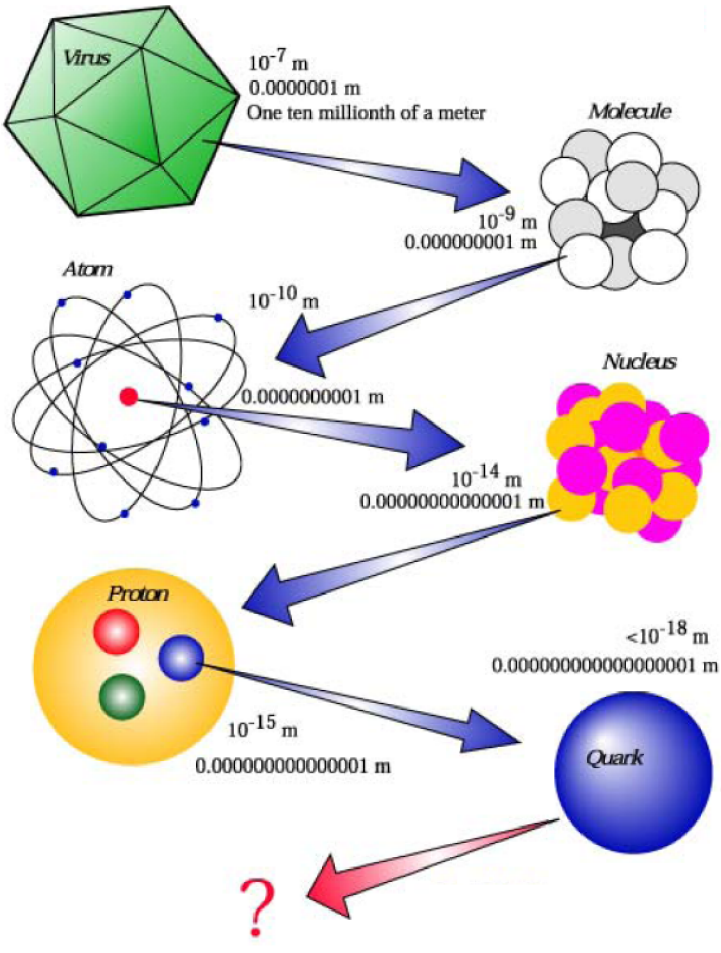
Code optimization in High Energy Physics

- challenges at the LHC -

Nuno Castro (nuno.castro@cern.ch)
António Onofre (antonio.onofre@cern.ch)
Juan Pedro Araque (jp.araque@cern.ch)

ATLAS LIP

3rd December 2013



$$\Delta x \Delta p \geq \frac{\hbar}{2}$$

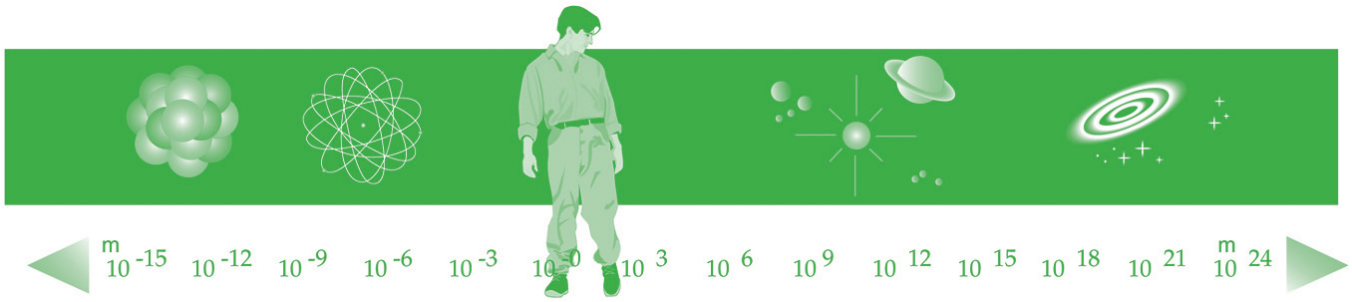
See at small scales

↓

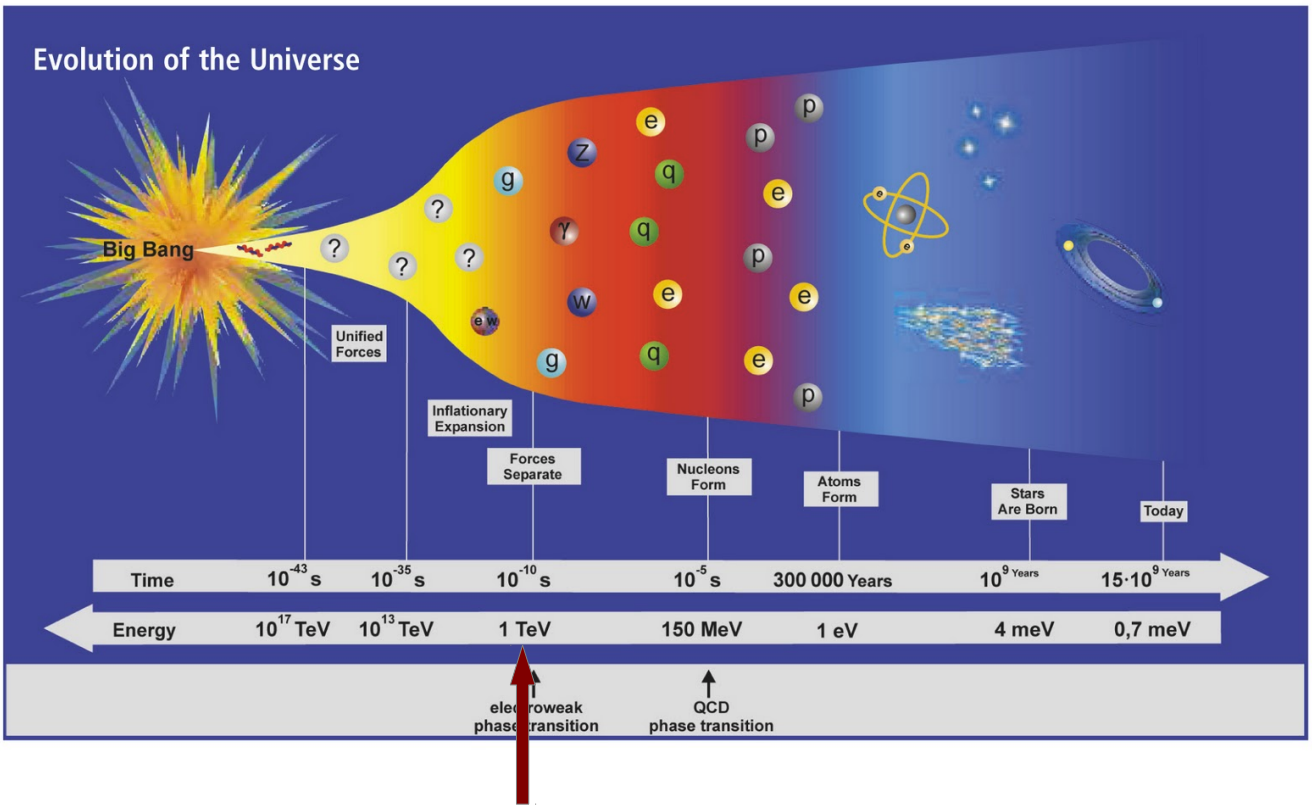
High *p*

High Energy Physics

Particle Physics

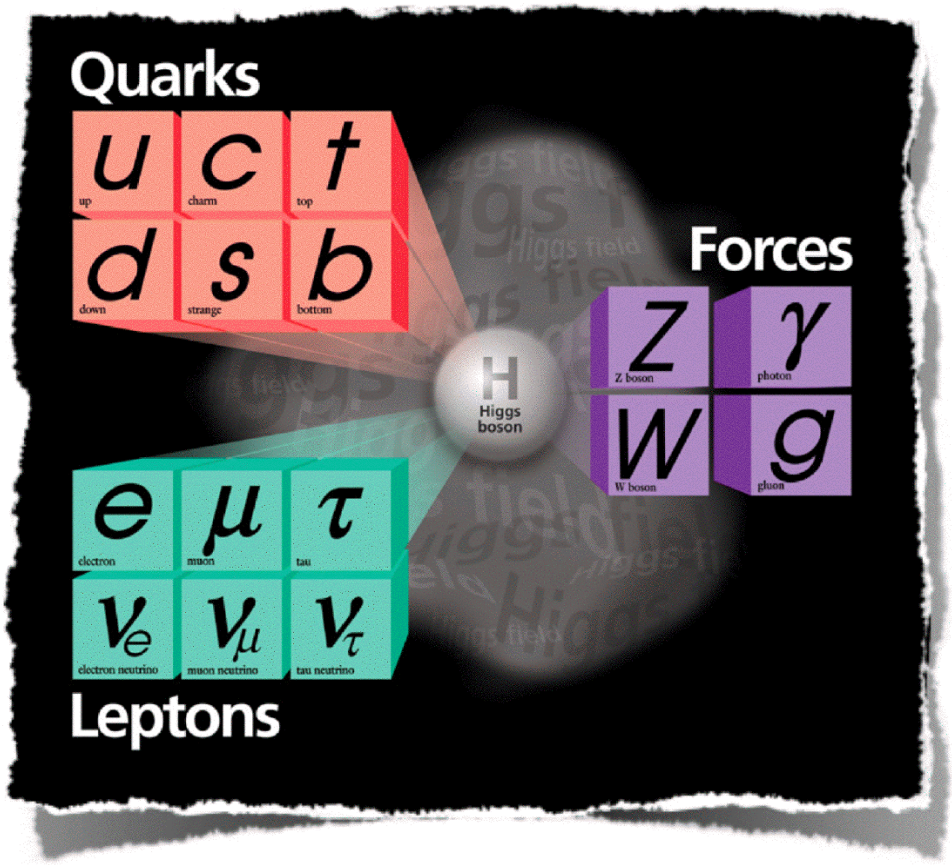


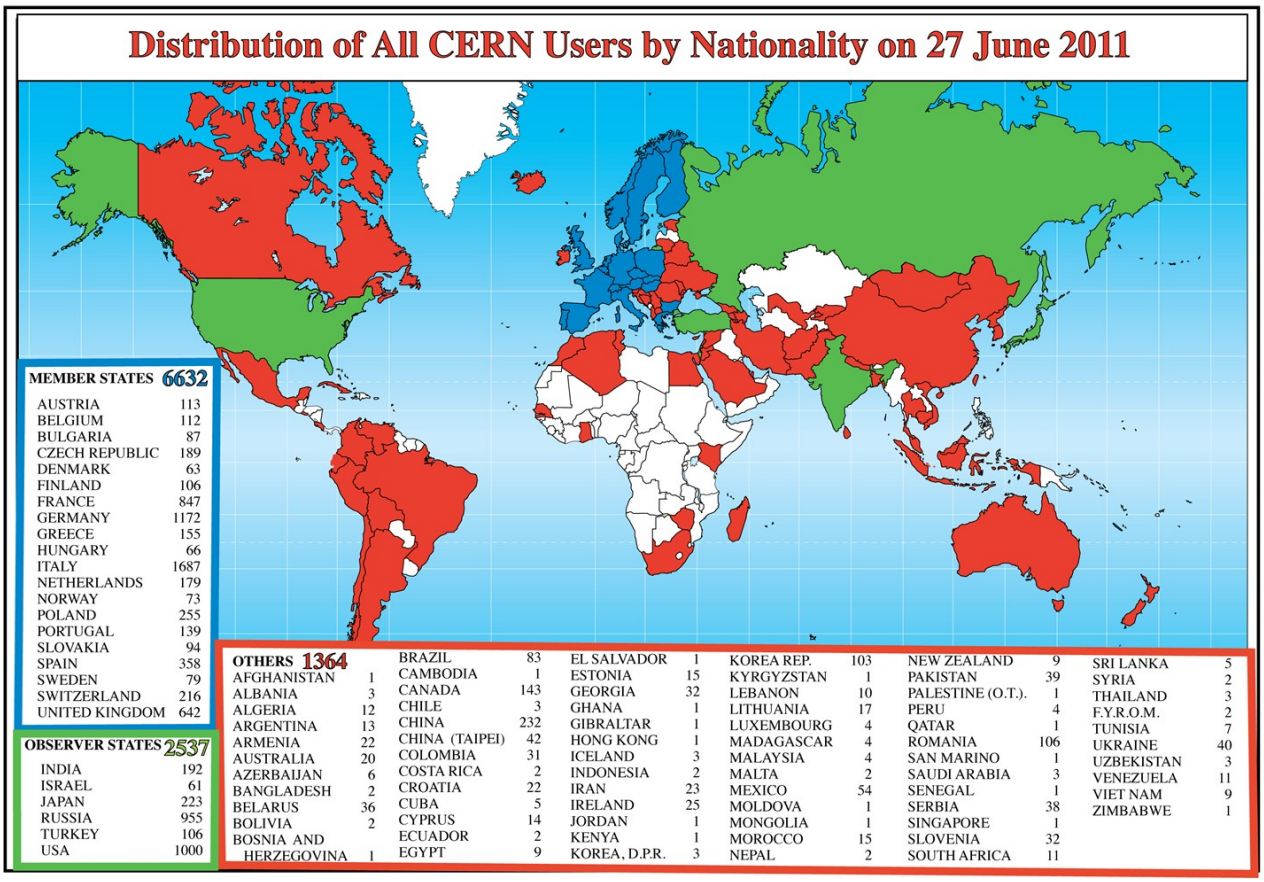
the 2 frontiers:
very big and very small



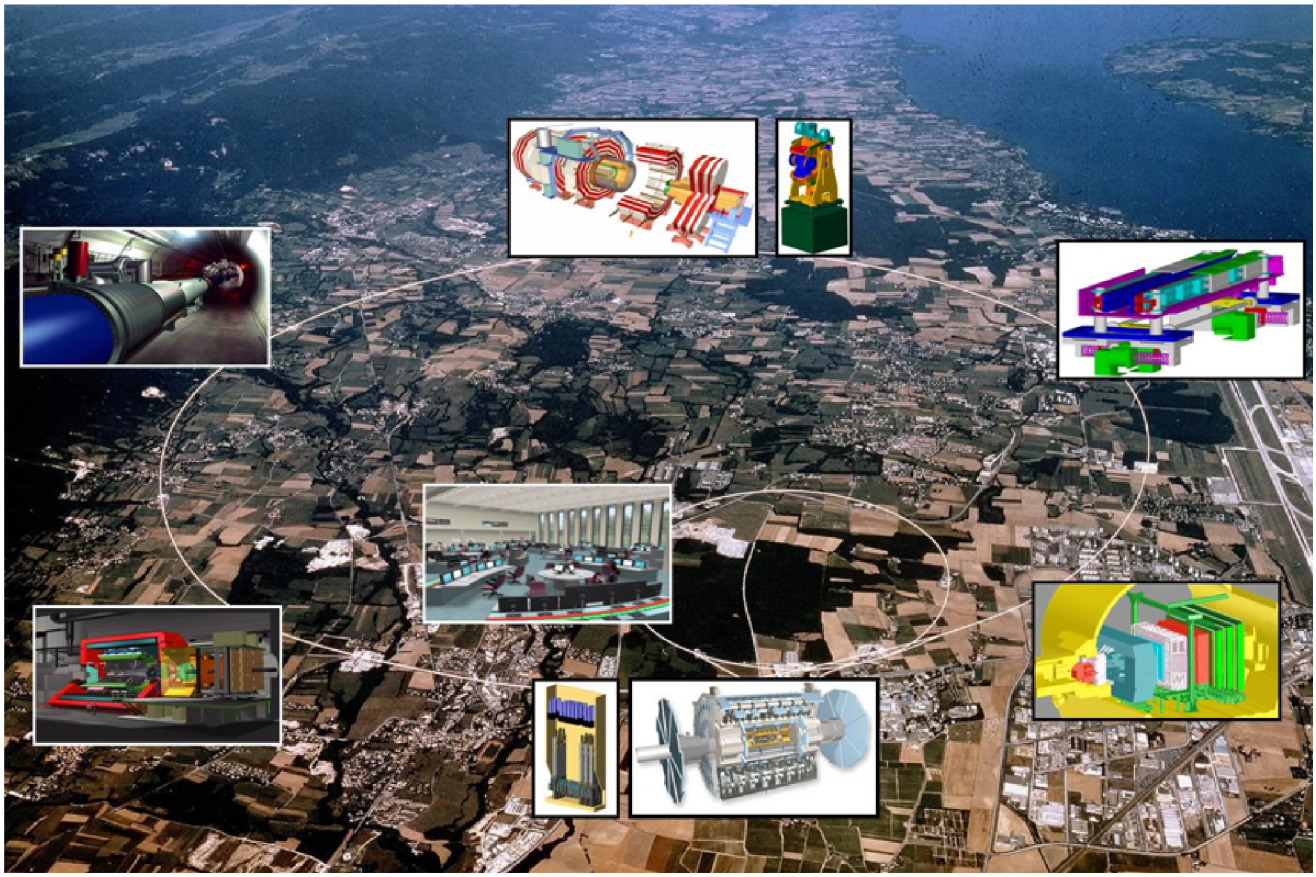
Large Hadron Collider (LHC)

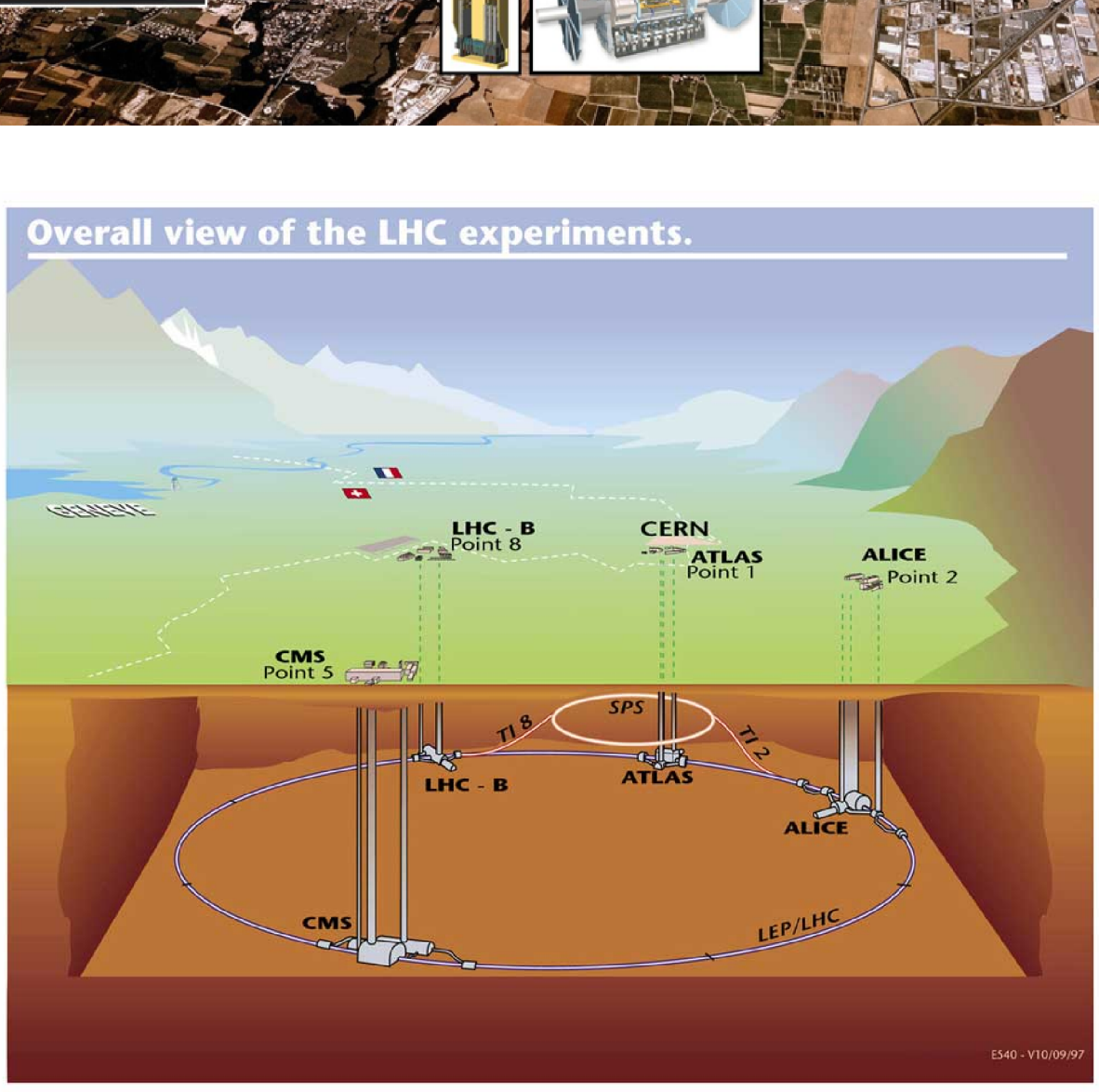
Standard Model of Particle Physics

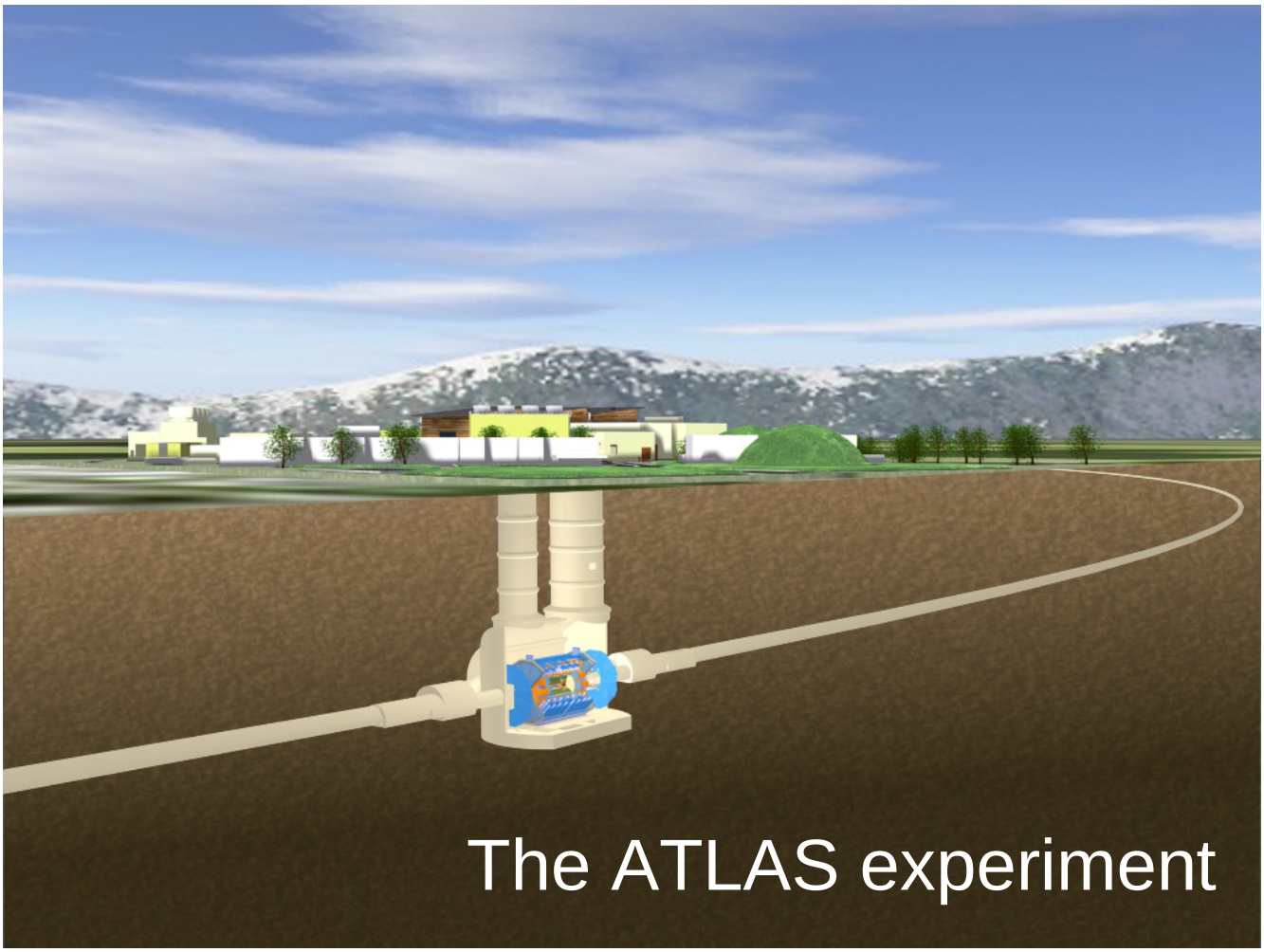




The LHC

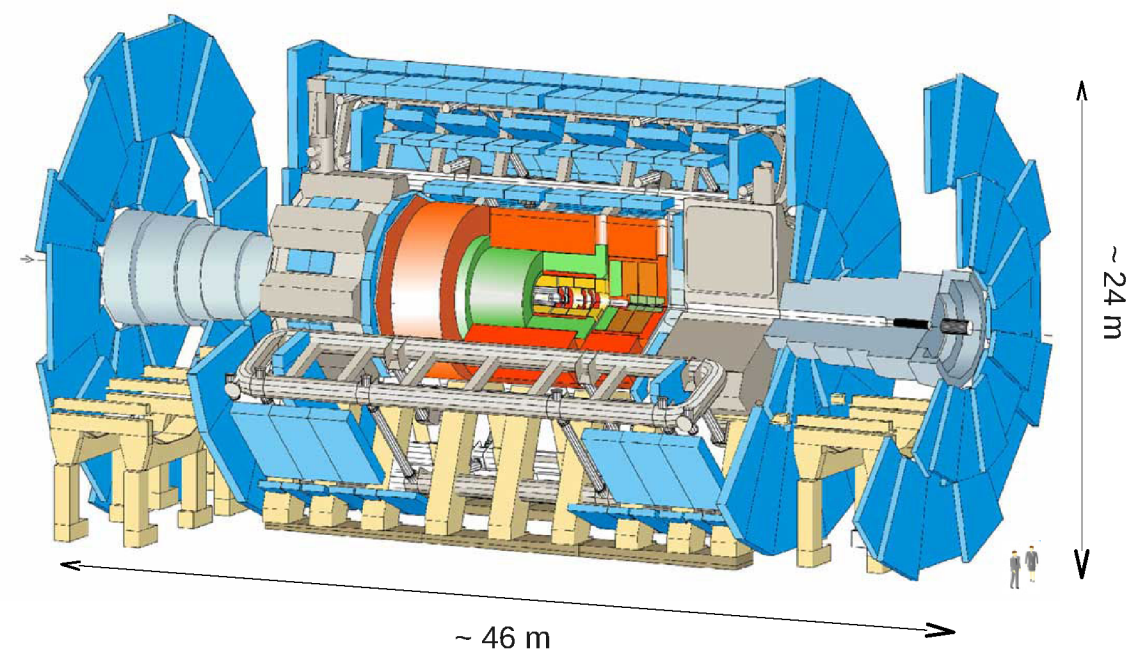



Overall view of the LHC experiments.
GENEVE
LHC - B
Point 8
CERN
ATLAS
Point 1
ALICE
Point 2
CMS
Point 5
SPS
TI 8
TI 2
LHC - B
ATLAS
ALICE
CMS
LEP/LHC
E540 - V10/09/97

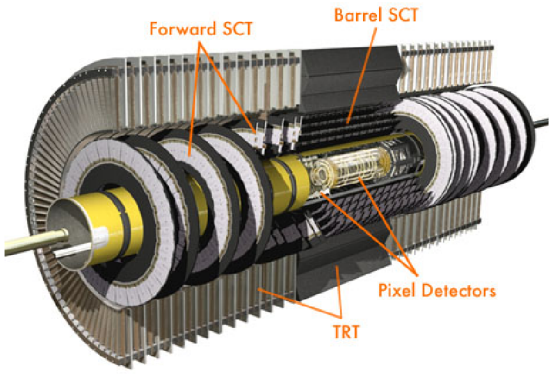

The ATLAS experiment

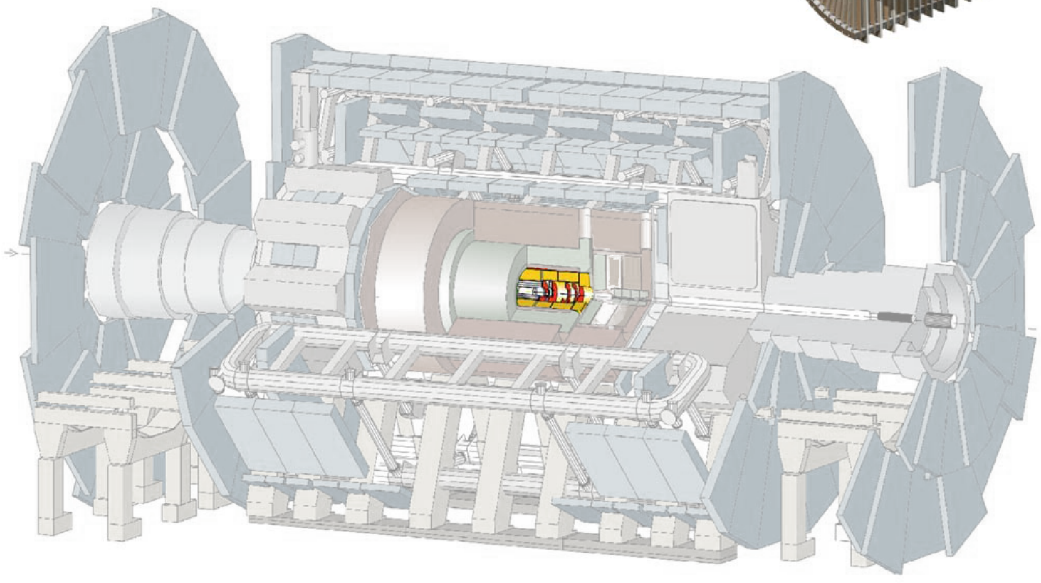
The ATLAS detector

~ 7000 t
~ 10^8 channels
~ 3000 km of cables
~ 40 x 10^6 collisions / s
~ 1 PB of data / s



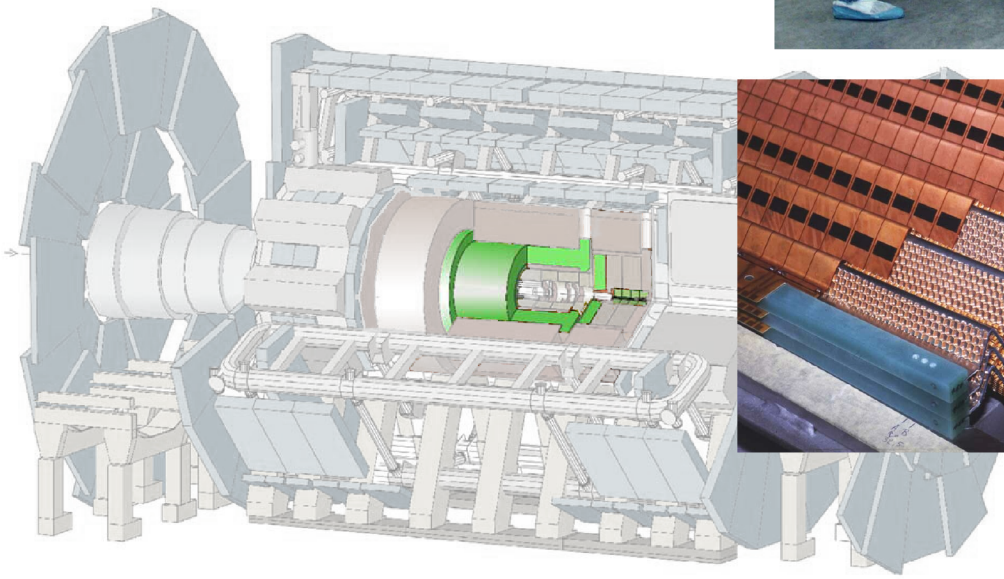
Inner detector

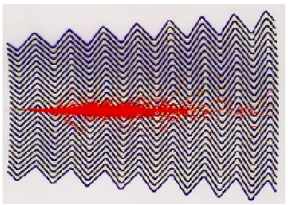




$\sigma/p_T \sim 3.8 \times 10^{-4}\ p_T\ (\text{GeV}) \oplus 0.015$
i.e. σ/p_T <2% for p_T < 35 GeV

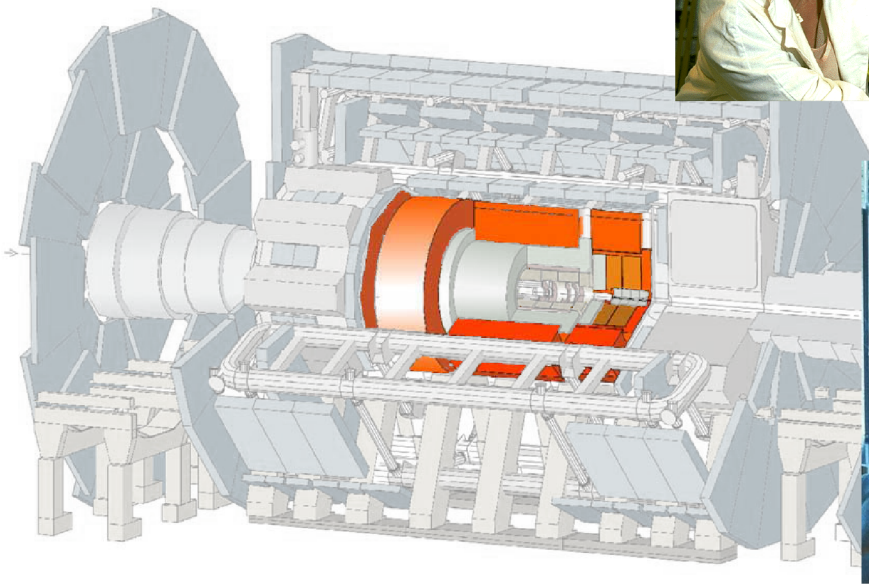
Electromagnetic calorimeter





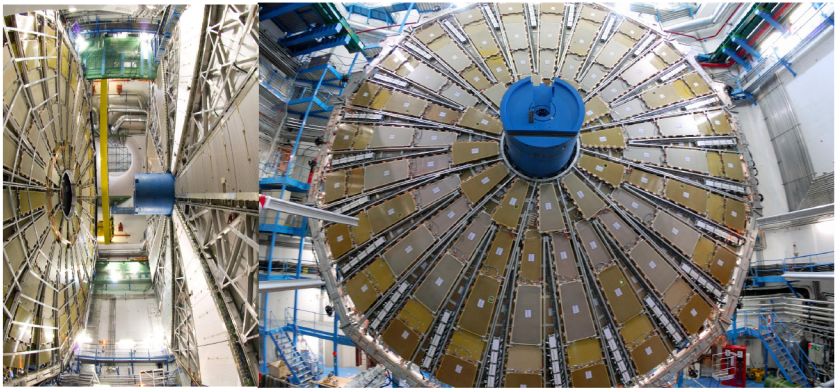
$\sigma/E \sim 10\%/\sqrt{E}$

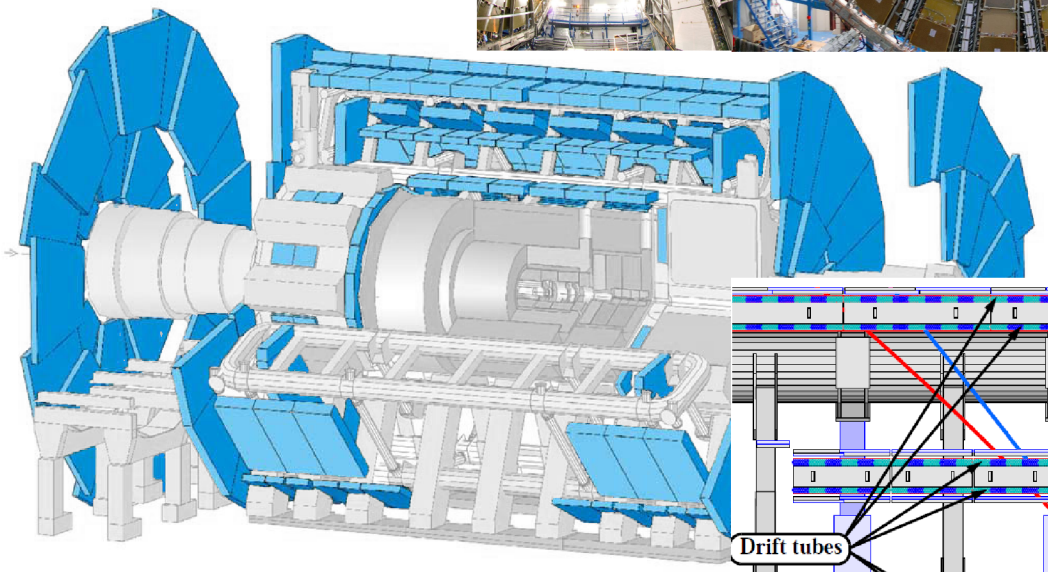
Hadronic calorimeter

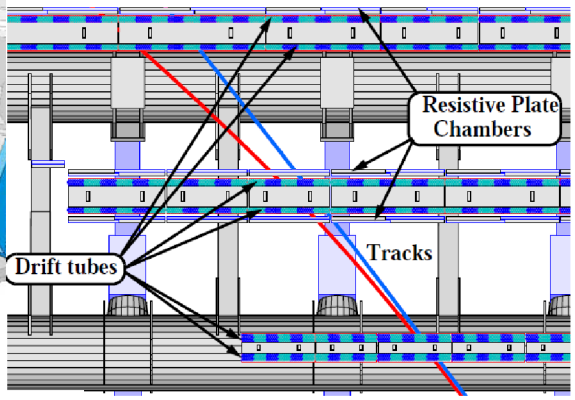


$\sigma/E \sim 50\%/\sqrt{E} \oplus 0.03$

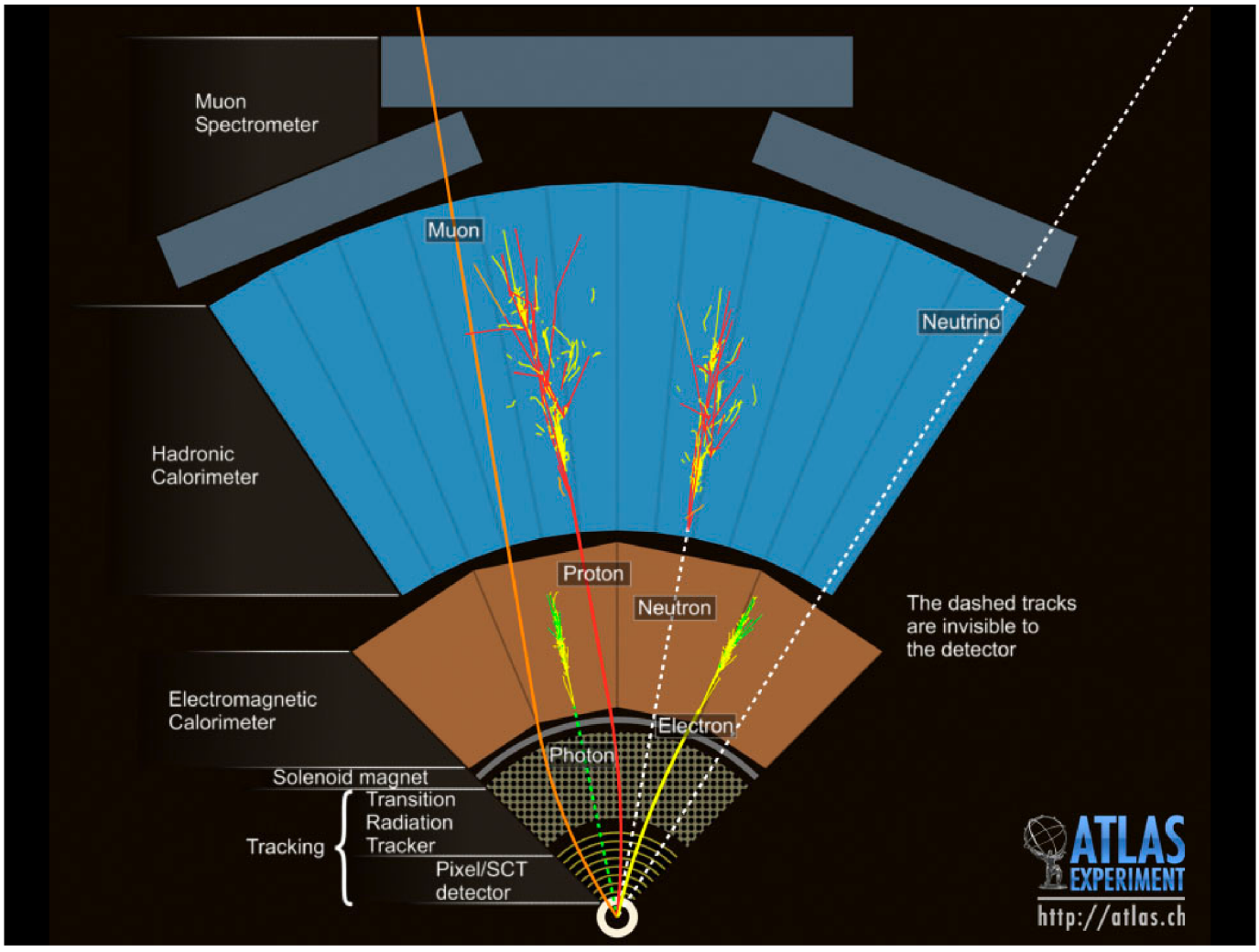
Muons detector



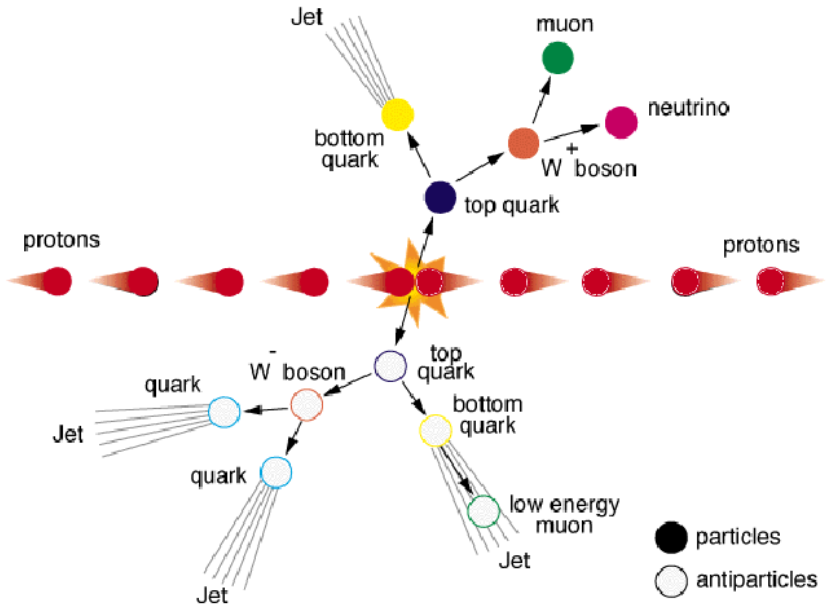




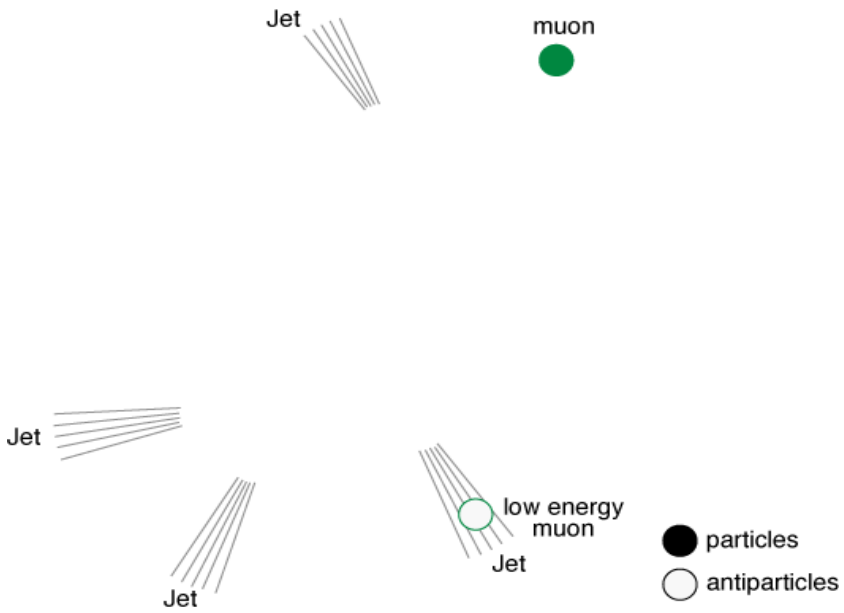
< 10% up to E_μ ~ 1 TeV

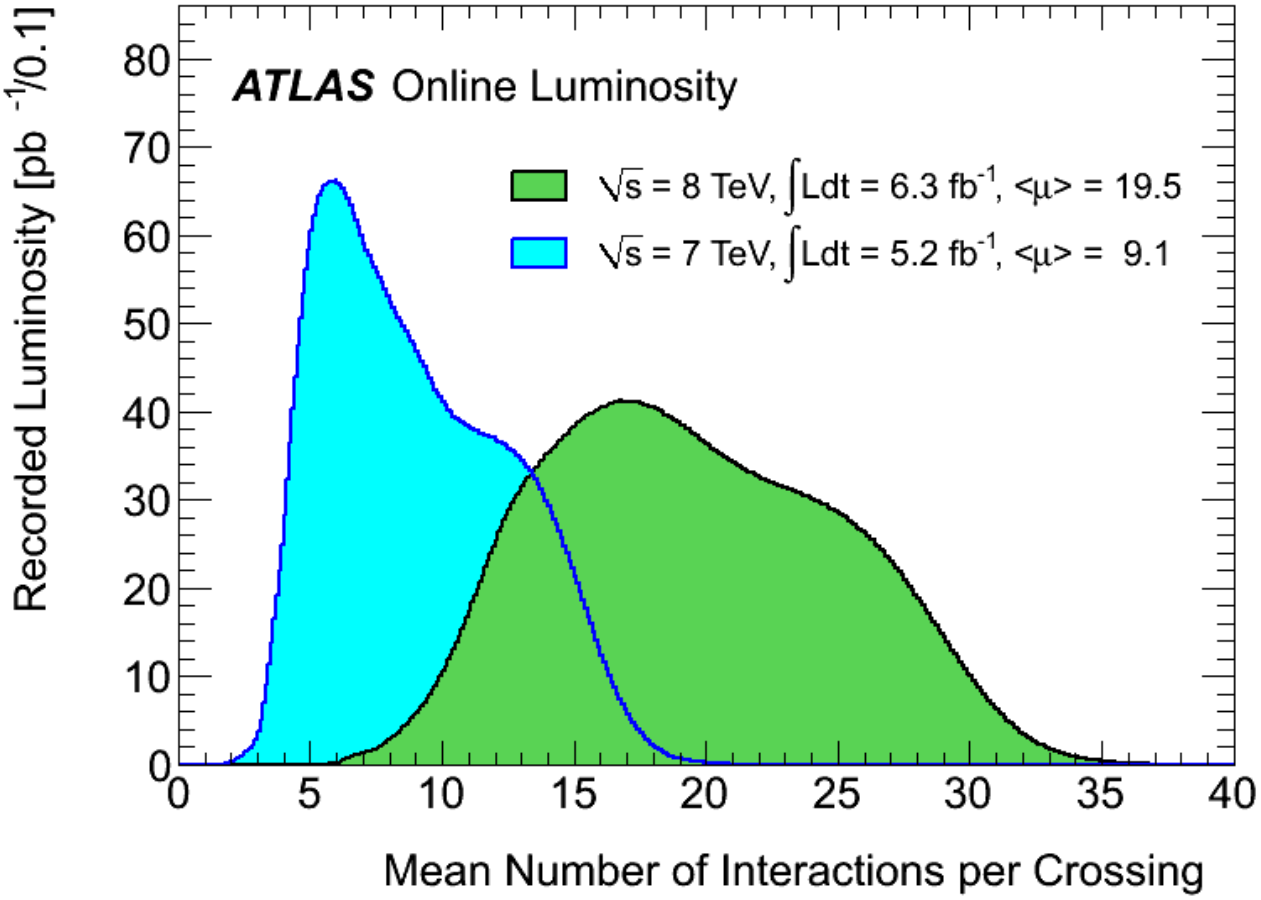


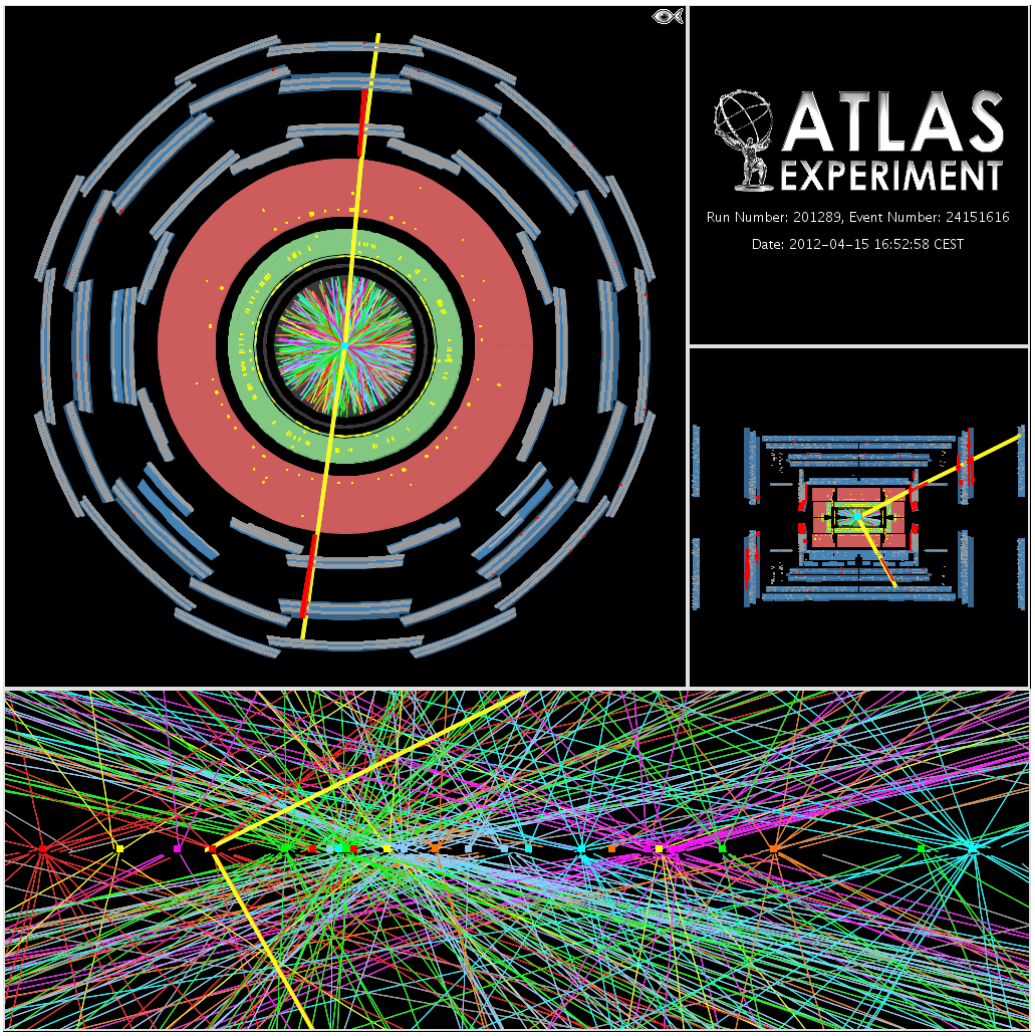
What happens in a collision?



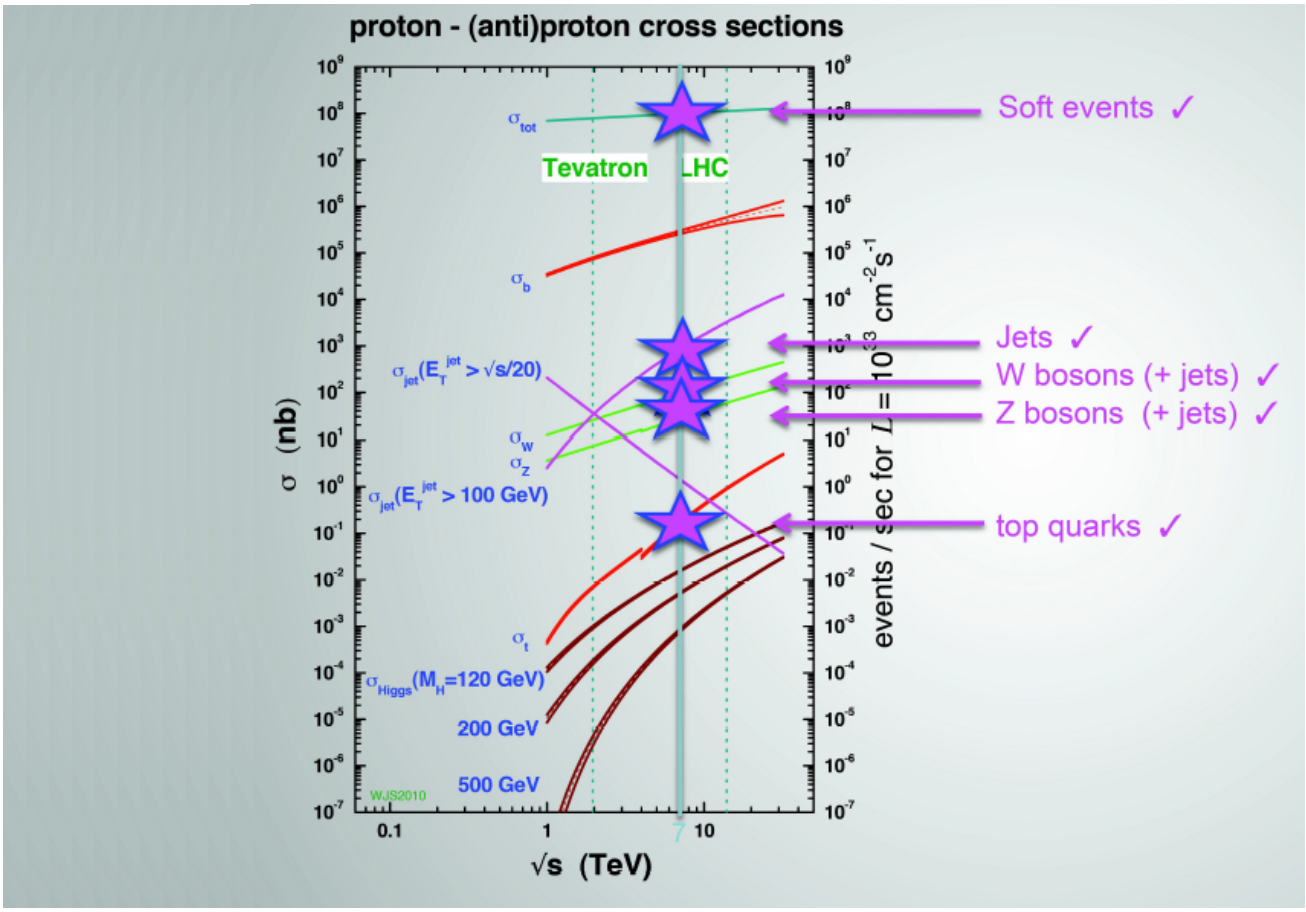
What do we see in the detector?




ATLAS Online Luminosity
√s = 8 TeV, ∫Ldt = 6.3 fb-1, <μ> = 19.5
√s = 7 TeV, ∫Ldt = 5.2 fb-1, <μ> = 9.1
Recorded Luminosity [pb -1/0.1]
Mean Number of Interactions per Crossing
0
10
20
30
40
50
60
70
80
0
5
10
15
20
25
30
35
40


ATLAS
EXPERIMENT
Run Number: 201289, Event Number: 24151616
Date: 2012-04-15 16:52:58 CEST

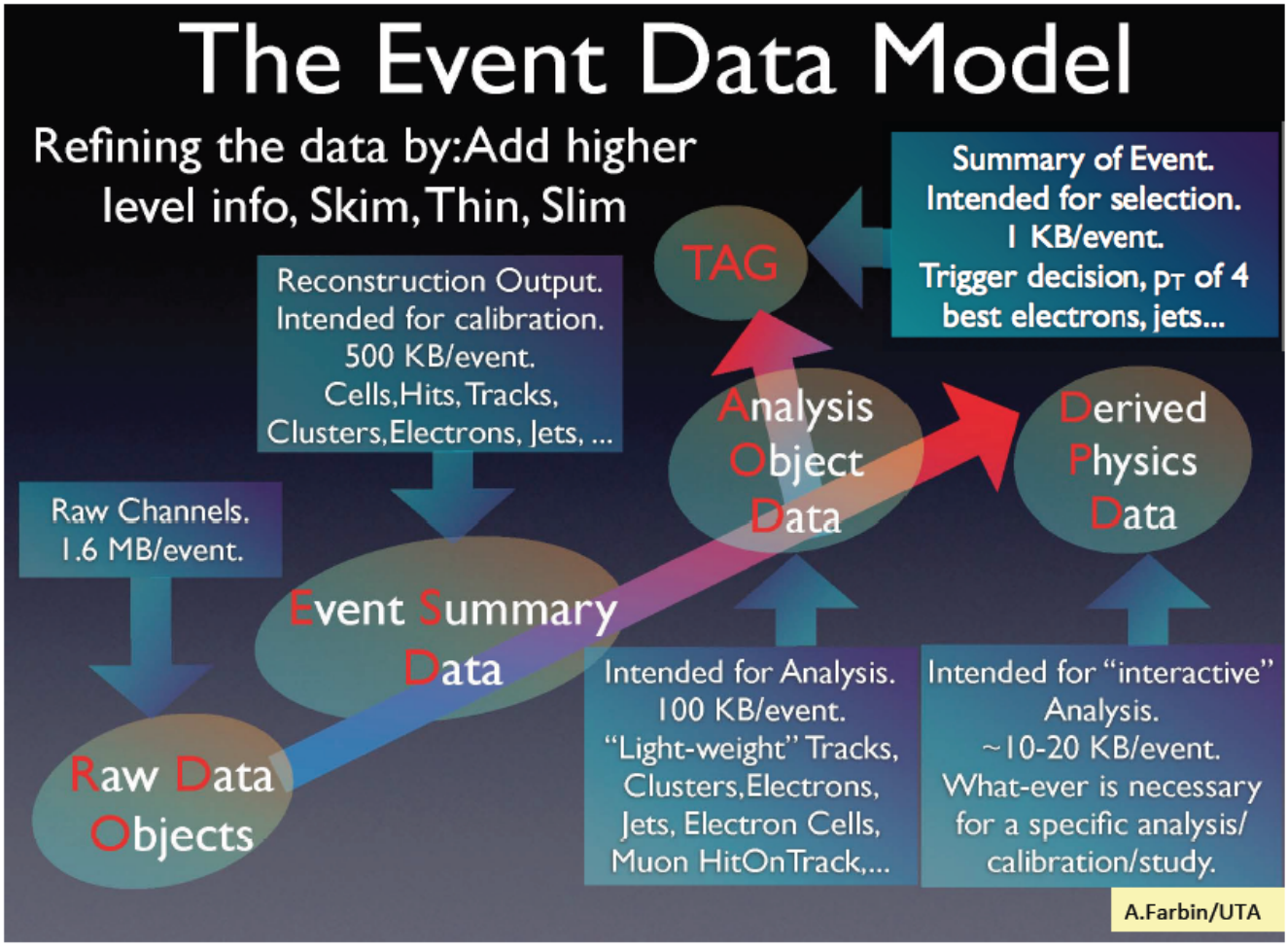
Summary of *pp* collisions results



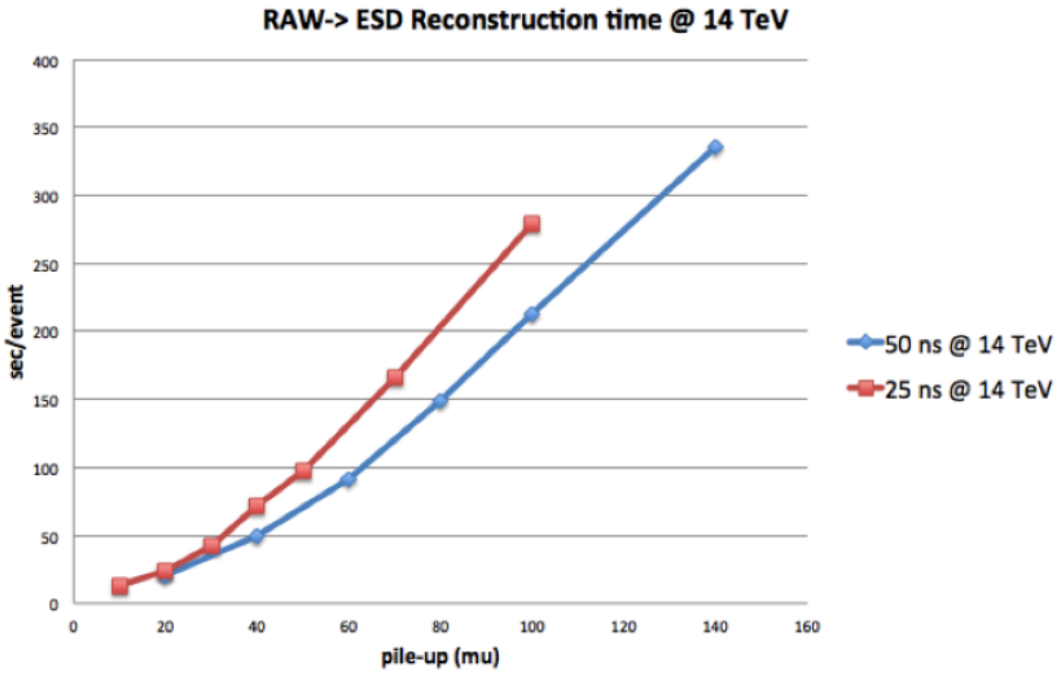
The ATLAS sw framework: ATHENA (in 2010)

Project	C/C++ code	C/C++ headers	PYTHON code	Total code
Core	390,000	43,000	240,000	860,000
Event	200,000	110,000	16,000	350,000
Conditions	280,000	90,000	21,000	620,000
Detector	38,000	6,100	8,400	140,000
Sum	910,000	250,000	280,000	2,000,000

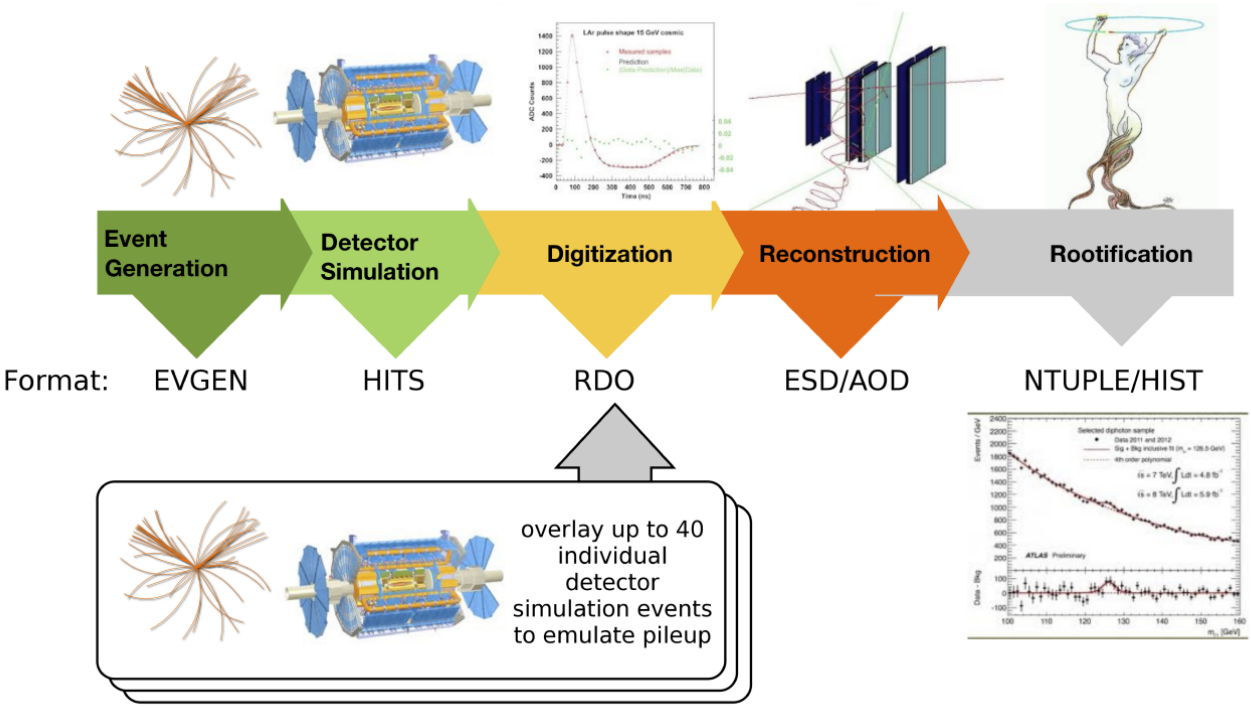
http://download.springer.com/static/pdf/255/art%253A10.1140%252Fepjc%252Fs10052-010-1429-9.pdf?auth66=1386170000_b50764175b665a0849a8fff546da53b3&ext=.pdf



Estimated reconstruction time in 2015

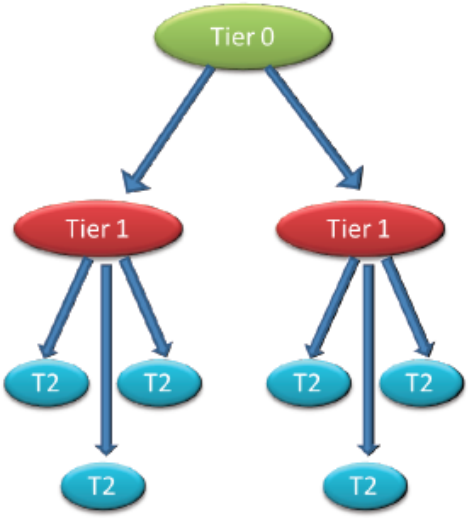


Monte Carlo simulation chain



Grid Data Processing

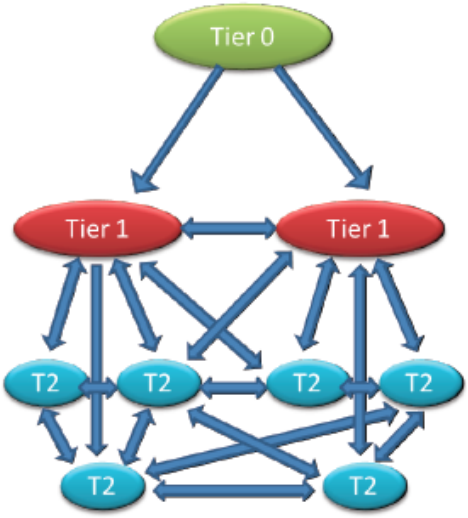
- ATLAS Grid Data Processing (GDP) uses Grids with three different interfaces split in ten "clouds" organized as large computing centres with tape data storage (Tier-1 sites) each associated with 5-6 other computing centres (Tier-2 sites)
  - ATLAS clouds evolve from the hierarchy (left) to the mesh (right)



15%: Tier-0 center at CERN
- Data recording
- First-pass data reconstruction
- Data distribution

40%: Tier-1 centers
- Permanent storage
- Reprocessing
- Group data processing

45%: Tier-2 centers
- Simulation
- End-user analysis

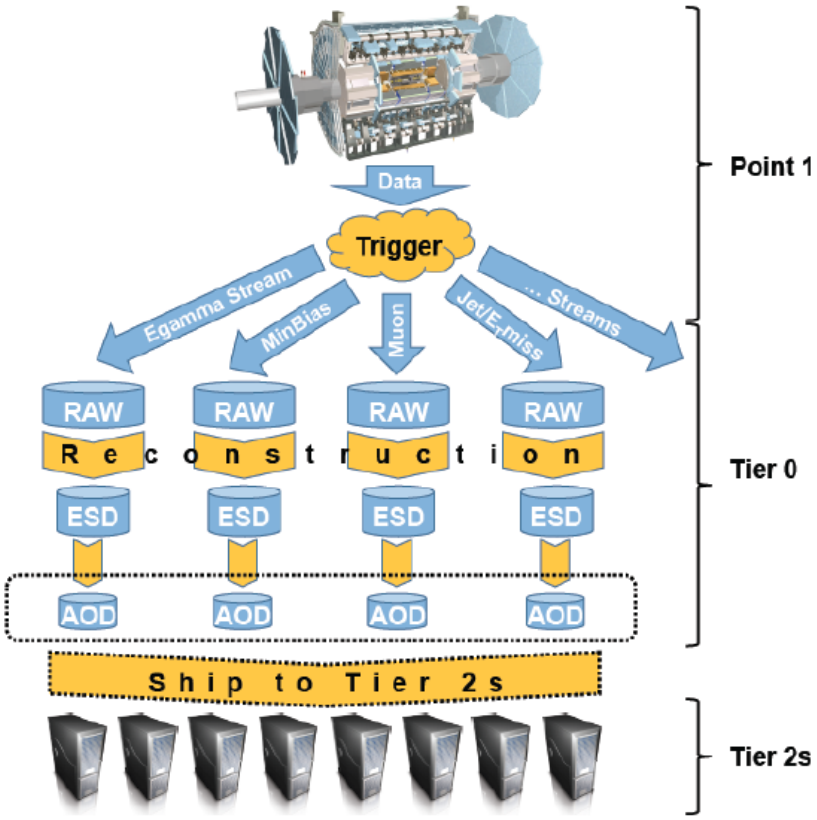


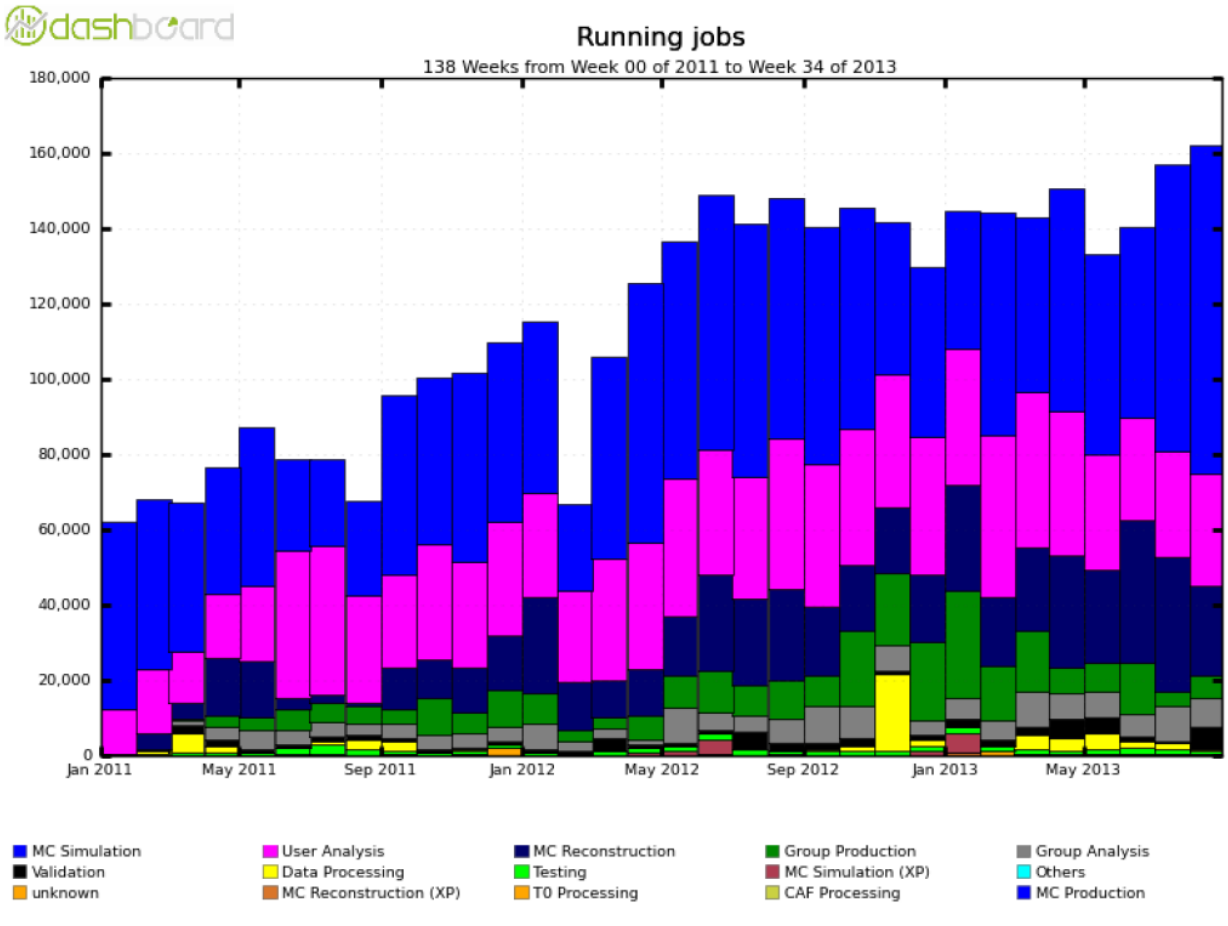
- GDP system empowered further data processing steps performed by dozens of ATLAS physics groups with coordinated access to computing resources worldwide

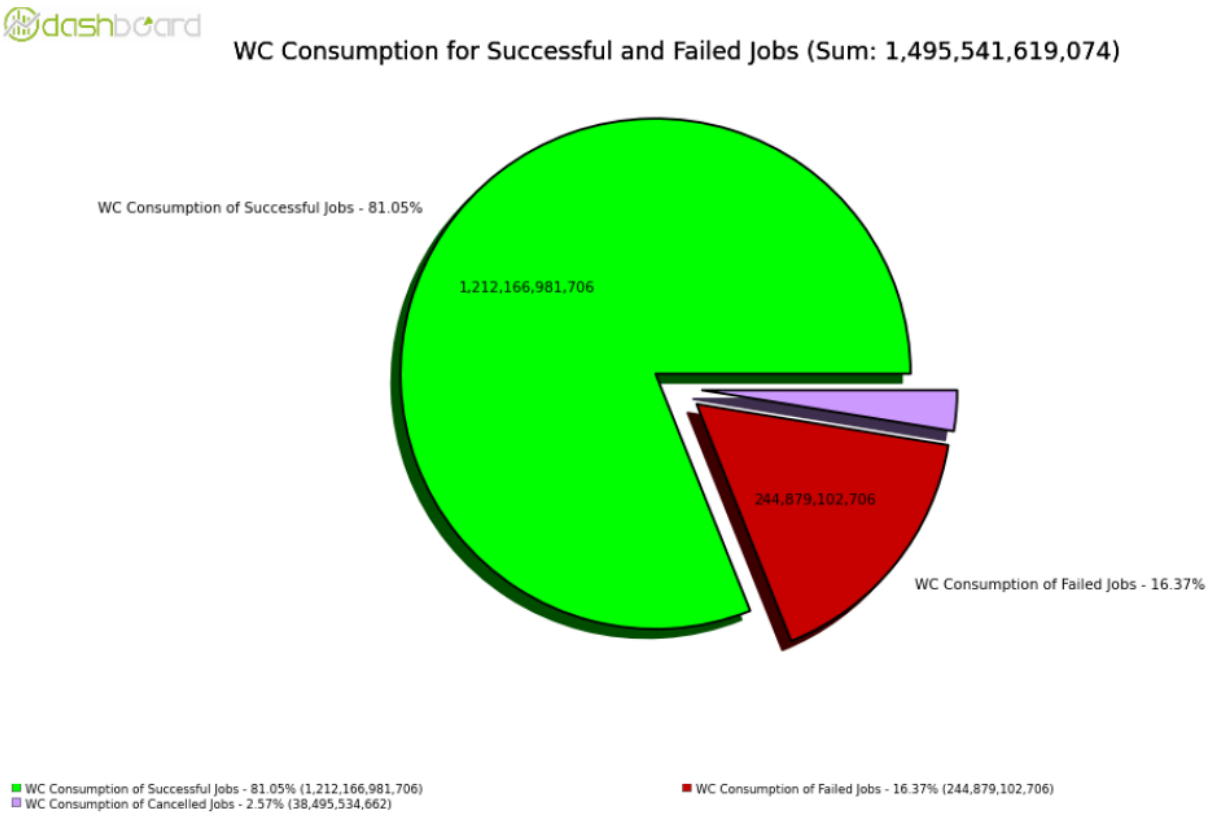
Data placement – original

Available to all users:

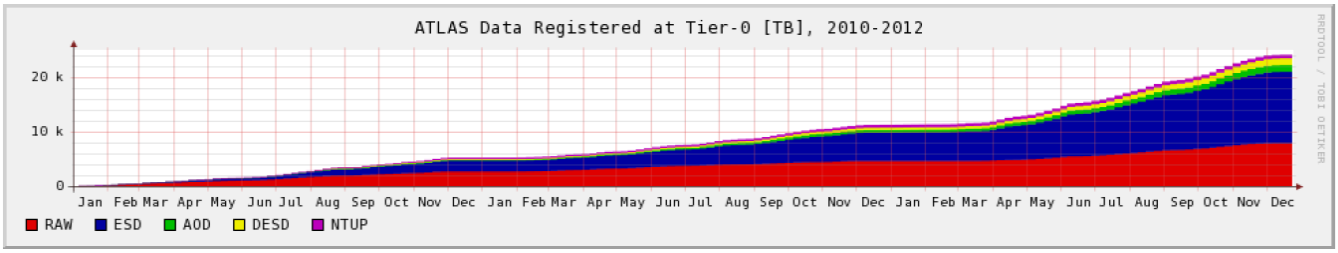
- Only data that are on disks at the tier-2s is readily available for user analysis.
- Only the AOD (and the TAG) is shipped to the tier-2s.
- Very detailed information that resides only in the ESD is NOT accessible to users.
- The AOD is still very large for frequent analyzes by users.

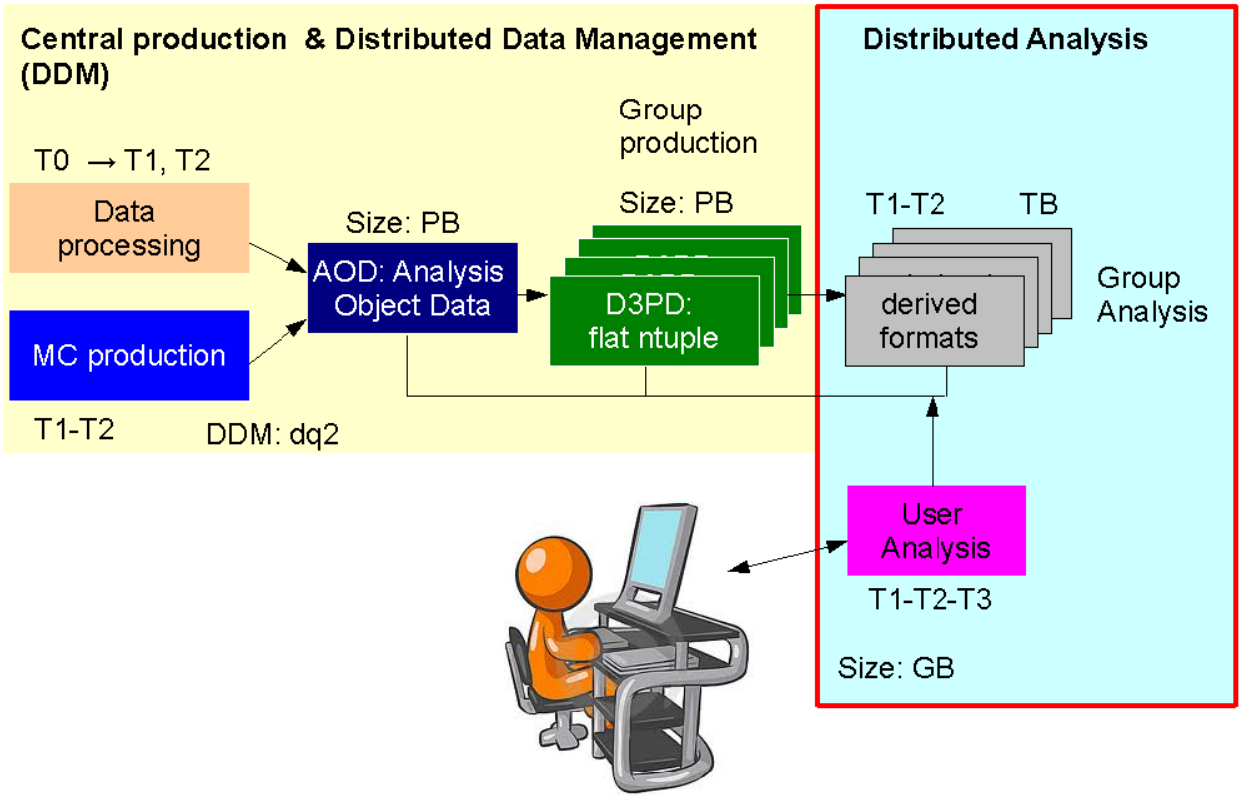




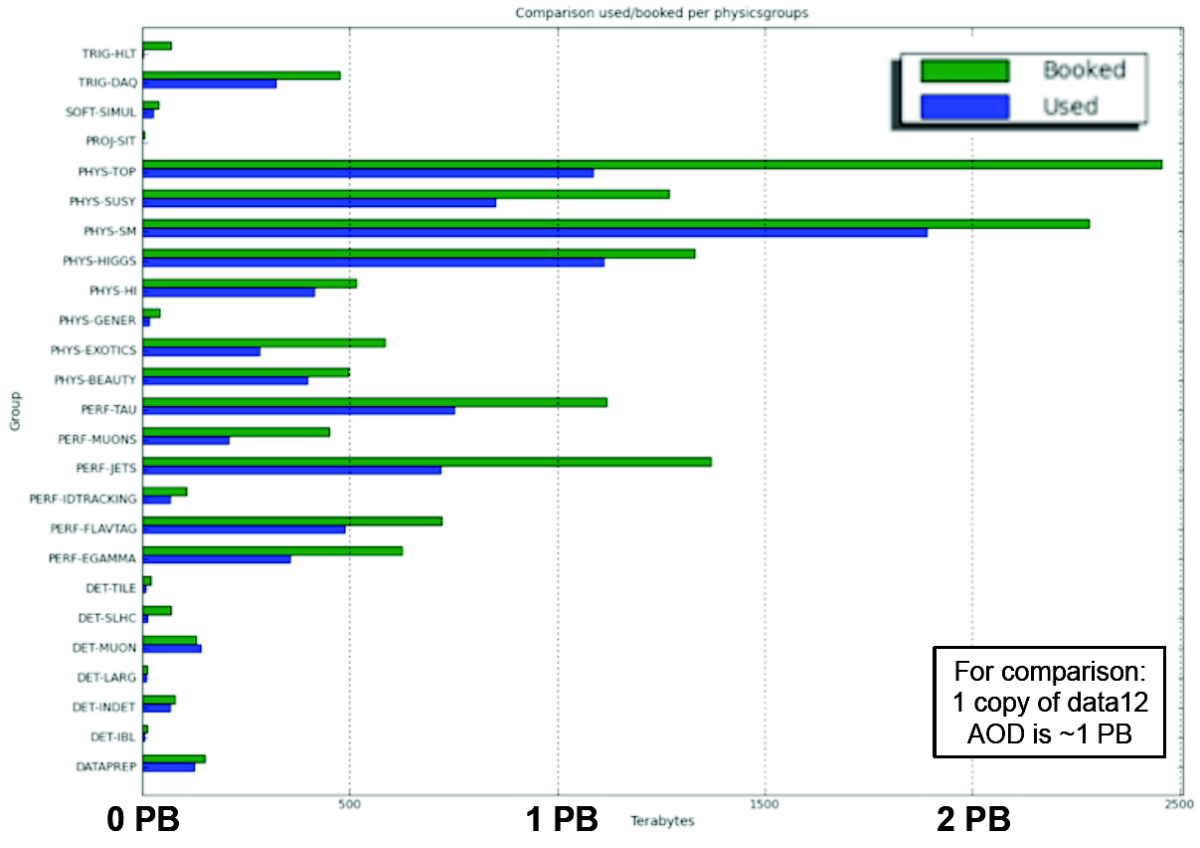


Cumulative amount of data processed by the ATLAS Tier-0





Group ntuples

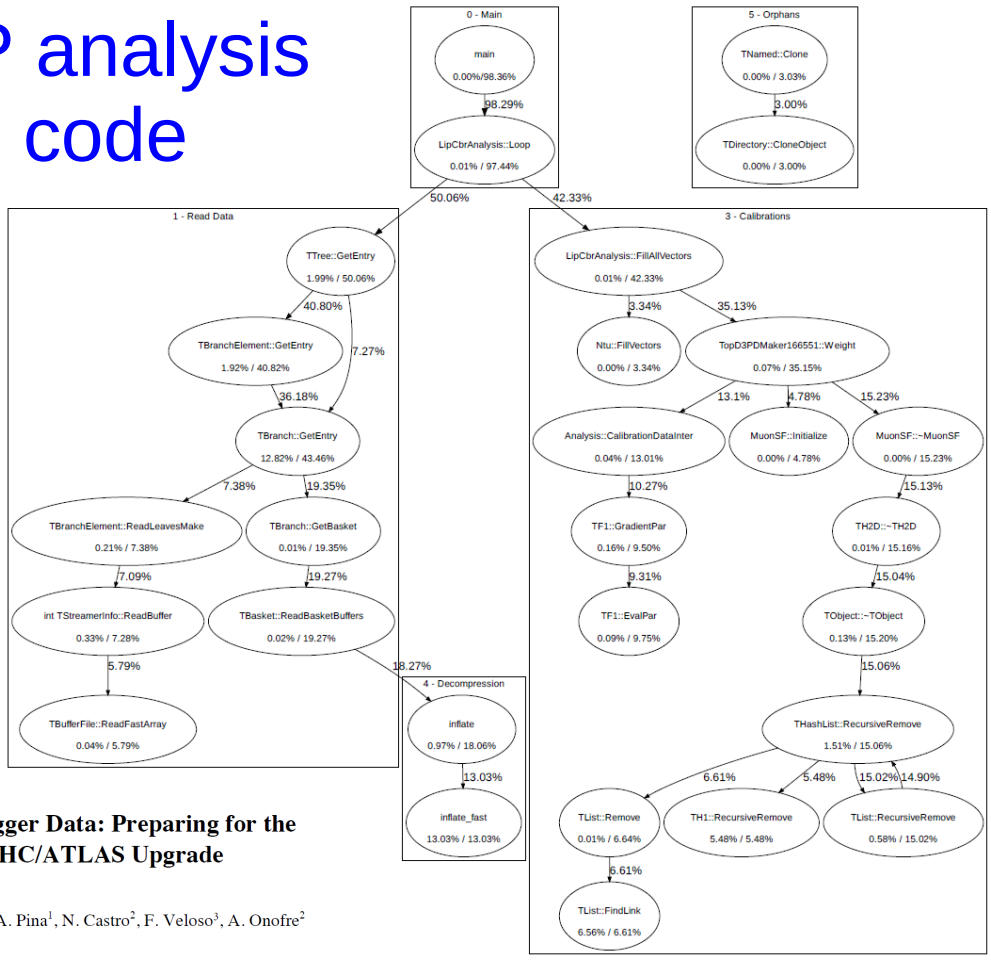


What do we do with the ntuples?

data
(either real data or MC)
analysis

LIP analysis code

LIP analysis code



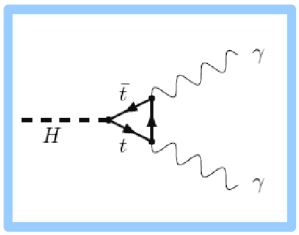
Even Bigger Data: Preparing for the LHC/ATLAS Upgrade

V. Oliveira[1], A. Pina[1], N. Castro[2], F. Veloso[3], A. Onofre[2]

What do we get in the end?

- many (small) txt files

- many root files with histograms (small compared to the input ntuples)



$H \rightarrow \gamma\gamma$

Outstanding Challenges

- **Analysis Code Optimization**
  - Use current analysis code, identify possible inefficiencies (way data is read, identify redundant operations, identify obsolete code, etc.) and correct them, without major changes in the underlying framework
  - We need to gain a significant factor in speed in order to remain competitive with increasing flow of data

Going Even Further:

- **I/O Optimization**
  - Is it possible to have a more efficient I/O handling within our framework (we read O(100 Gb) of data/per analysis)?
  - We crucially depend on ROOT package (not trivial to optimize!) http://root.cern.ch/drupal/
- **Local Resources Optimization:**
  - Parallel ROOT Facility (PROOF)
  - ...