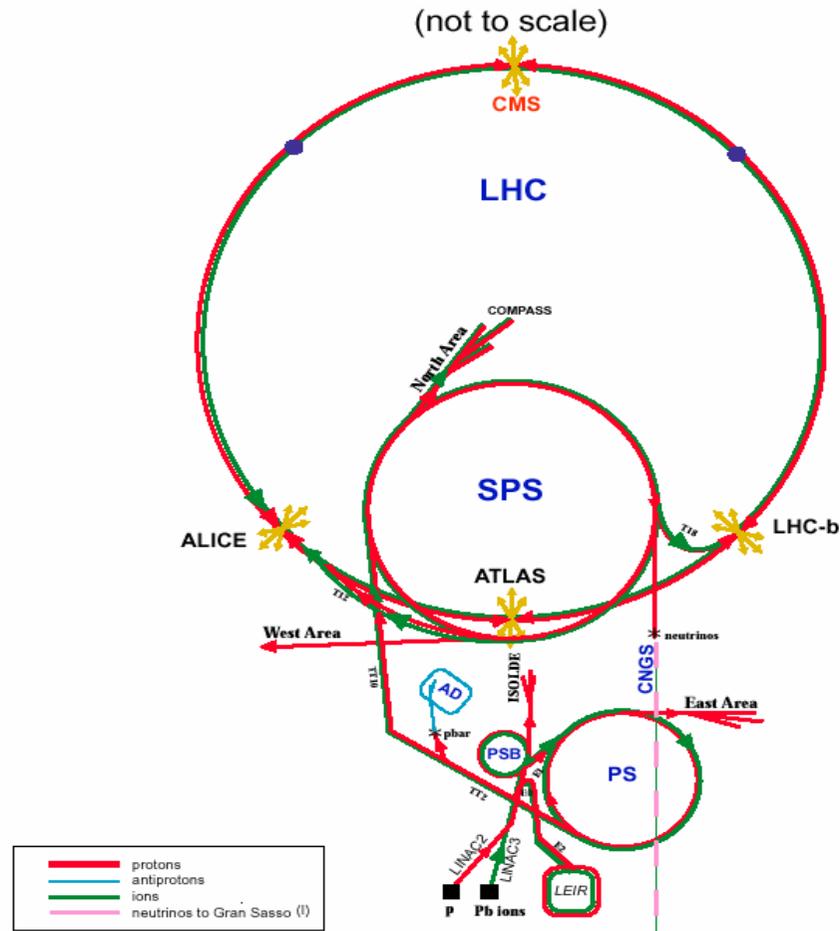
- Le LHC et le CERN
- Les expériences sur le LHC
- Le but scientifique au LHC
- Les modes calculatoires et les données
- Les « Data-Challenges » avec ALICE, CMS et ATLAS
- La suite ...Conclusion

# Le CERN: vue aérienne

Carte de situation



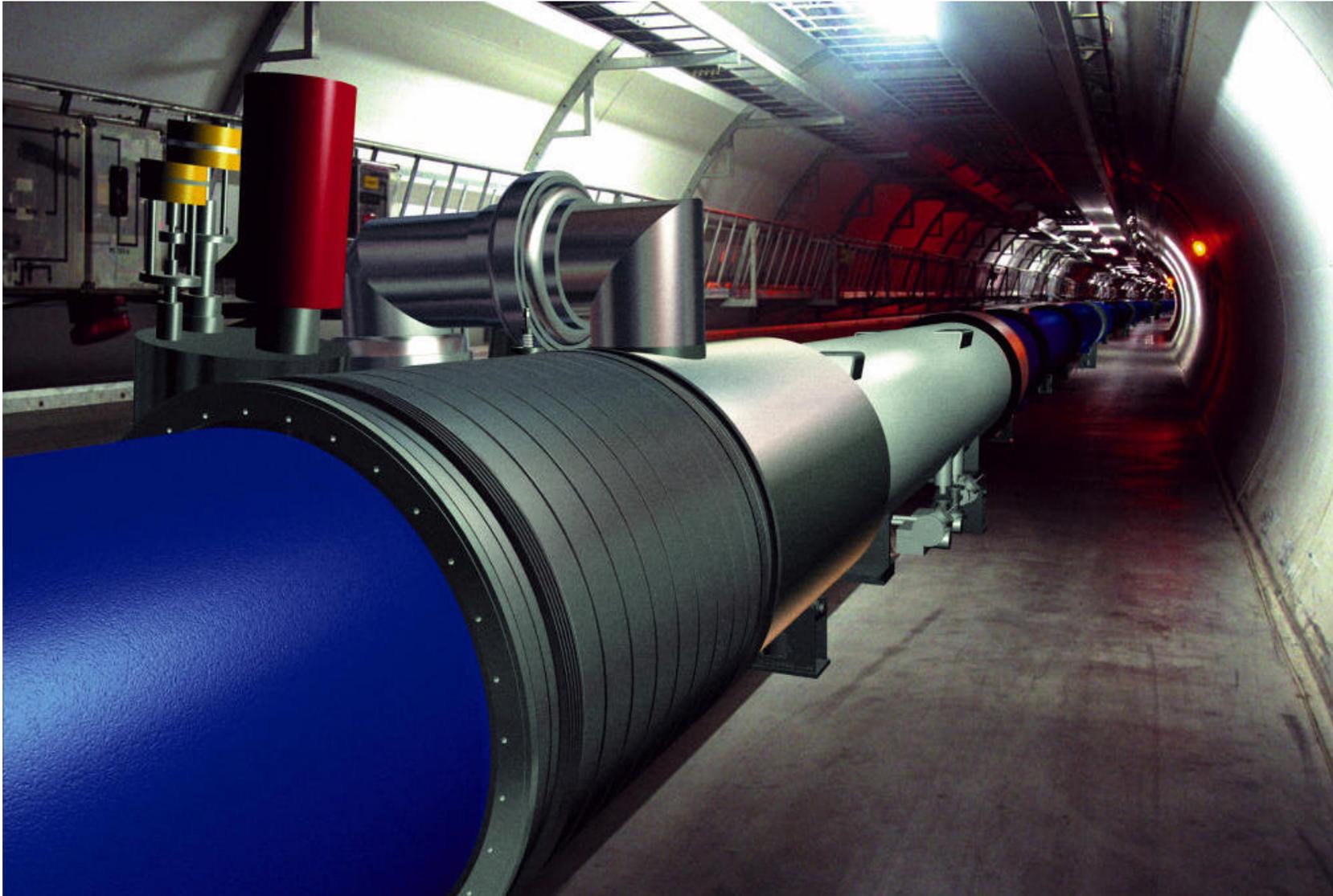
# Les accélérateurs du CERN et le LHC



LHC: Large Hadron Collider  
 SPS: Super Proton Synchrotron  
 AD: Antiproton Decelerator  
 ISOLDE: Isotope Separator OnLine DEvice  
 PSB: Proton Synchrotron Booster  
 PS: Proton Synchrotron  
 LINAC: LINear ACcelerator  
 LEIR: Low Energy Ion Ring  
 CNGS: Cern Neutrinos to Gran Sasso

Rudolf LEY, PS Division, CERN, 02.09.96  
 Revised and adapted by Antonella Del Rosso, ETT Div.,  
 in collaboration with B. Desforges, SL Div., and  
 D. Manglunki, PS Div. CERN, 23.05.01

# Le tunnel du LHC



CUIC 03, 17-18 juin 2003

Data Challenges au LHC

F. Ohlsson-Malek  
LPSC/IN2P3

# L'échelle des choses

## Instruments



Accelerators  
LHC, LEP



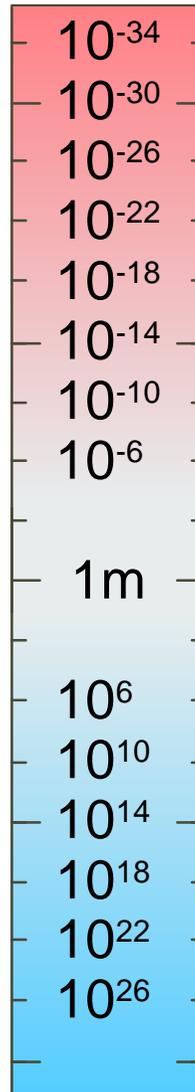
(Particle beams)  
Electron  
Microscope  
Microscope



Telescope



Radio  
Telescope



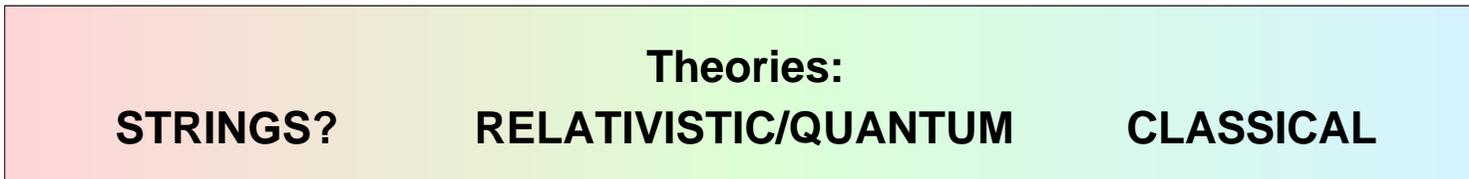
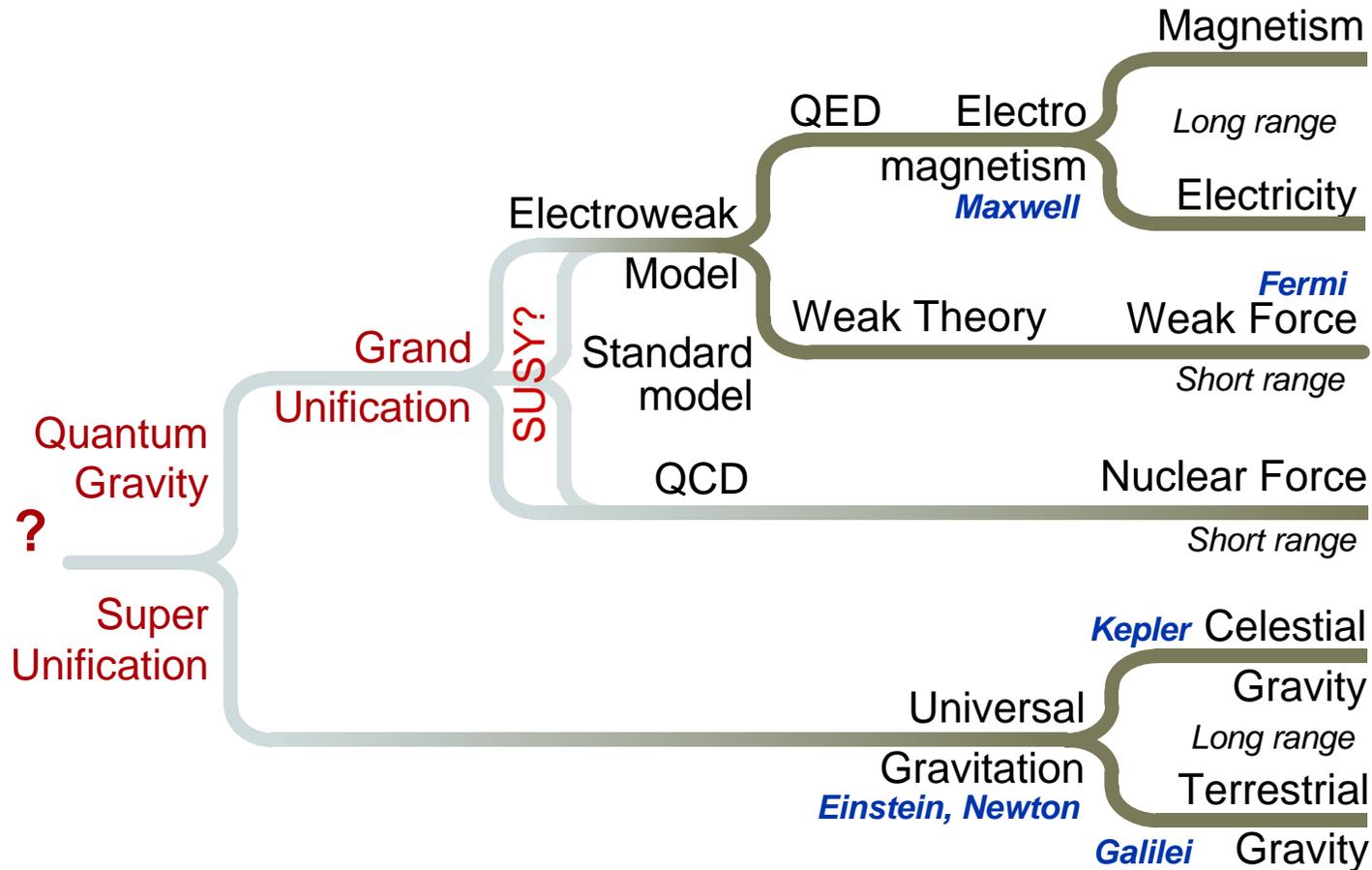
SUSY particle?  
Higgs?  
Z/W (range of nuclear force)  
Proton  
Nuclei (range of weak force)  
Atom  
Virus  
Cell

## Observables



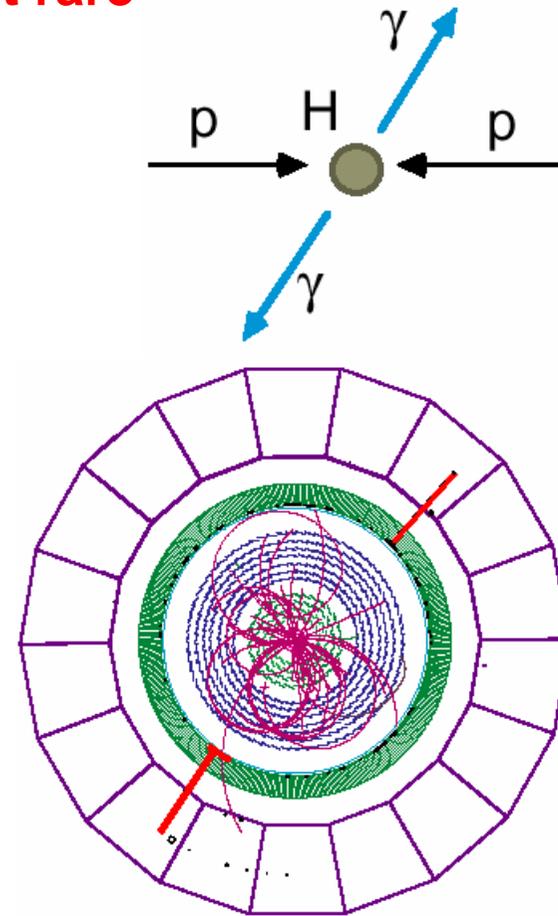
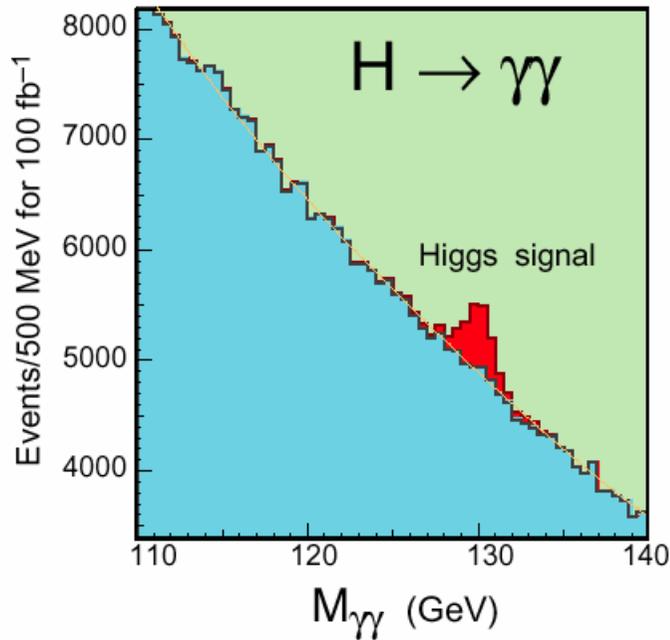
Earth radius  
Earth to Sun  
Galaxies  
Radius of observable  
Universe

# Les forces

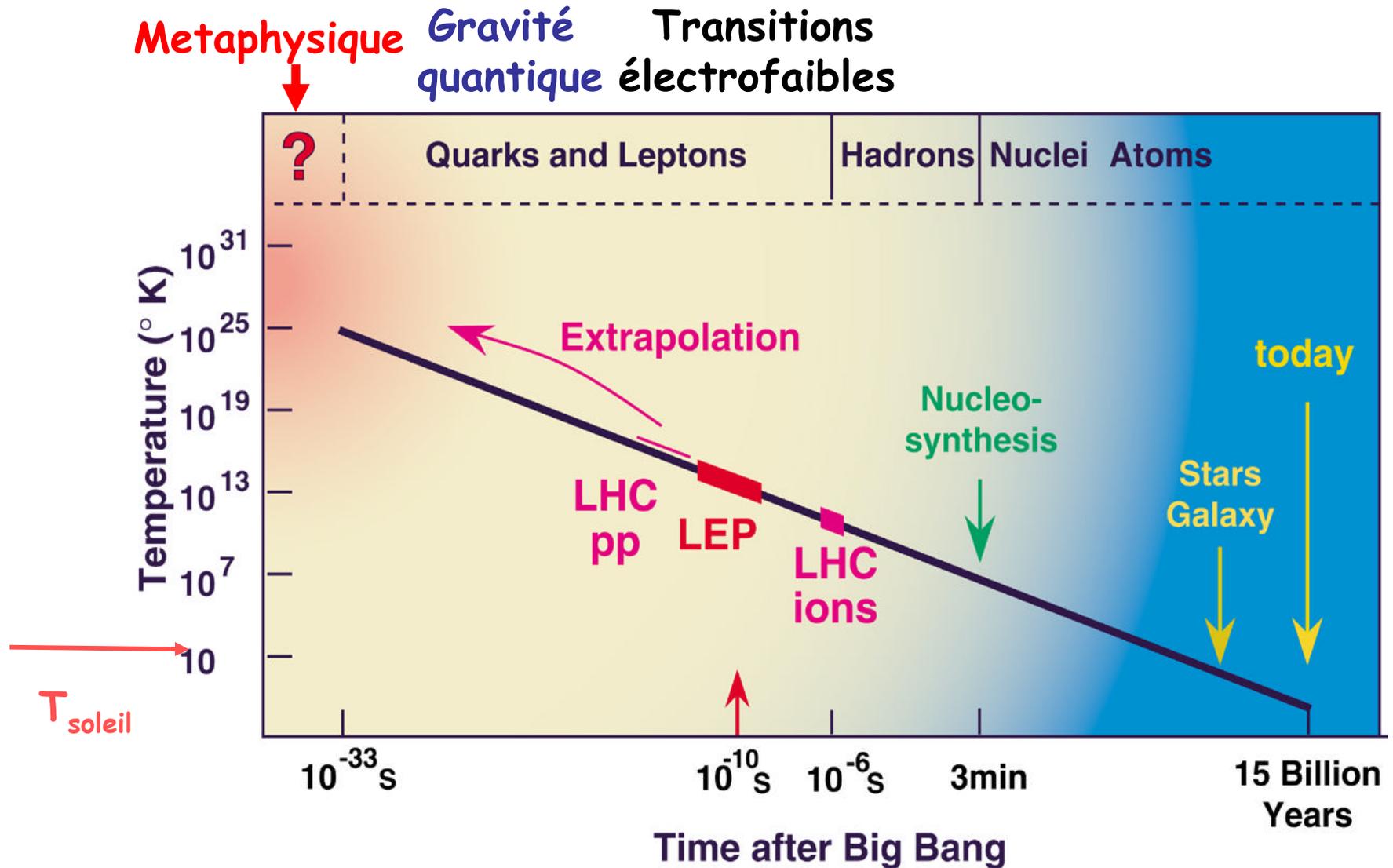


# Le but: avoir un signal clair d'une découverte attendue

**Higgs via  $H^{\text{R}}_{gg}$ : la désintégration est rare**

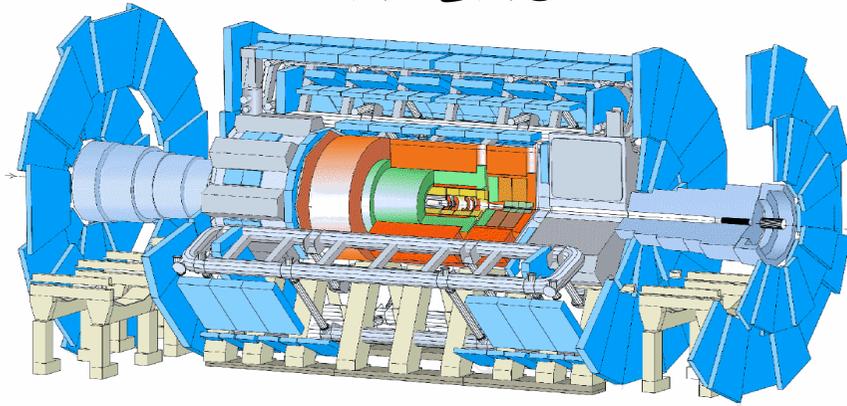


# La physique des particules

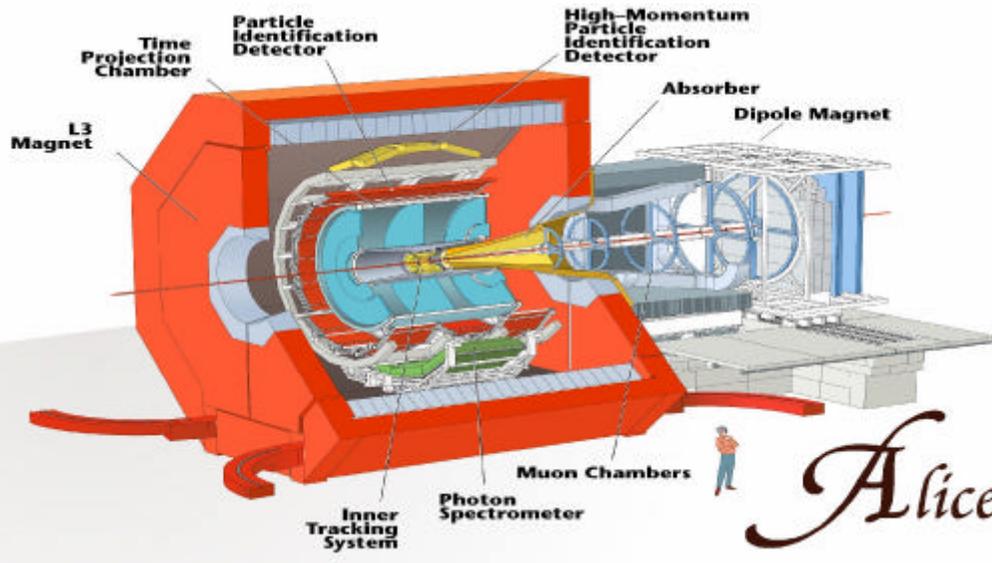
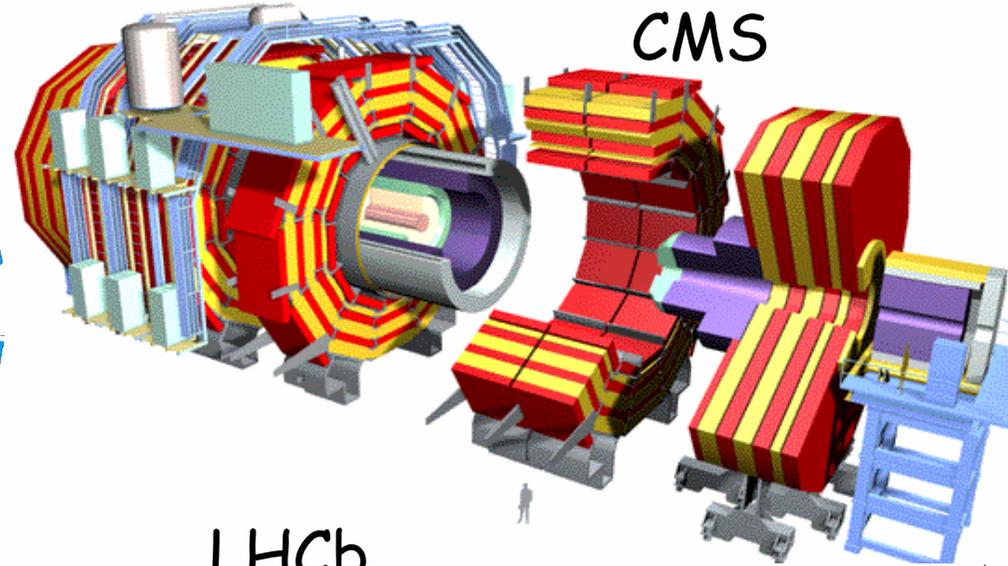


# Les expériences au LHC

ATLAS

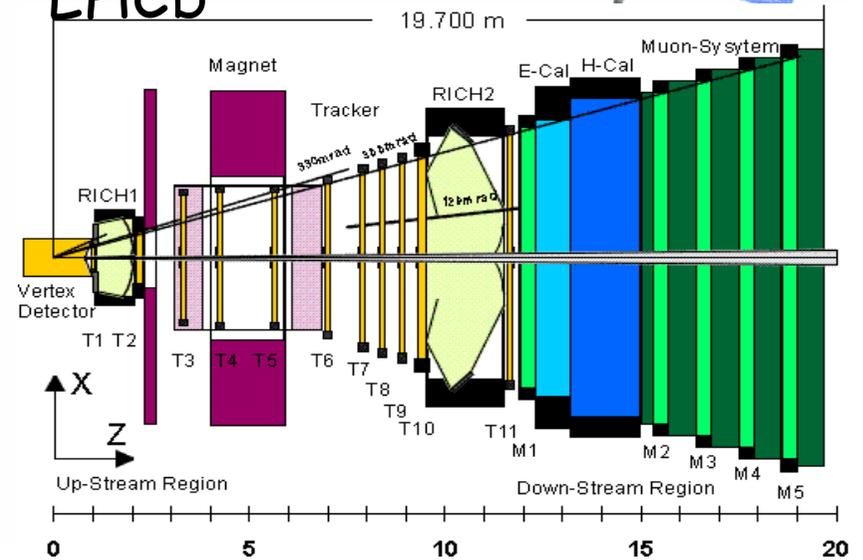


CMS

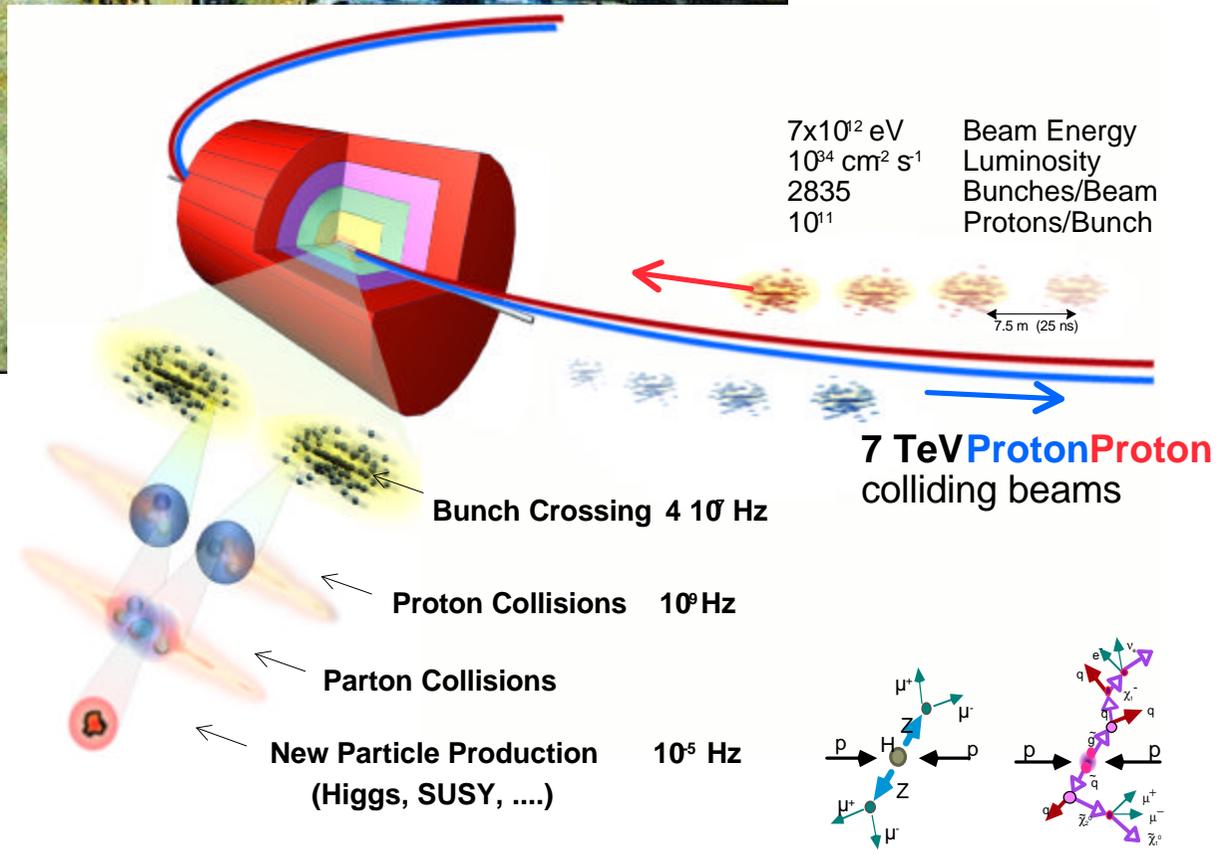
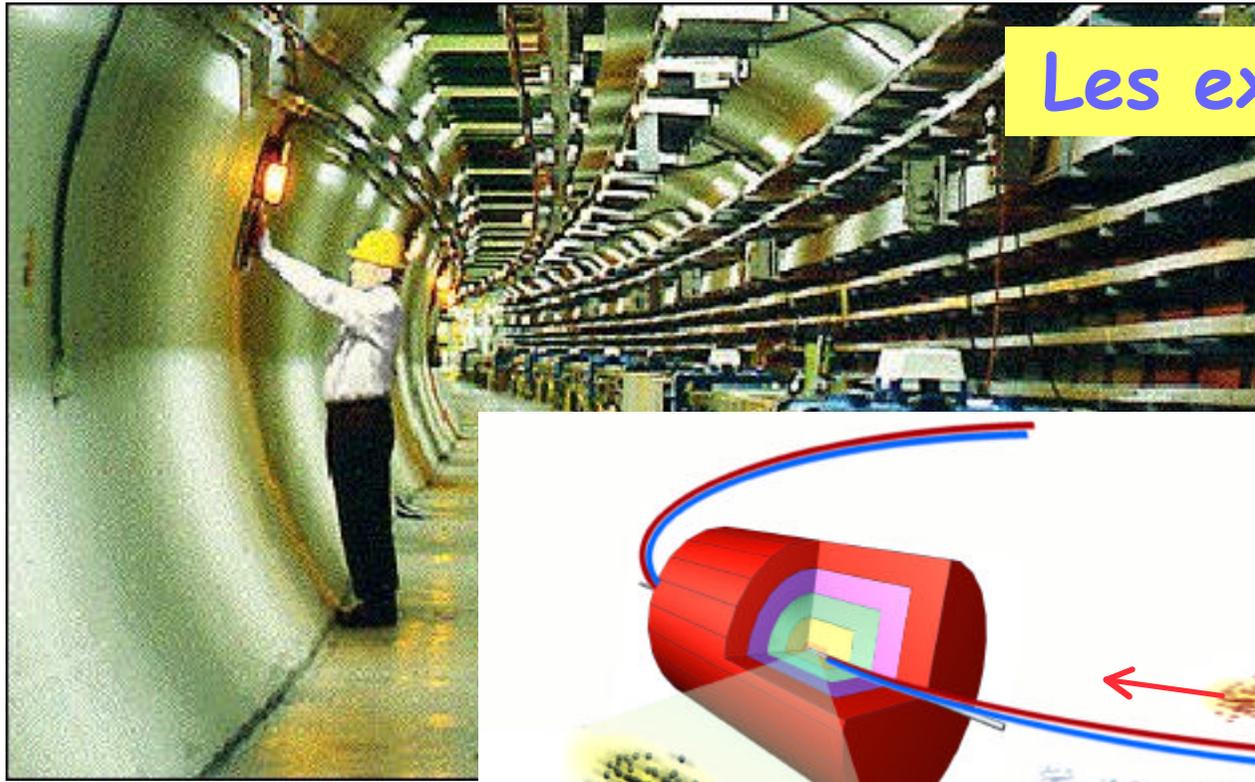


*Alice*

LHCb

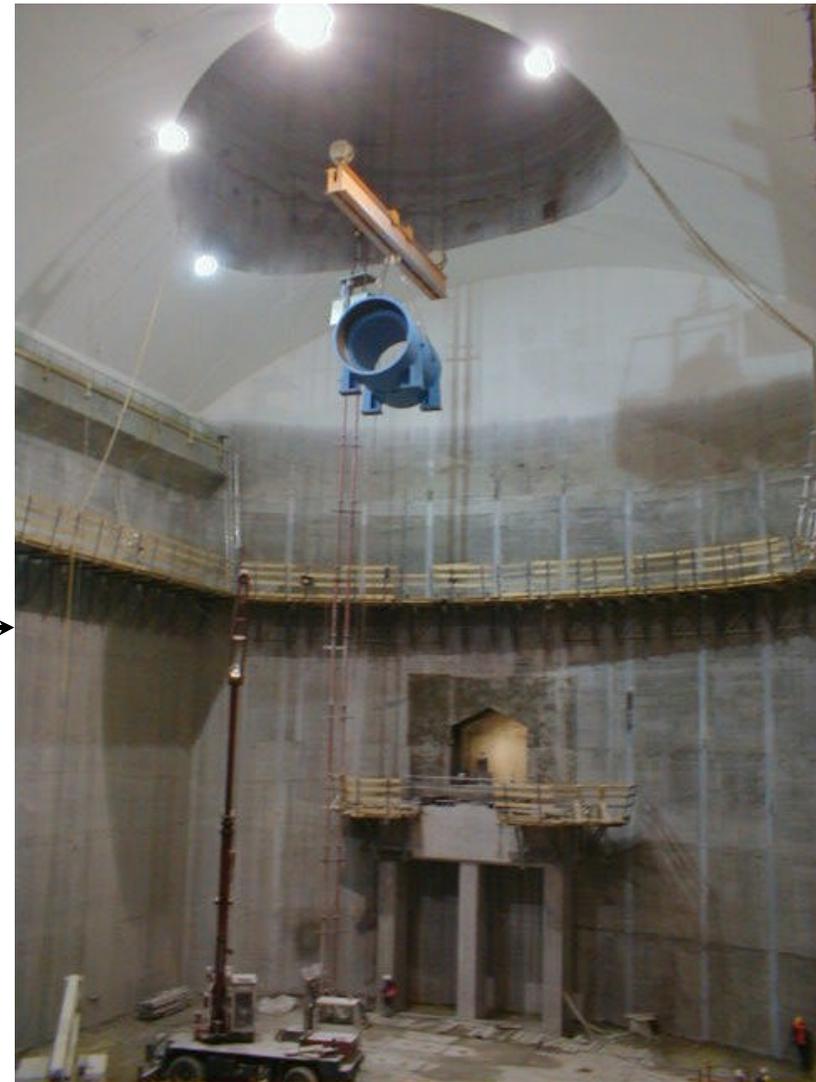
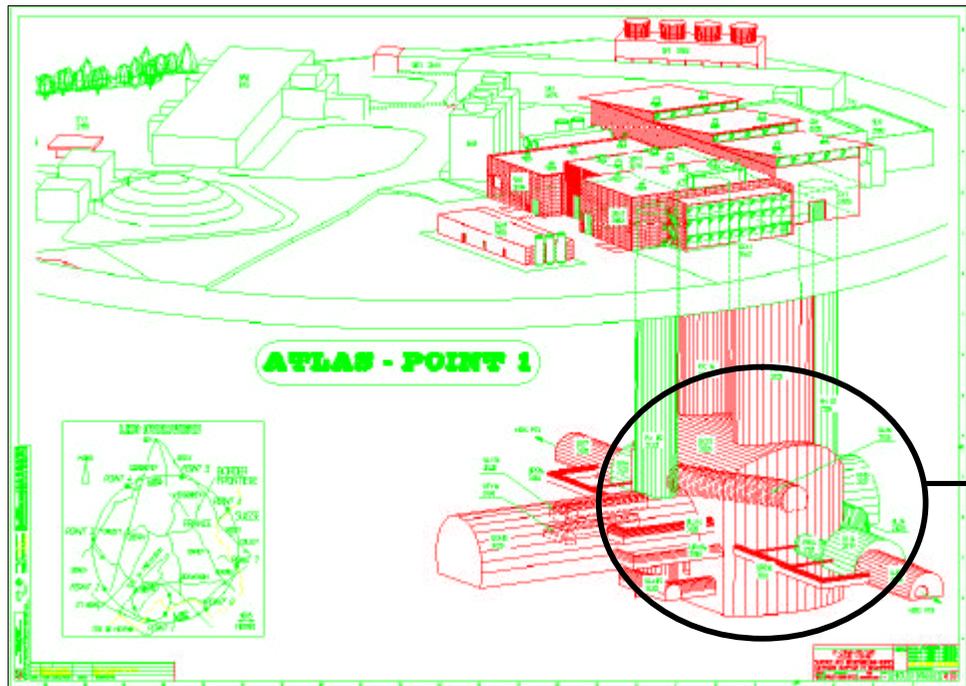


# Les expériences au LHC



**Selection of 1 event in 10,000,000,000,000**

# ATLAS: bientôt



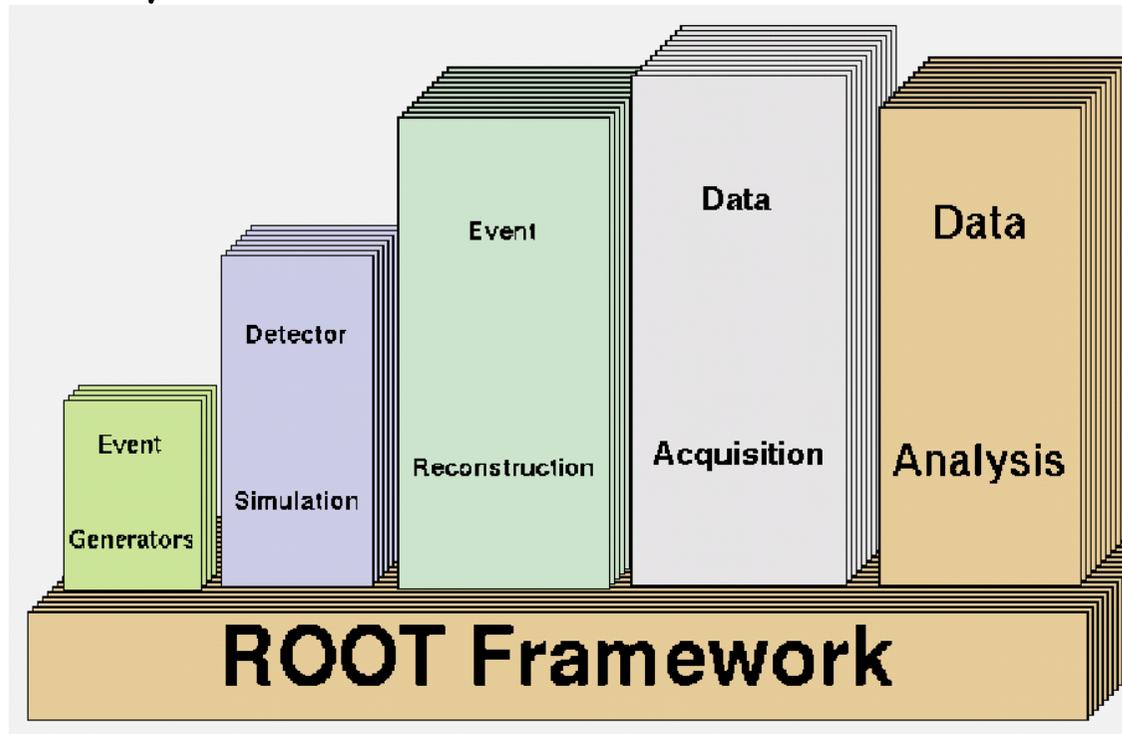
# Utilisateurs du CERN

**Europe: 267 instituts, 4603 utilisateurs**  
**Ailleurs: 208 instituts, 1632 utilisateurs**



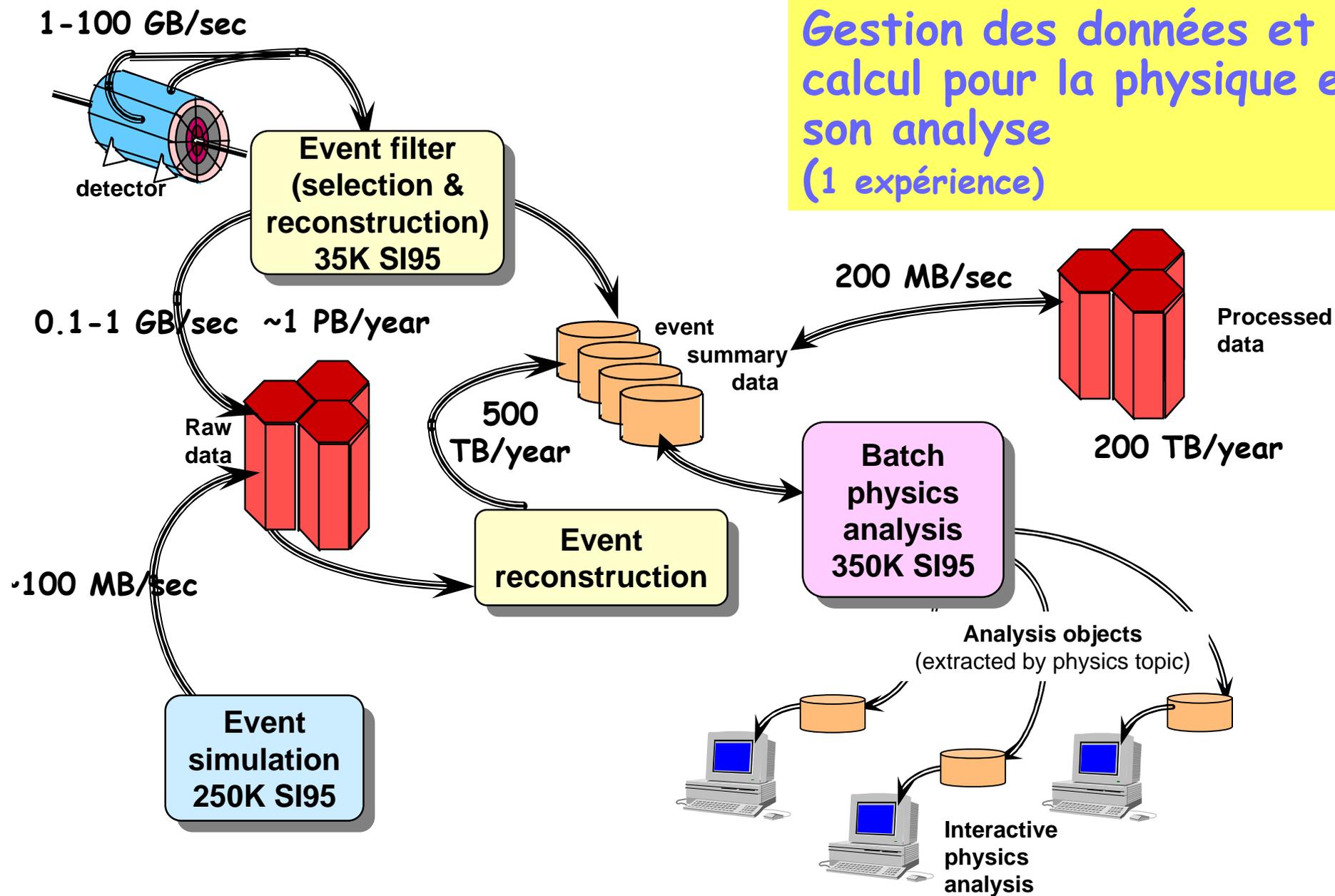
## Les modes calculatoires habituels

- Génération: modèles théoriques, phénoménologie
- Simulation
- Reconstruction
- DAQ
- Analyse



Un exemple de l'architecture du logiciel de l'expérience ALICE

**Gestion des données et calcul pour la physique et son analyse (1 expérience)**



# Données d'une expérience: exemple ATLAS



- Chaque événement consiste en une taille de  $\sim 1-1.5$  MB
- Après la sélection On-Line (HLT), les événements seront écrits sur un support permanent au taux de 100-200 Hz
- Capacité totale "raw" data:  $\approx 1$  PB/an
- Pour construire et analyser: "Modèle Logiciel Complexe et Mondialement distribué" + "Event Data Model"
  - Raw Data @ CERN
  - Données reconstruites "distribuées"
  - **TOUS** les membres de la collaboration doivent avoir accès à **TOUTES** les copies PUBLIC des données (à une vitesse d'accès raisonnable)

## Défi

# Data Challenges ou le Défi

- "Data Challenges" sont un moyen de tester un prototype d'infrastructure avant le début de l'expérience (2007)
- **ATLAS prévoit 1 "Data Challenge" par an, en augmentant la complexité et le taux de données à traiter**
- Chaque "Data Challenge" consiste en:
  - Génération d'événements de physique
  - Simulation (détecteur) de ces événements
  - Simulation du Bruit de Fond
  - Simulation de la réponse du détecteur et du pile-up (empilement)
  - Reconstruction des événements
  - Analyse des données reconstruites
- Les données peuvent être (re-)distribuées à des sites de production différents à tout moment et à chaque étape

} **le vrai défi !**

# Data Challenges

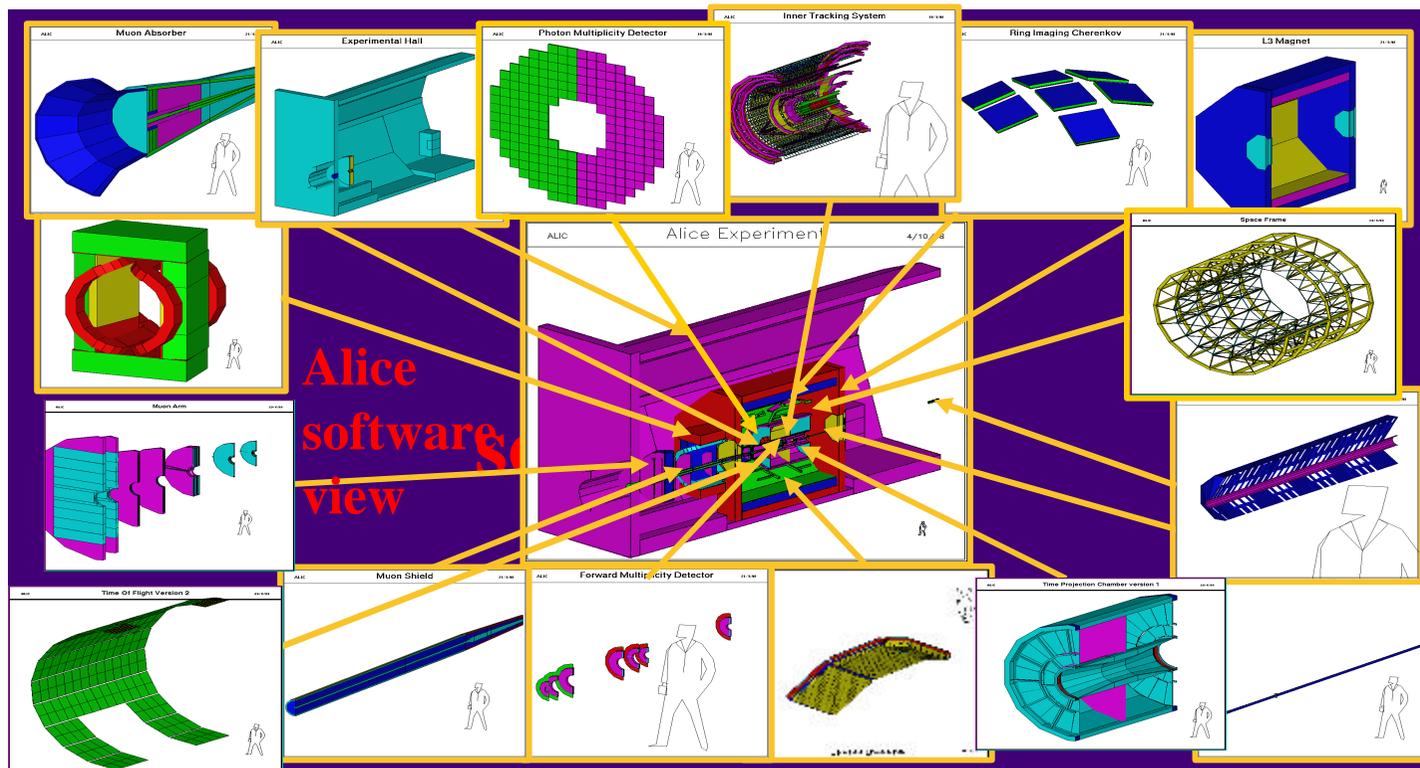
## S'assurer

- de la « bonne » description du détecteur
- de la simulation

# Description du détecteur

La clé pour comprendre la physique:

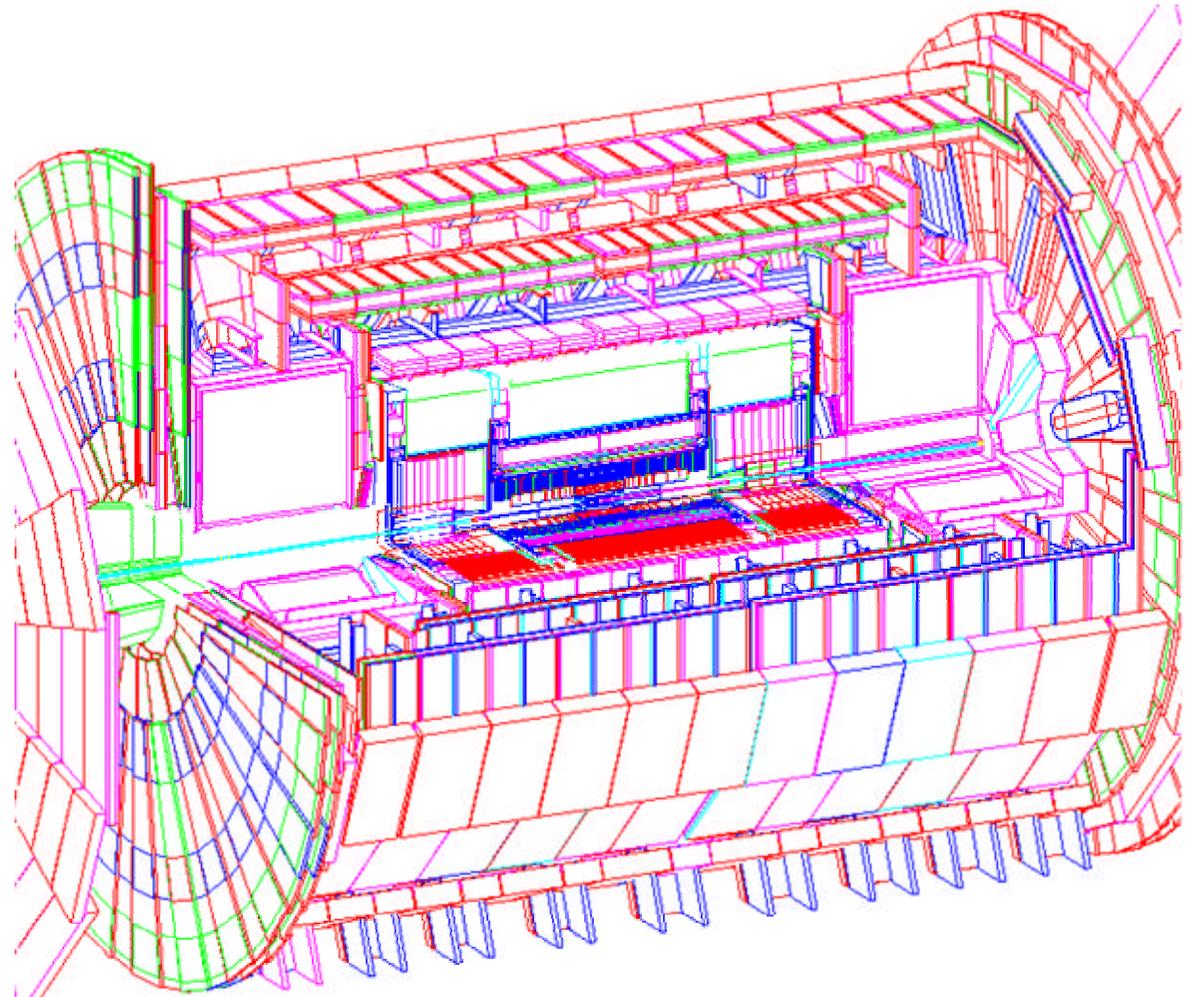
une description complète et fiable du détecteur et la compréhension du signal reconstruit VS signal généré





# ATLAS: La géométrie

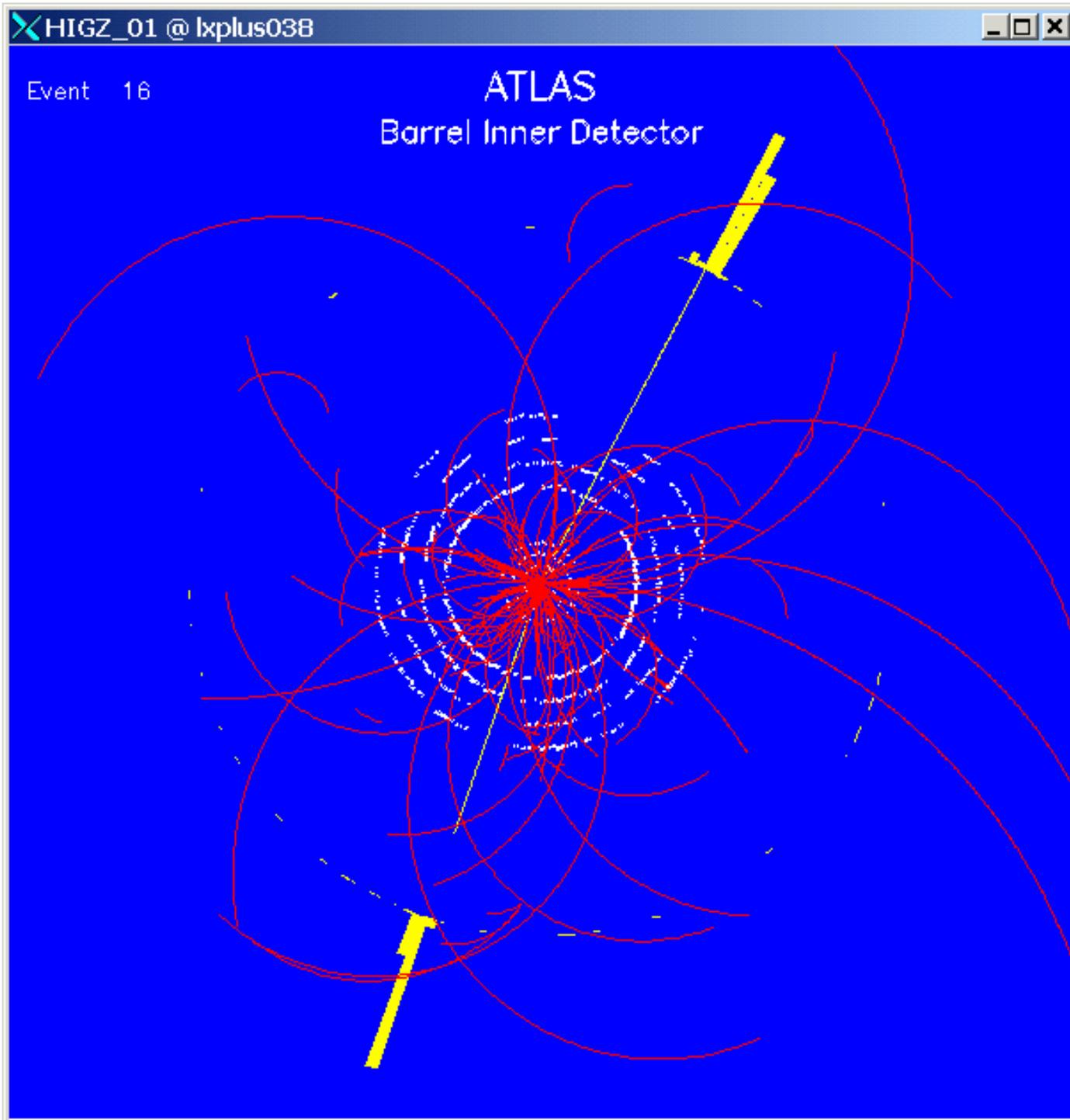
- 25,5 millions de copies de volumes distincts
- 23 000 objets (volumes) différents
- 4 673 types de volumes différents
- Quelques centaines d'événements Pile-up à gérer
- ~ 1 million de hits par événement



# Data Challenges

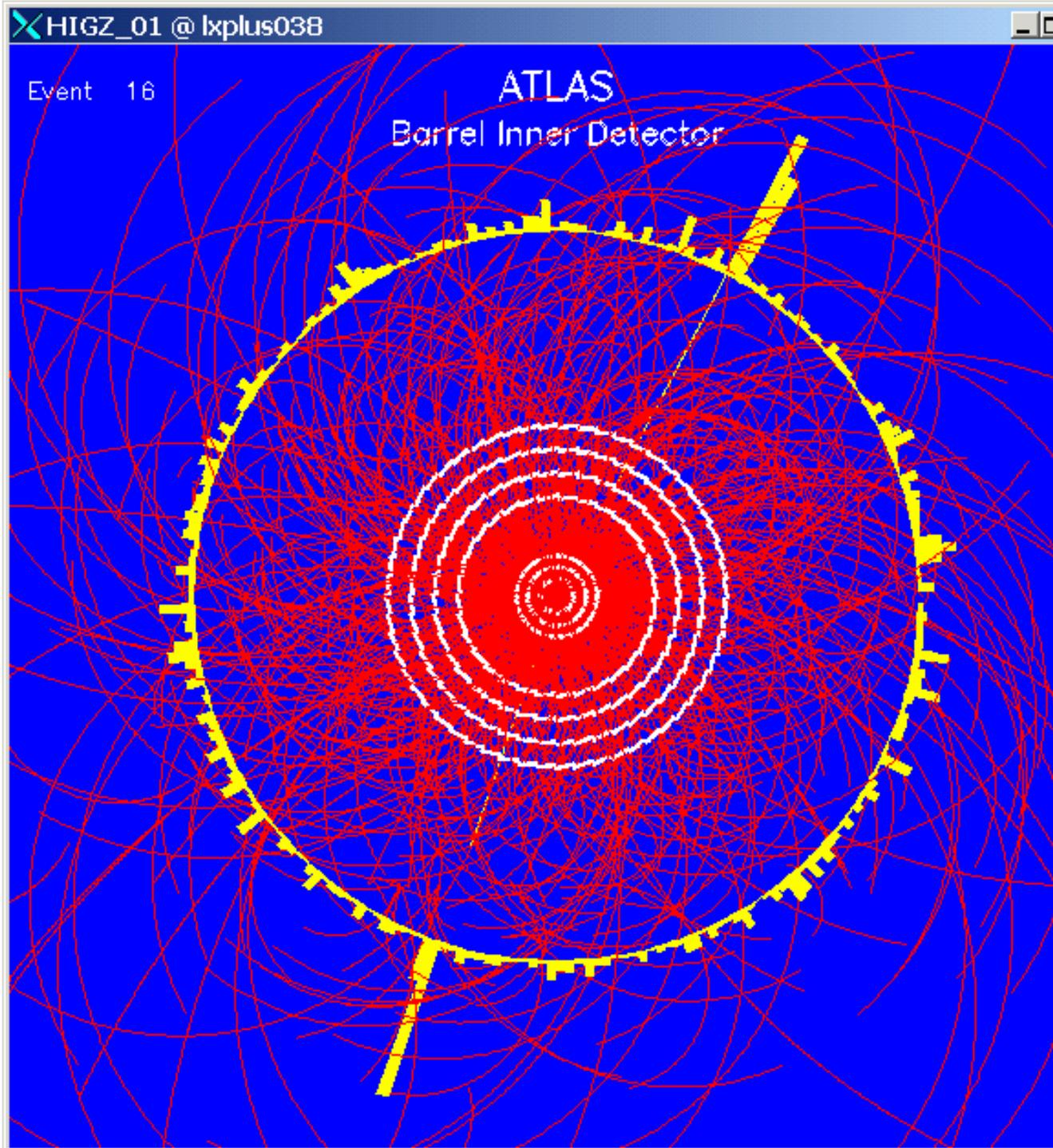
## Tenir compte

- du pile up (empilement)
- de la luminosité



$H \rightarrow \gamma\gamma$

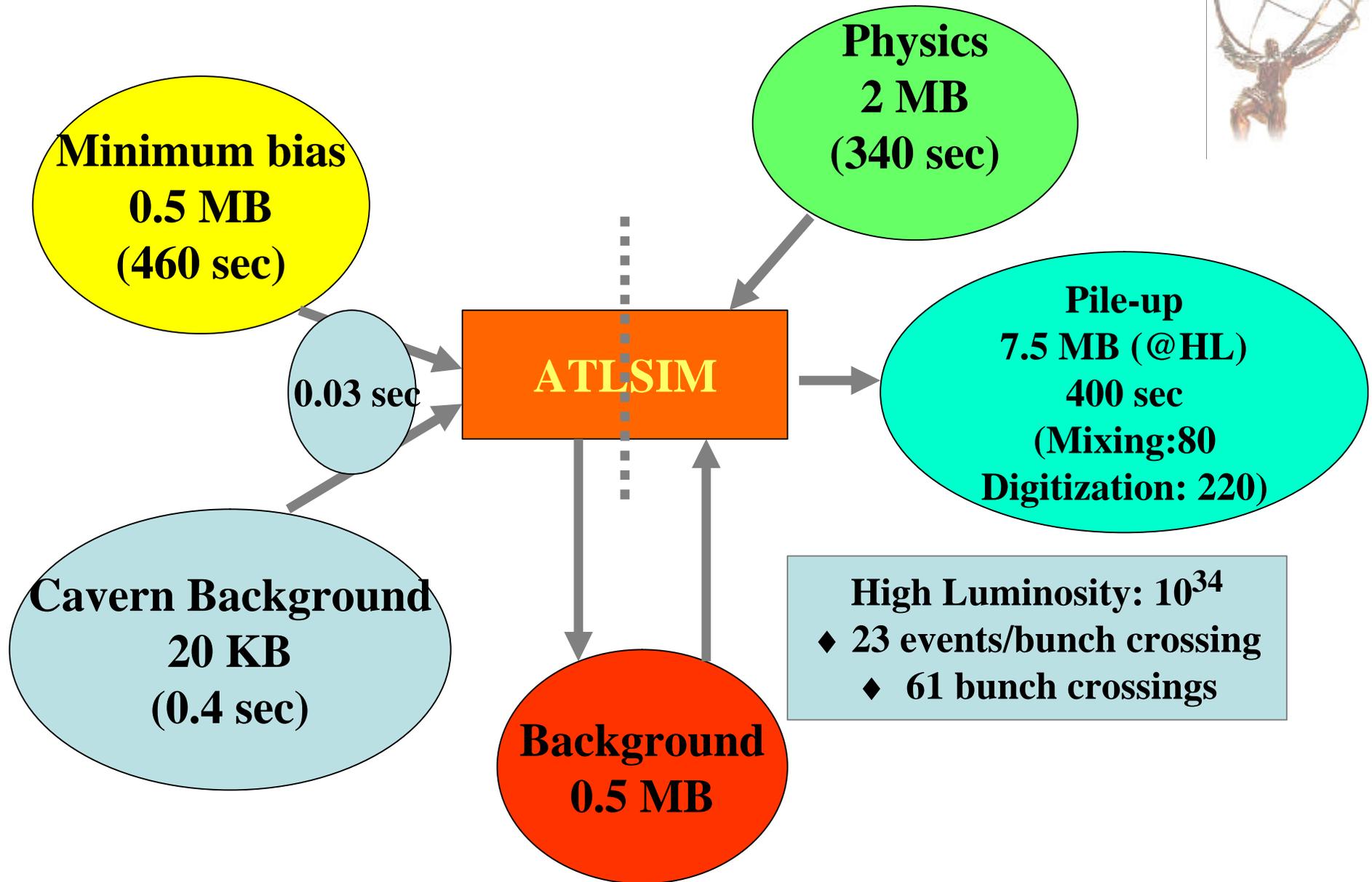
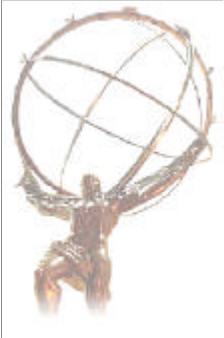
pas de  
pile-up



$H \rightarrow \gamma\gamma$

pile-up à  
 $L = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$

# Comment simuler le Pile-up



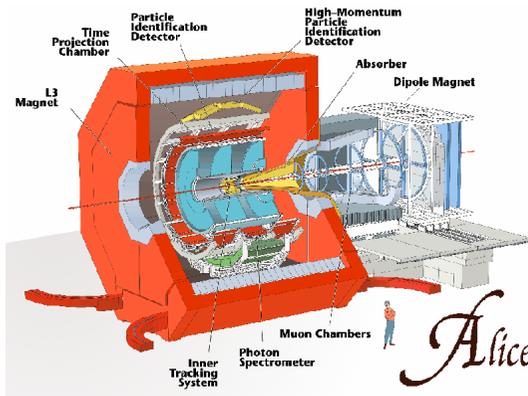
# Data Challenges

## Simuler

- le trigger (déclenchement)
- l'acquisition de données (DAQ)



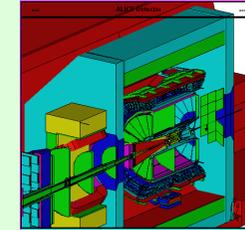
# ALICE Data challenges



Raw Data

Simulated Data

## AliRoot



GEANT3  
GEANT4  
FLUKA

## DAQ

ROOT I/O

CERN  
TIER 0  
TIER 1



## ROOT

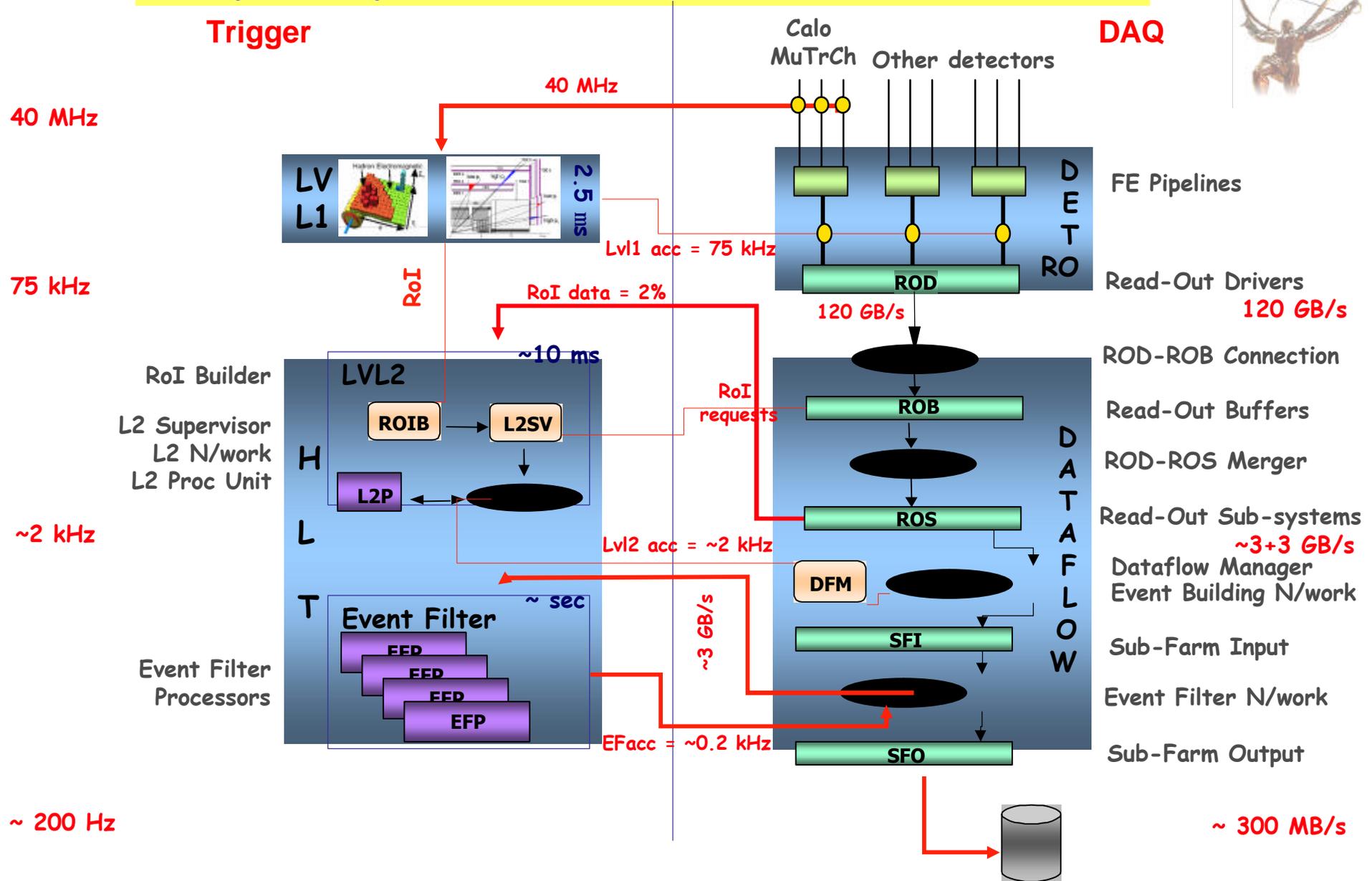
## CASTOR

## GRID

Regional  
TIER 1  
TIER 2



# Le principe de l'acquisition de données



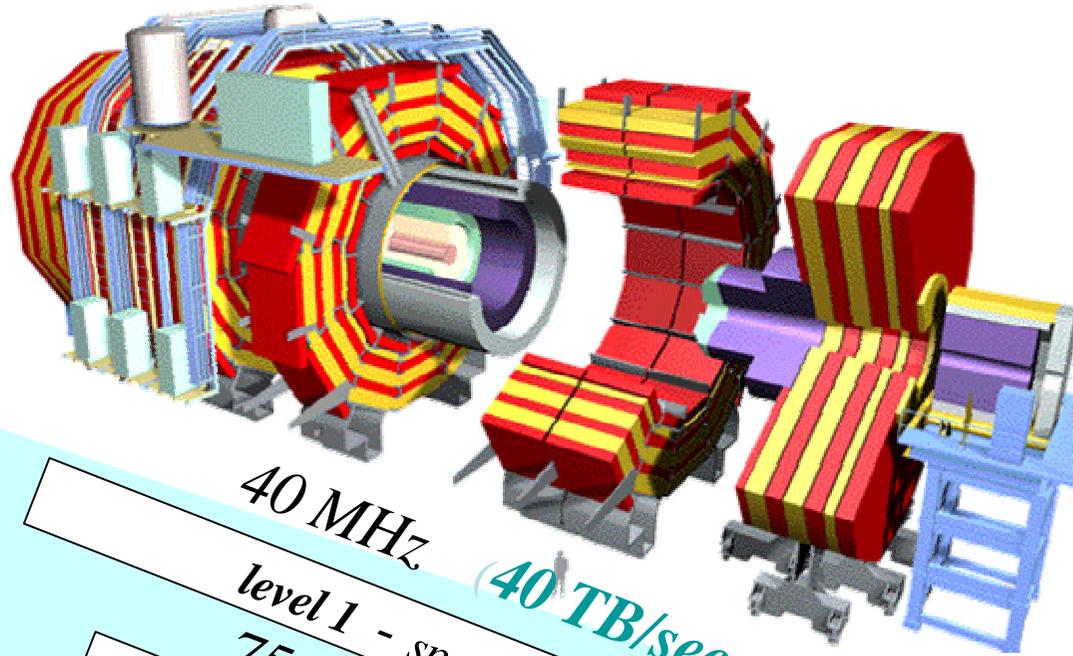
# Data Challenges

Les « grands » nombres:

- le taux de données
- le stockage



# Le défi de 2007



# Le défi de 2007

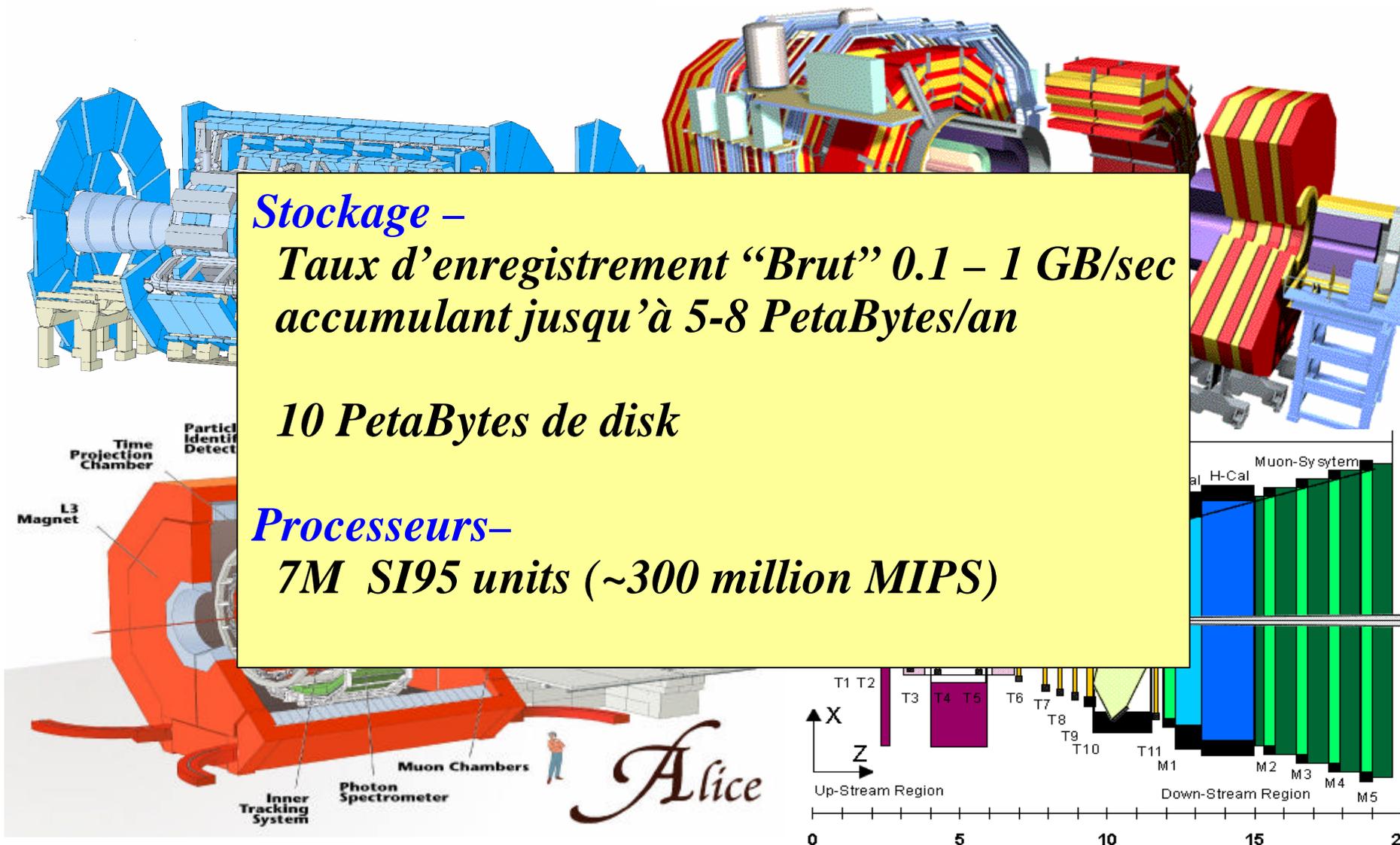
## *Stockage –*

*Taux d'enregistrement "Brut" 0.1 – 1 GB/sec  
accumulant jusqu'à 5-8 PetaBytes/an*

*10 PetaBytes de disk*

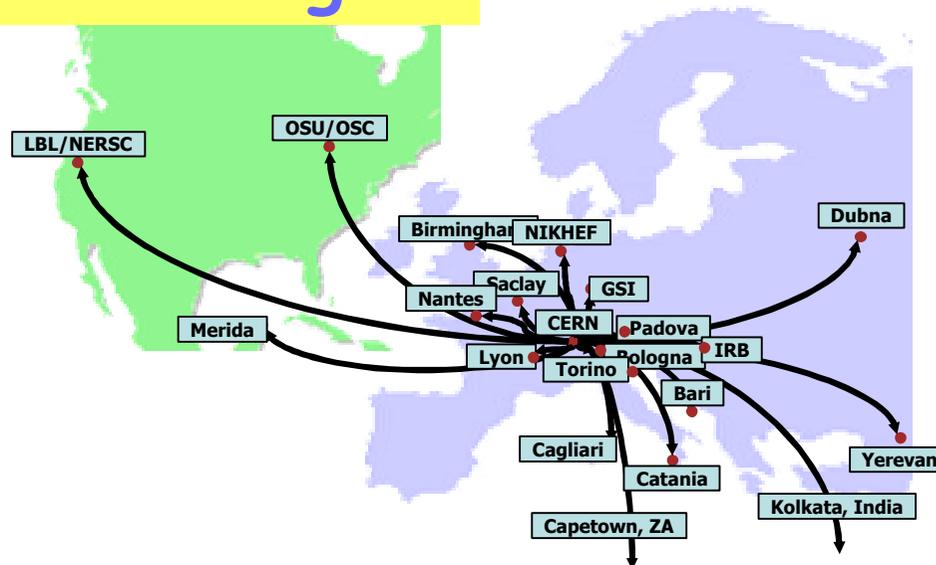
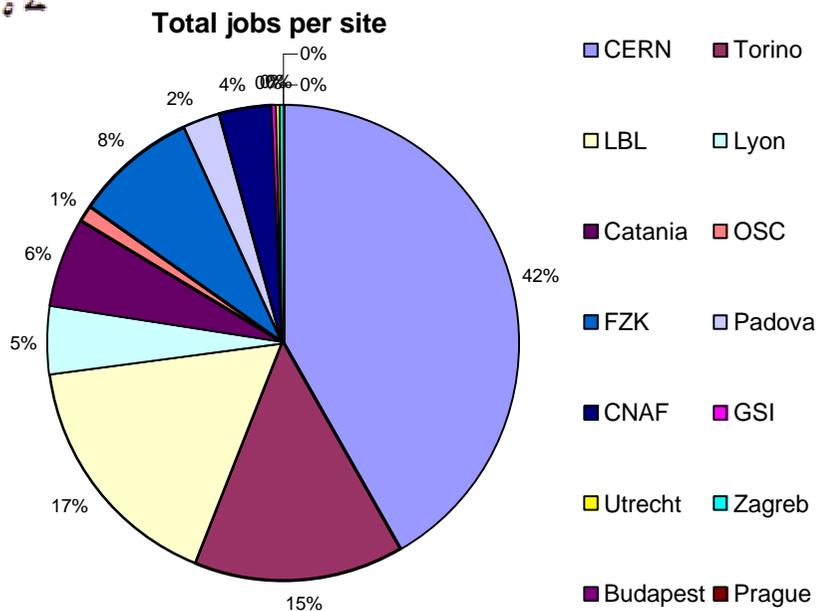
## *Processeurs–*

*7M SI95 units (~300 million MIPS)*





# ALICE Data challenges



- ◆ 32 sites configurés
- ◆ 5 sites procurant le stockage des données
- ◆ 22773 jobs validés
- ◆ Jusqu'à 450 jobs concurrents



## CMS Data Challenges: Production 2002

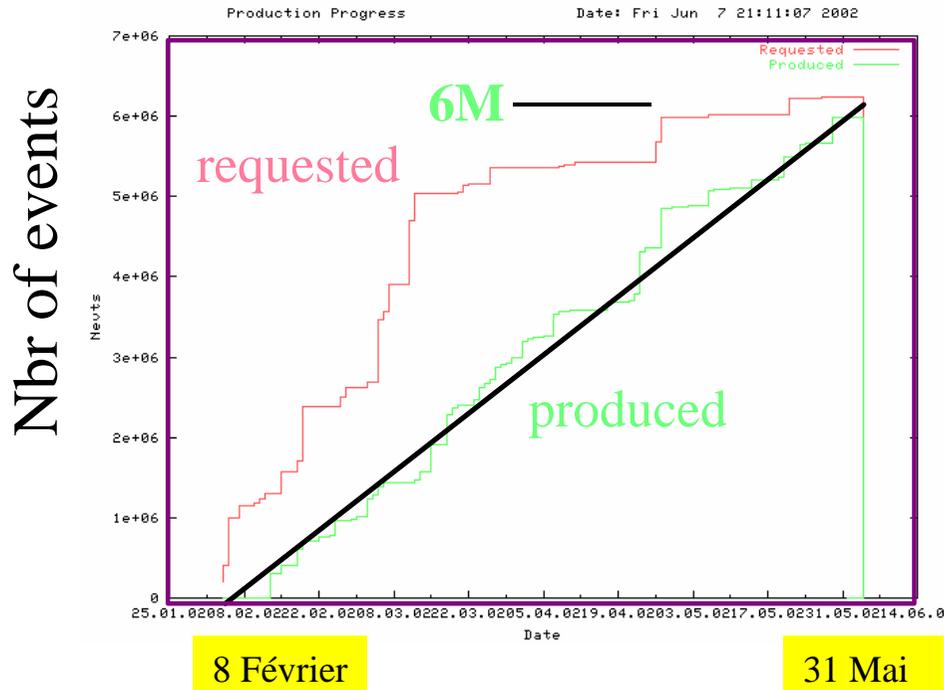
Nombre de Centres régionaux	11
Nombre de Centres de calcul	21
Nombre de CPU	~1000
Nombre de Fichiers	~11,000
Quantités de données	17 TB

### Participants:

Bristol/RAL, Caltech, CERN, FNAL, IC, IN2P3, INFN, Moscou, UCSD, UFL, WISC

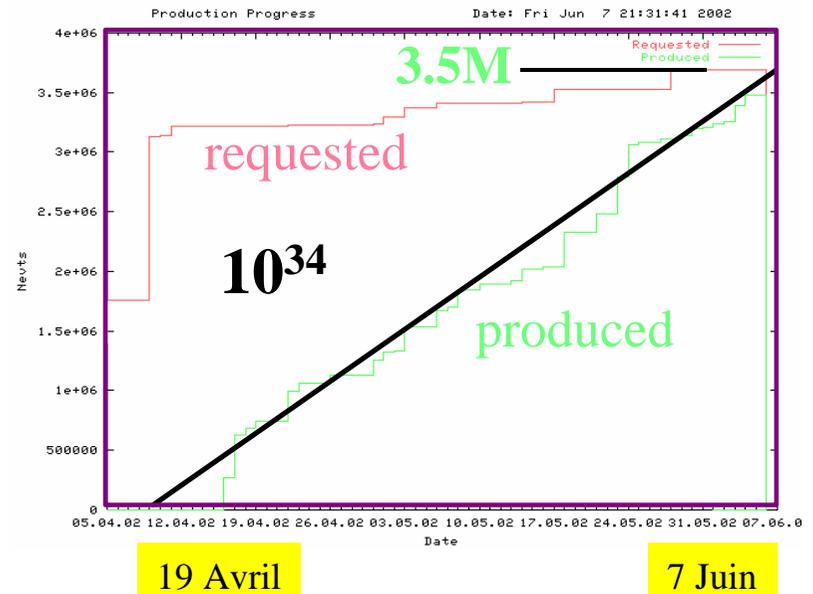


# CMS Data Challenges: Production 2002



Digitisation Haute luminosité:  
1.4 secondes / ev. pendant 2 mois

CMSIM:  
1.2 secondes / ev. Pendant 4 mois



# ATLAS Data Challenges 1, Phase I



- Délivrer des événements à la communauté HLT (trigger de haut niveau): **but:  $\sim 10^7$  événements**
- Mettre en place le Monte Carlo de génération (physique) et de la simulation du détecteur et valider la chaîne
- Mettre en place le MonteCarlo de production, "distribué"
  - **"ATLAS kit" (rpm)**
  - Scripts et outils (monitoring, bookkeeping)
    - **AMI database; Magda replica catalogue; VDC**
  - Quality Control et Validation de la chaîne complète

# ATLAS Data Challenges 1, Phase I



3200 CPU's  
110 kSI95  
71000 CPU days

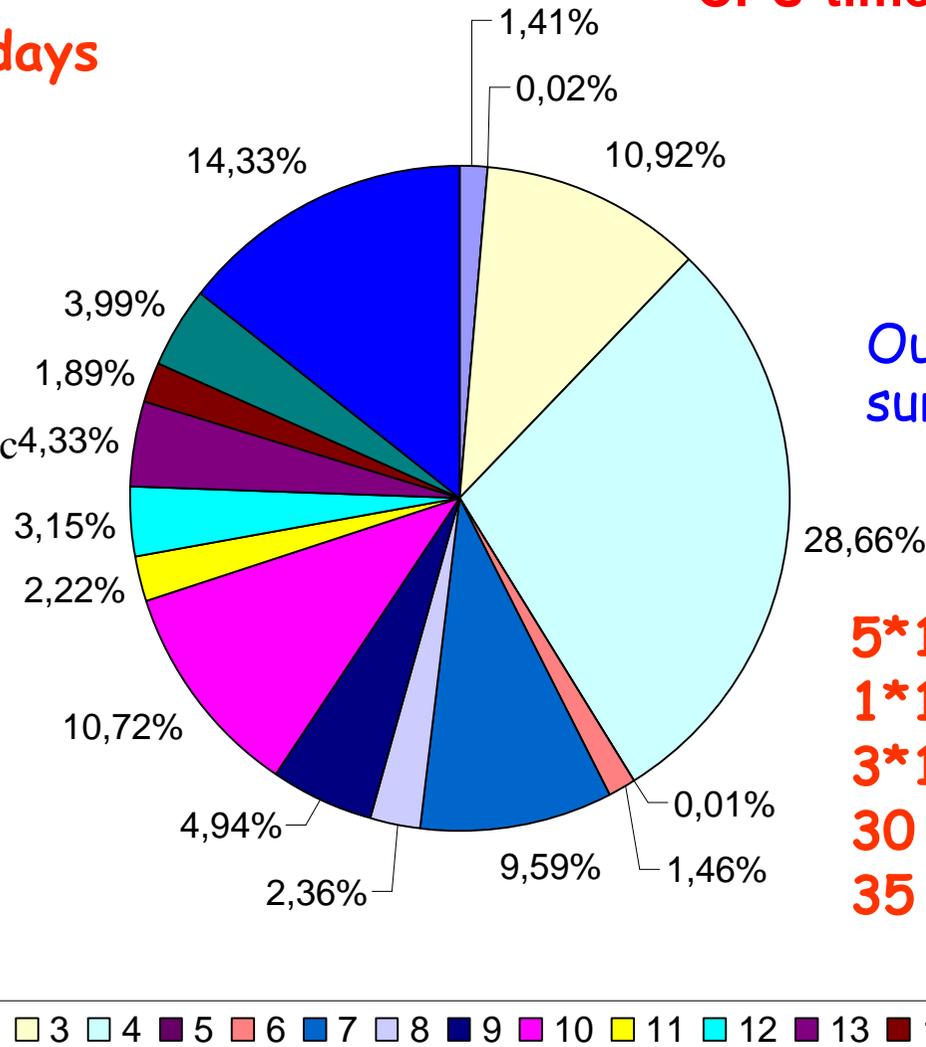


50% du CPU estimé  
pour 1 centre régional

Contribution to the overall  
CPU-time (%) per country

39 Instituts  
18 Pays

1. Australia
2. Austria
3. Canada
4. CERN
5. Czech Republic
6. France
7. Germany
8. Israel
9. Italy
10. Japan
11. Nordic
12. Russia
13. Spain
14. Taiwan
15. UK
16. USA



Outils grid utilisés  
sur 11 sites

5\*10<sup>7</sup> ev. générés  
1\*10<sup>7</sup> ev. simulés  
3\*10<sup>7</sup> single particles  
30 Tbytes  
35 000 files

# ATLAS Data Challenges 1, Phase II



- Délivrer des données avec et sans **pile up** au HLT
- Introduction et test du nouveau "Event Data Model" (EDM)
- Production de données pour la physique et le "Computing Model studies"
- Tester la **Reconstruction** des données
- Utilisation de GRID de manière plus large

# ATLAS DC1 Phase 2 : Nov-Dec 2002



**7800 'CPU's'**

**170 kSI95**

**56 Institutes in 21 countries participate**

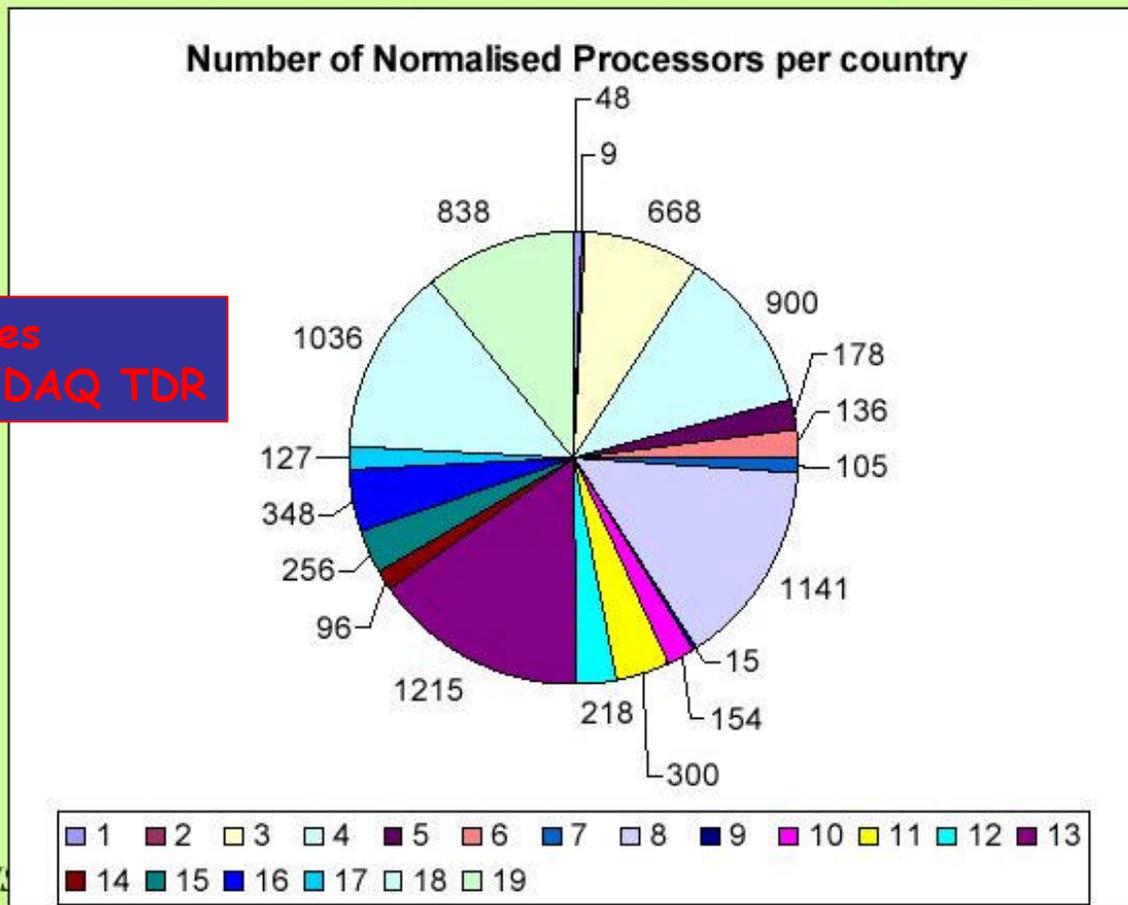
**20 Institutes use Grid middleware in production**

**In preparation for upcoming Grid tests : Grid middleware already installed at most places**

1. Australia
2. Austria
3. Canada
4. CERN
5. China
6. Czech Republic
7. France
8. Germany
9. Greece
10. Israel
11. Italy
12. Japan
13. NorduGrid
14. Poland
15. Russia
16. Spain
17. Taiwan
18. UK
19. USA

**More data samples  
Pile-up for HLT/DAQ TDR**

*Alois Putzer, I*

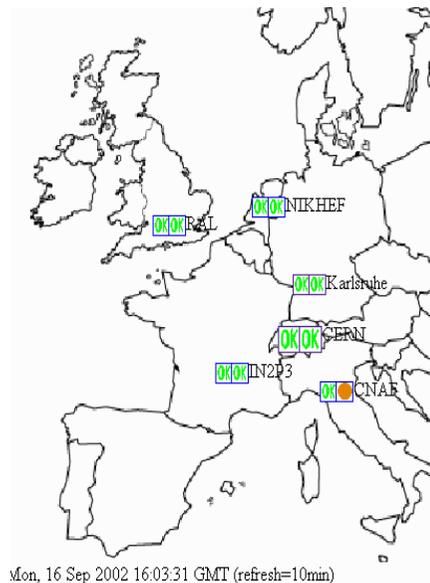
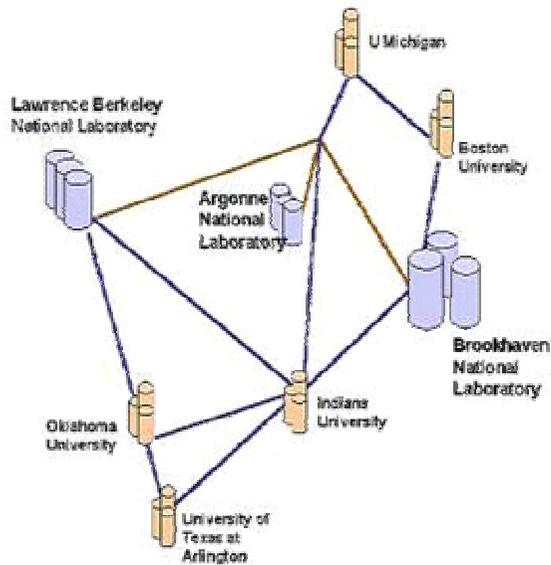


## Après ATLAS DC1 ...



- Mettre en place une infrastructure pour la production de masse
- Avoir un "logiciel" de la reconstruction prêt et validé
- Inclure les codes HLT
  - Lvl1, Lvl2 & Event Filter
- Données concentrées sur 8 "sites"
- La Production sur "standard batch" et/ou "GRID"

# GRID dans ATLAS DC1



## US-ATLAS

Une partie de la phase 1  
Production complète phase 2

CUIC 03, 17-18 juin 2003

## EDG Testbed Prod

Reproduction d'une  
partie de la phase 1

Data Challenges au LHC

## NordGrid

Toute la production  
Phase 1 et 2

F. Ohlsson-Malek  
LPSC/IN2P3

## La suite ... Conclusion

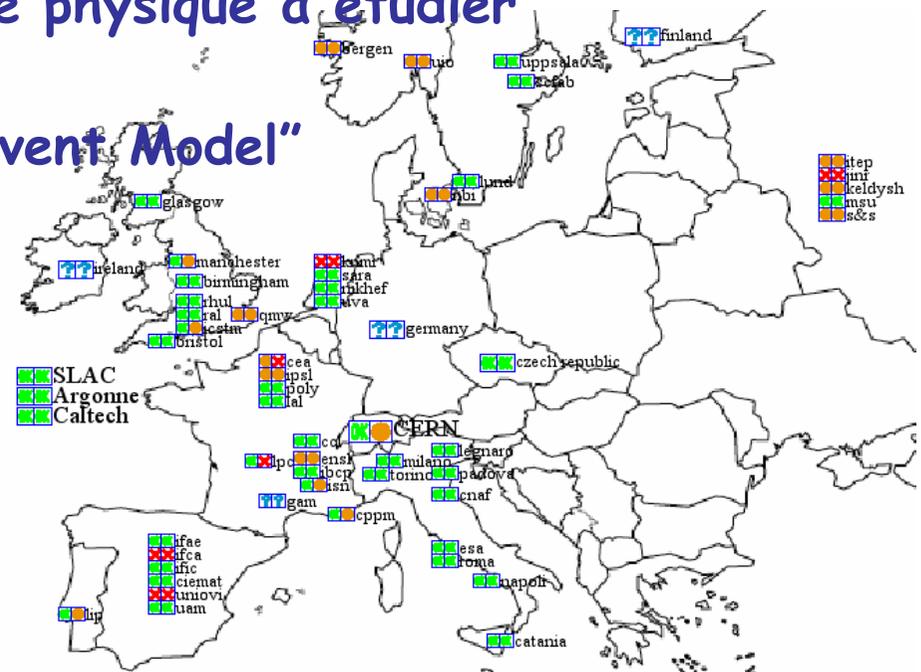
DCs plus complexes + Data-Grid



# CMS Data challenges 2004

## Reconstruction de 50 million d'événements

- Correspond à 25 Hz à  $2 \times 10^{33} \text{ cms}^{-2} \text{ s}^{-1}$  pendant 1 mois
- Supposés être des événements dans le centre Tier-0, i.e. avec HLT
- Définir des lots d'événements de physique à étudier
- Validation de la physique
- Préparer l'analyse <sup>®</sup> tester l'"Event Model"
  - Calibration et alignement
- Intégration et développements d'outils **EDG dans CMS** et en permettre la production



Fri, 07 Jun 2002 13:55:04 GMT (refresh=10min)

# ATLAS DC2-3-4-...



- **DC2:**
  - Période: T4/2003 - T2/2004
  - But
    - Déploiement complet de EDM & Detector Description
    - **Geant4 replacing Geant3**
    - Procédures de tests the calibration et d'alignement
    - **Utilisation des outils logiciels du LCG (POOL, ...)**
    - Utilisation systématique de GRID
    - Préparer des lots de données de physique plus grands
  - échelle
    - Comme pour DC1:  $\sim 10^7$  év. complètement simulés
- **DC3: T3/2004 - T2/2005**
  - Buts à définir; échelle: 5 x DC2
- **DC4: T3/2005 - T2/2006**
  - Buts à définir; échelle: 2 X DC3

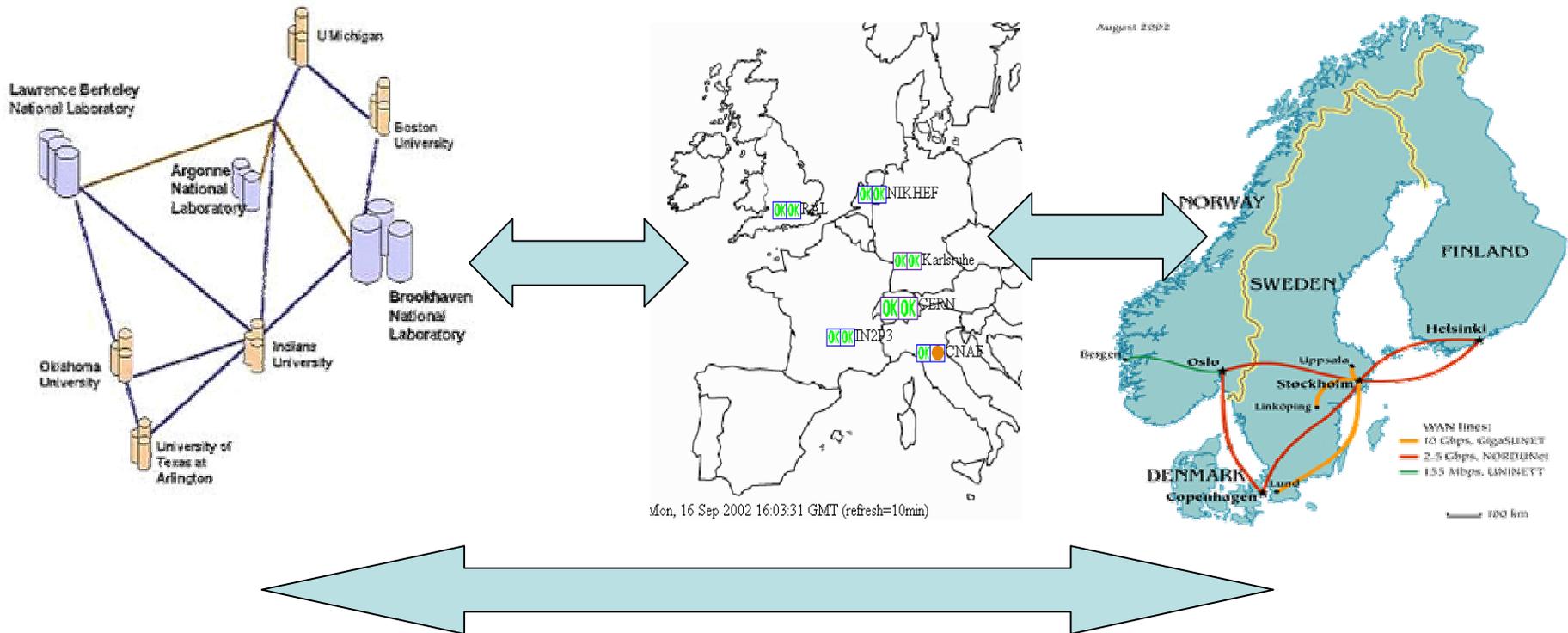
2007



# Défi

LCG

# GRID dans ATLAS ...



**US-ATLAS**

**EDG**

**NorduGrid**

**Et plus .....**

CUIC 03, 17-18 juin 2003

Data Challenges au LHC

F. Ohlsson-Malek  
LPSC/IN2P3