Physics Validation of the LHC Software

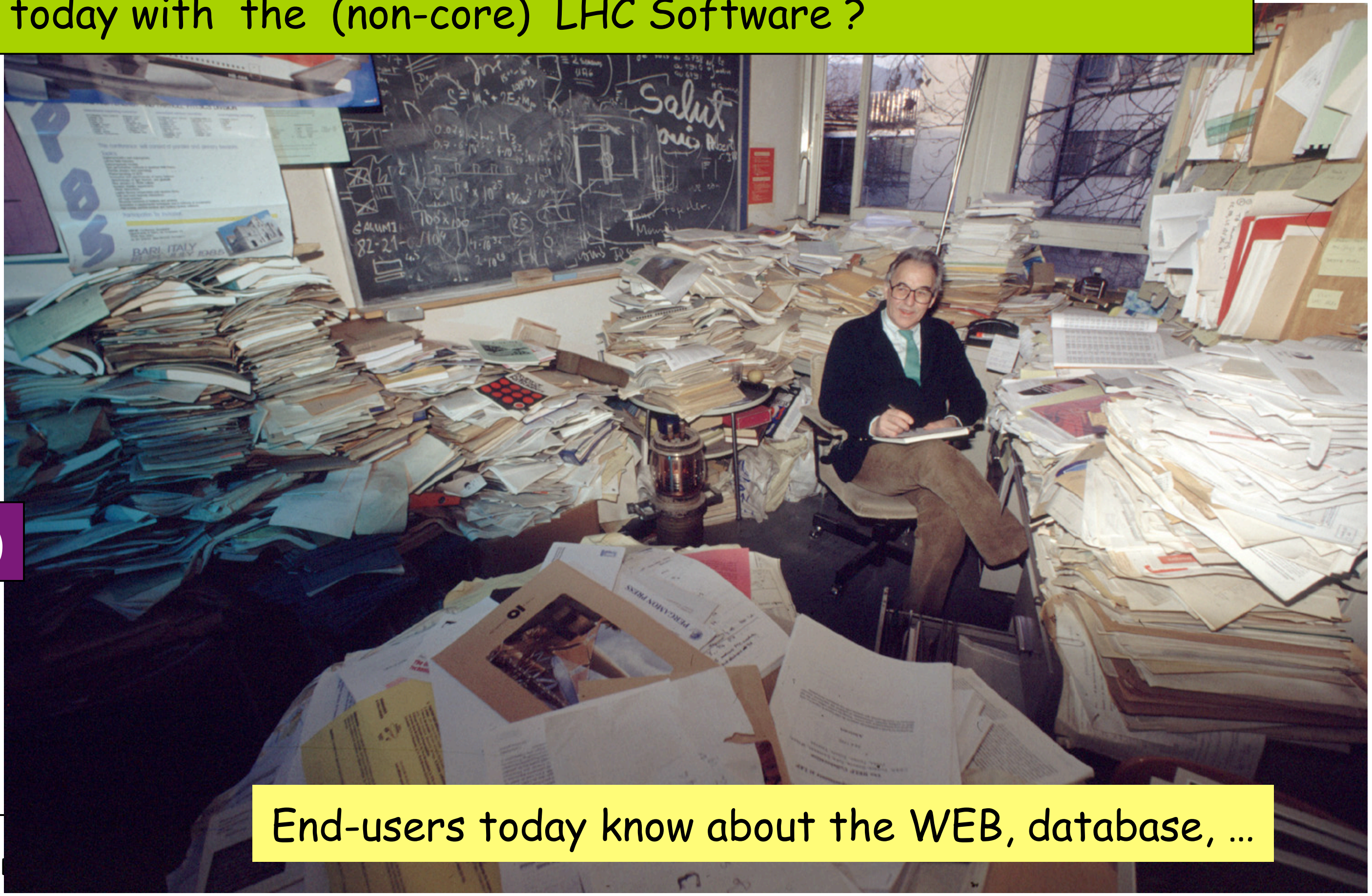
❶ Software requirements and challenges from LHC environment, detectors/triggers and physics (physics aspects only, technical aspects not covered here)

❷ Examples of software validation and performance from simulation, reconstruction, analysis

❸ Where do we stand today with the (non-core) LHC Software ?

The point of view of a physicist and end-user

Fabiola Gianotti (CERN)



End-users today know about the WEB, database, …

❶

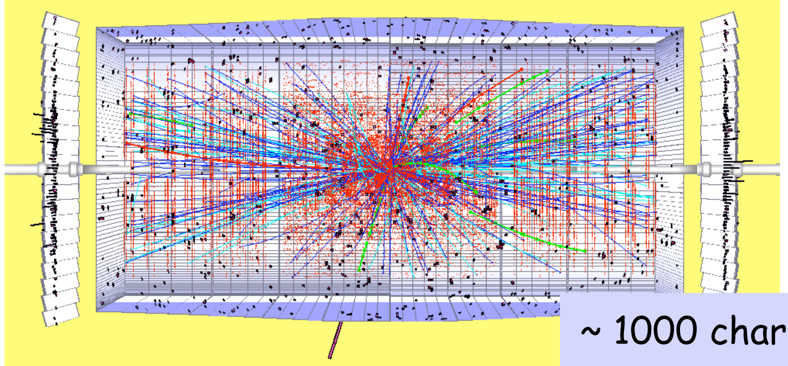
Main requirements and challenges of LHC software

Compared to previous machines : -- much more difficult environment
-- much more demanding triggers
-- much better detector performance
-- much more ambitious and broad physics goals

- Unprecedented particle energy range : ~ 0.1 GeV (ALICE) → few TeV (ATLAS, CMS)
  → detector simulation, reconstruction, …
- Unprecedented particle multiplicities :
  -- pile-up at 10^{34} cm^{-2} s^{-1} → ~20 pp collisions / bunch x-ing (every 25 ns) in ATLAS/CMS
  -- high-E heavy-ion collisions → ~ 10000 charged particles per event in ALICE TPC

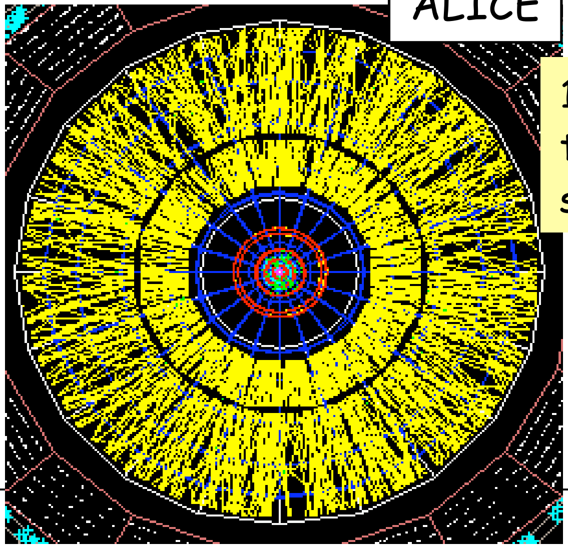
→ pile-up simulation, pattern recognition, ….

H → eeμμ event in CMS at 10^{34} $cm^{-2}s^{-1}$



~ 1000 charged particles

ALICE



1% of all tracks shown here

- Unprecedented triggers → data access, fast reconstruction, …

	ATLAS, CMS pp, $L=10^{34}$	LHCb pp, $L=2\times10^{32}$	ALICE central PbPb, $L=10^{27}$
Interaction rate	10^9 Hz	10^7 Hz	8 kHz
Input rate to HLT	~ 100 kHz	10^6 Hz	< 1 kHz
Rate to storage	100-200 Hz	~ 200 Hz	~ 50 Hz
Event size	~ 1-2 MB	~ 100 kB	~ 25 MB

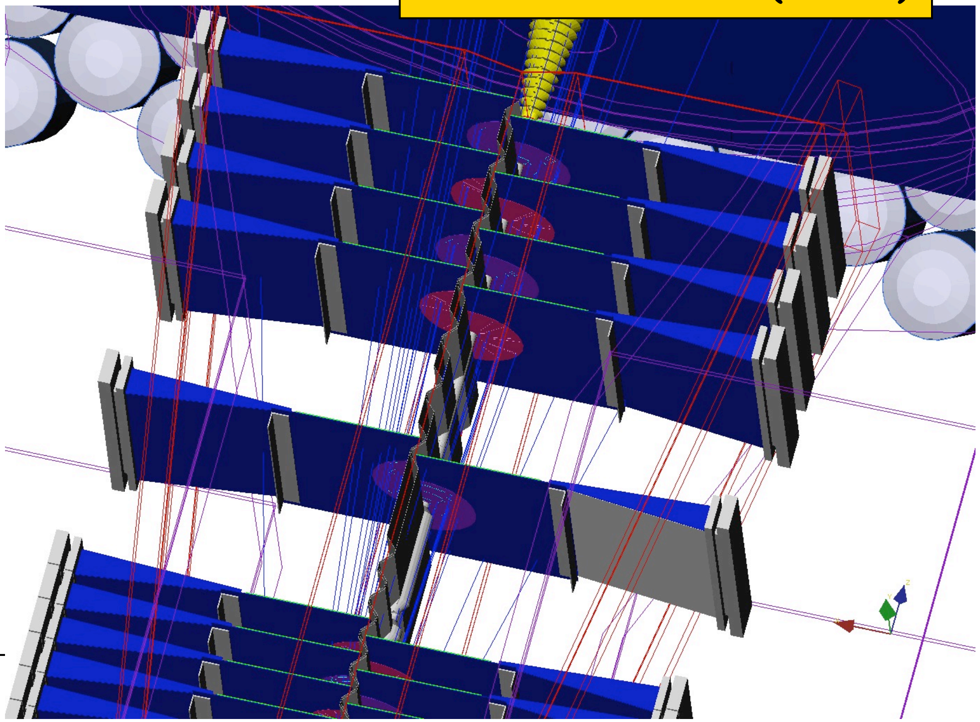
Software based

(latency: 1ms -1s)

- Unprecedented detectors:
  -- large variety of technologies
  -- number of channels : $>10^8$ ATLAS/CMS
  -- excellent performance (resolutions, measurement accuracies, particles identification) : 0.1%-1%

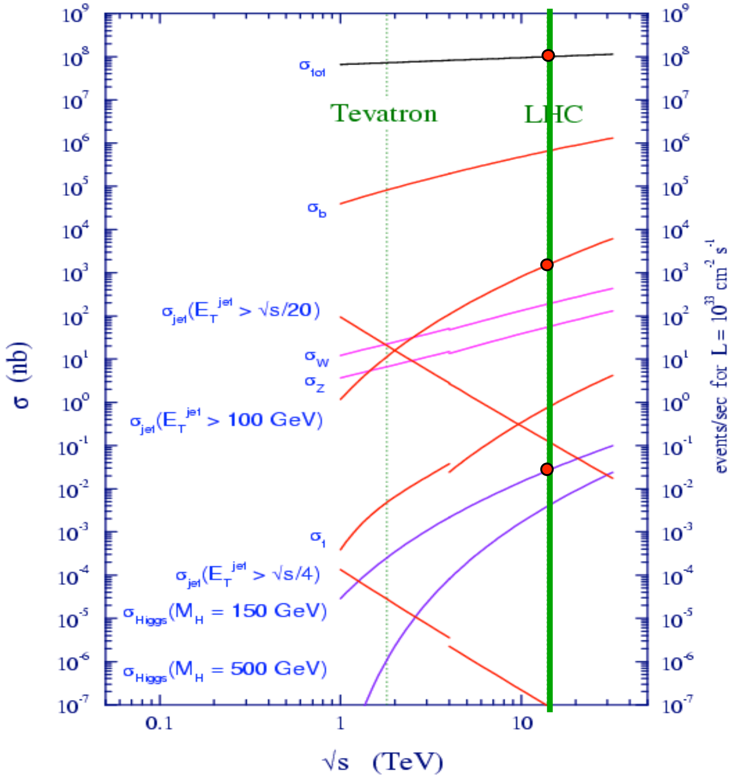
→ simulation, detector description, calibration, reconstruction, …

Simulation of the LHCb Vertex Detector (VELO)

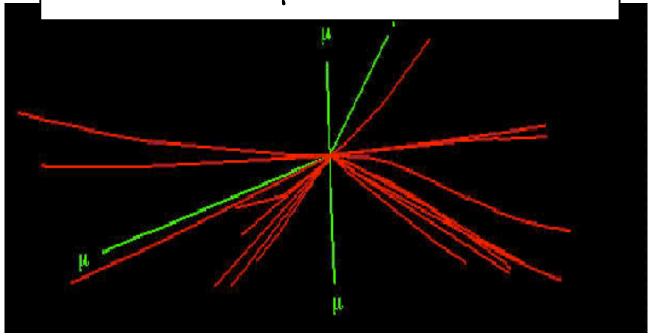


- Unprecedented physics goals :
  - -- precise measurements (e.g. m_W, m_{top}, B-decays) with higher accuracies than before
  - -- extract tiny new signals from huge backgrounds (e.g. 1 H→ 4l event every ~ 10^{13} pp collisions)

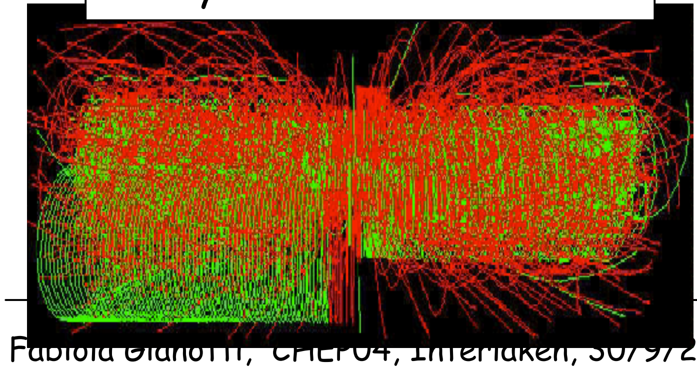
    Note : S/B ratios typically ≥ 100 worse at LHC than at the Tevatron for channels accessible to both
  - -- explore the "unknown" up to the multi-TeV scale through huge number of topologies



One H → 4μ event like this



Every 10^{12} events like that



⇒ Need precise/robust/redundant understanding of detector performance and physics (e.g. backgrounds to New Physics)

⇒ -- many Monte Carlo generators

-- several levels of detector simulations full (Geant4, FLUKA), parametrized, fast

-- many reconstruction algorithms

→ software (framework !) modularity and flexibility

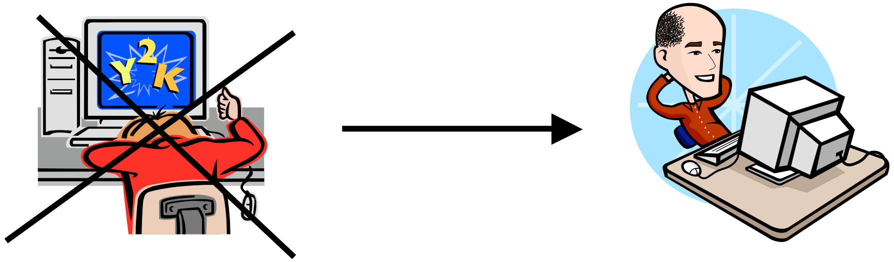
2 main "physics" requirements for the LHC software :

❶ Cope with these unprecedented conditions and challenges, i.e.:

do not become the limiting factor to trigger and data taking, detector performance and physics reach

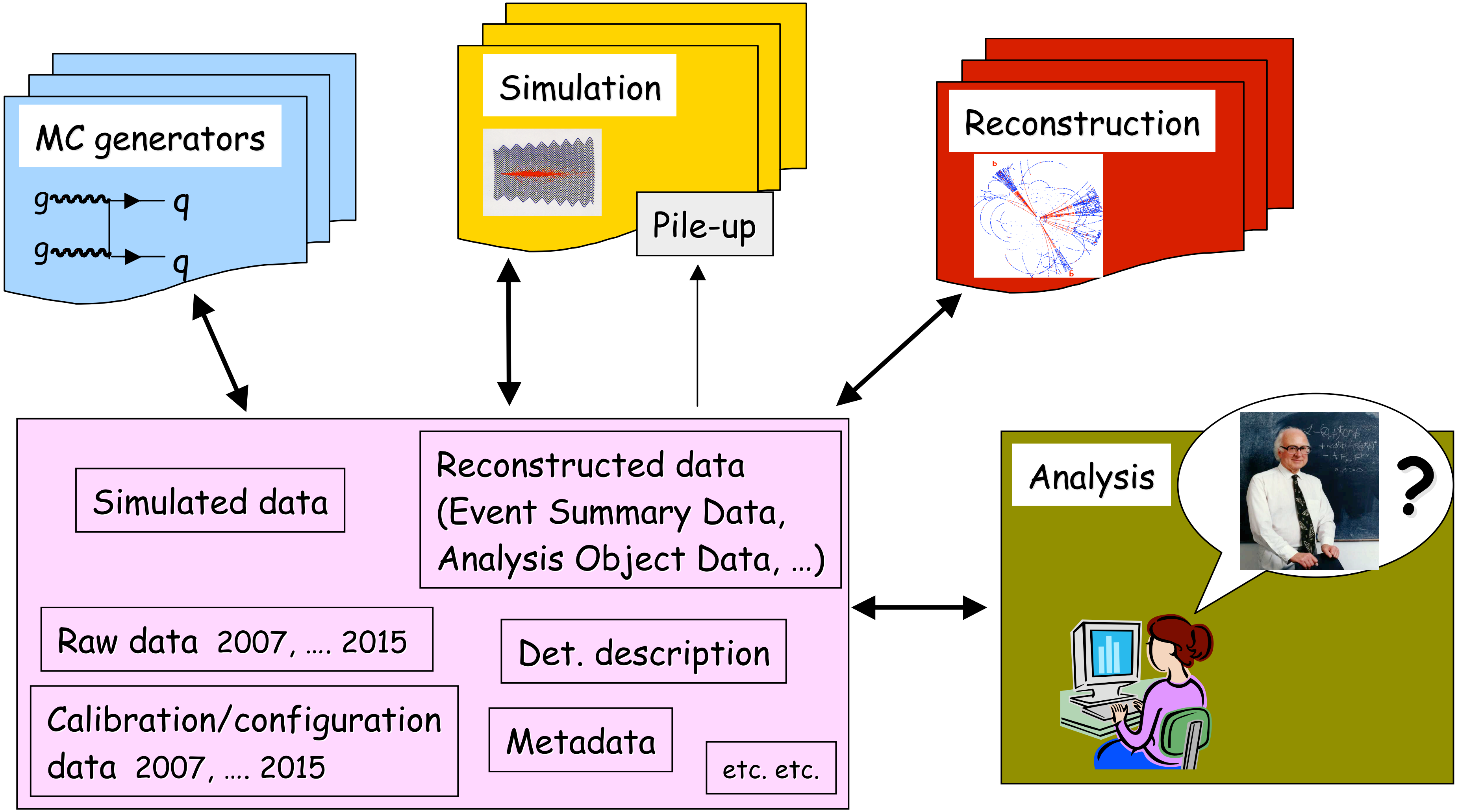
❷ In spite of complexity, be easy-to-use

Each one of the ~ 4000 LHC physicists (including people from remote/isolated countries, physicists who have built the detectors, software-old-fashioned senior physicists) should be able to run the software, modify part of it (reconstruction, …), analyze the data, extract physics results



Users want:
Simplicity (simple interfaces)
Stability
Interactivity

❷ Main component of LHC software: a simplified physics-oriented point of view



MC generators

Simulation

Pile-up

Reconstruction

Simulated data

Reconstructed data
(Event Summary Data,
Analysis Object Data, ...)

Raw data 2007, ... 2015

Det. description

Calibration/configuration
data 2007, ... 2015

Metadata

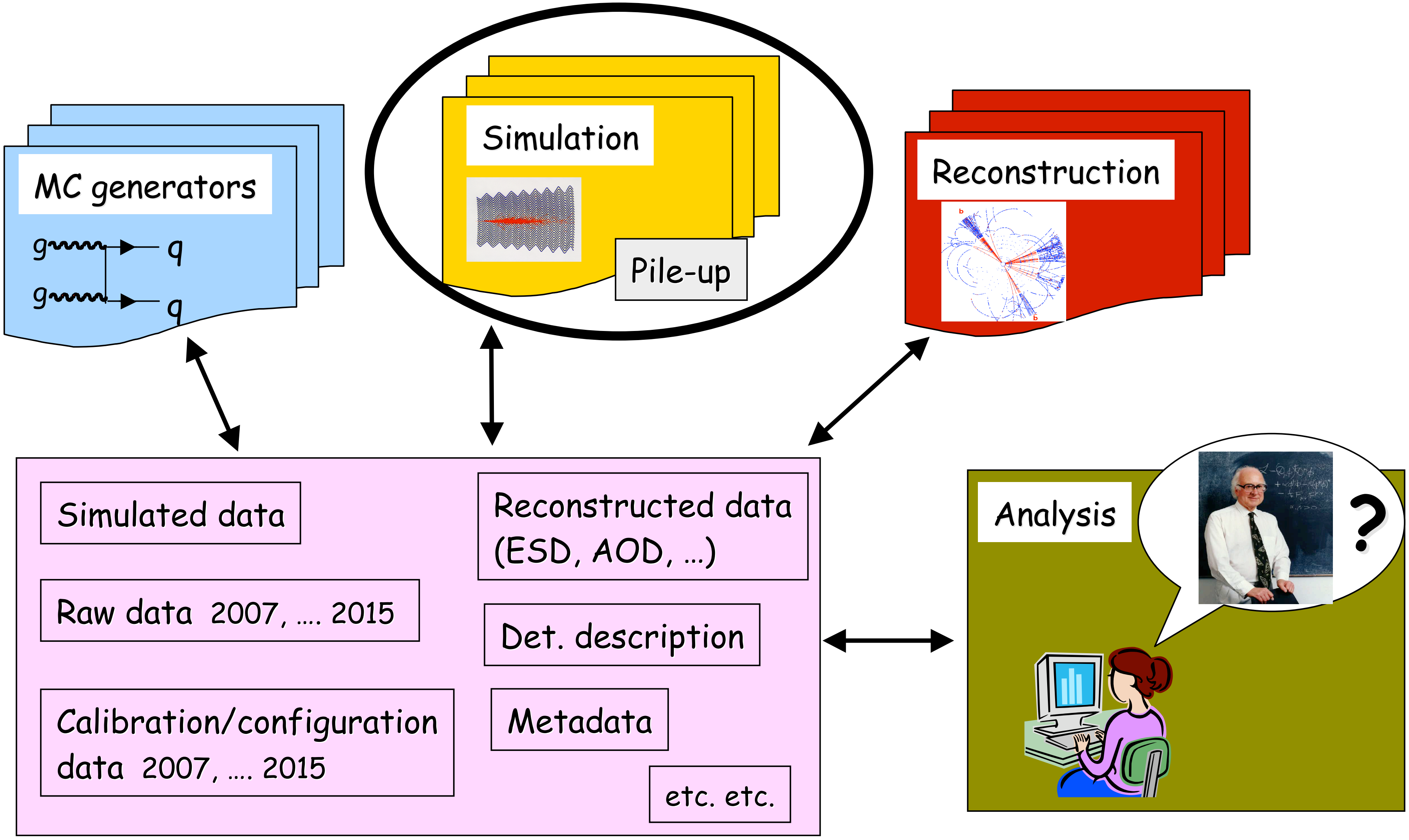
etc. etc.

Analysis

Here : a few examples from simulation, reconstruction, analysis, ...

Note : lot of experiment-common LCG software
→ good also for physics (robustness/reliability, easier cross-checks among experiments, etc.)

SIMULATION



MC generators

Simulation

Pile-up

Reconstruction

Simulated data

Raw data 2007, …. 2015

Calibration/configuration data 2007, …. 2015

Reconstructed data (ESD, AOD, …)

Det. description

Metadata

etc. etc.

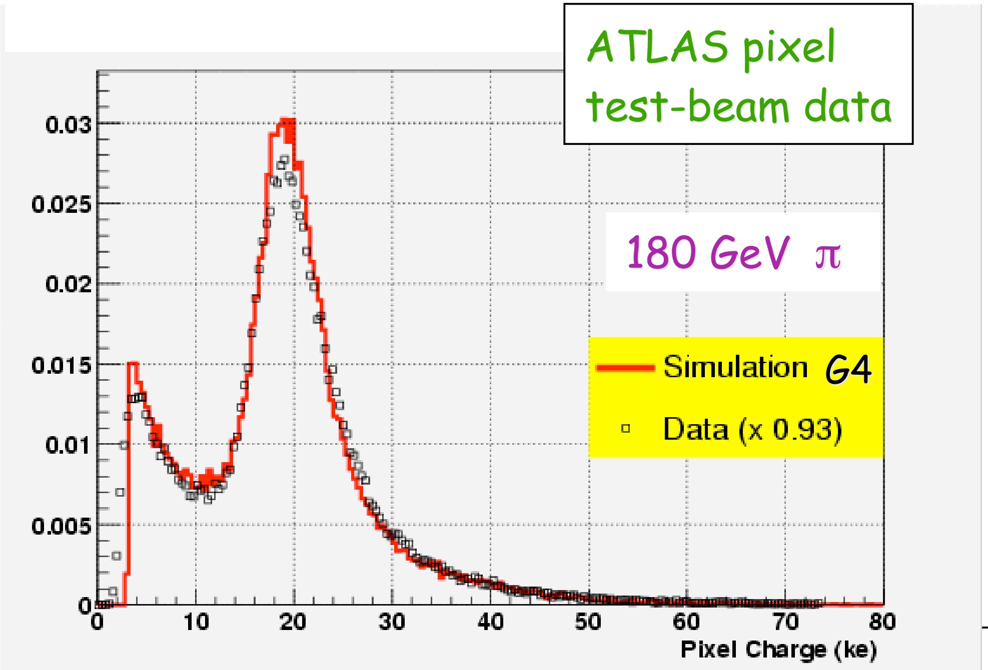
Analysis

Trackers : thin detector layers (e.g. 300 μm Si sensors) → need to model individual microscopic collisions down to ~ 10 eV/gas, ~ keV/Si for precise estimate of occupancy → of detector performance, aging, efficiency of pattern recognition, ...

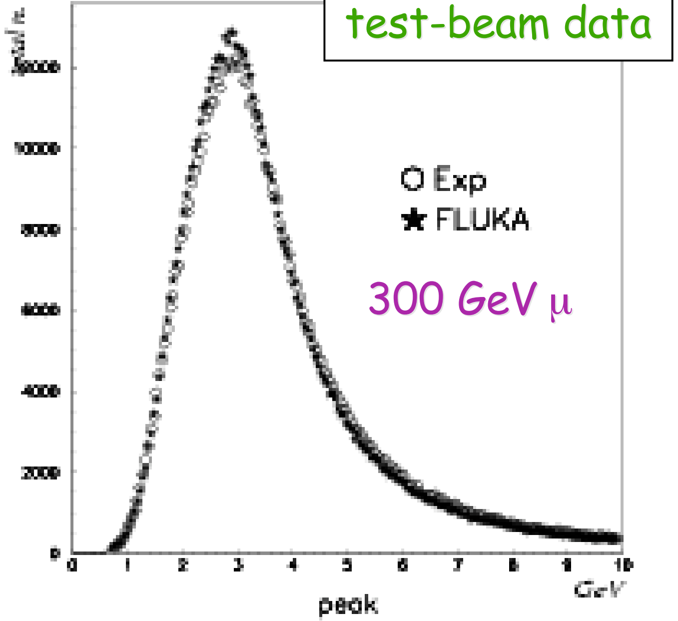
Muon Spectrometers: need to describe background hits from
- -- high-E μ : catastrophic E-losses in upstream calorimeters and shower punch-through
- -- radiation background in the cavern: ~ 1 MeV neutrons, 300-500 keV γ

→ impact on trigger rates, detector performance and aging, pattern recognition, ..

Required precision : ~ % in most cases


ATLAS pixel test-beam data

180 GeV π


ATLAS Tilecal test-beam data

300 GeV μ

Calorimeters :

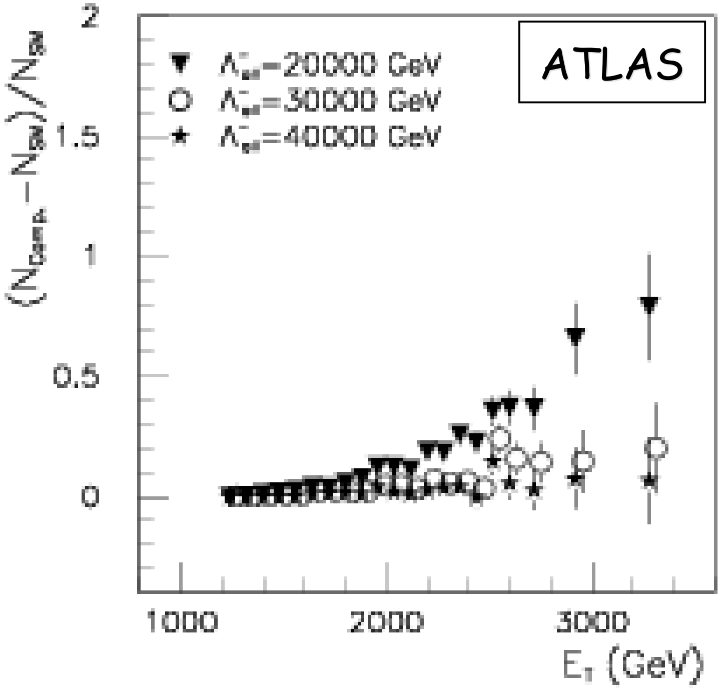
-- e/π/μ test-beam data available for E ~ 1-300 GeV

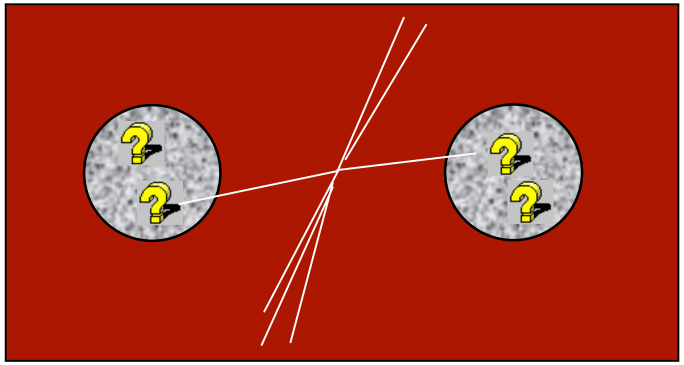
-- "calibration" samples at LHC, e.g. Z (→ ll) +jets, cover up to few hundreds GeV

Validate simulation over this range and use it to predict detector response at E ~ TeV (where New Physics is expected !)

Example :

Are quarks really point-like ?





If quarks are composite : new qq → qq interactions with strength ~$1/\Lambda^2$, $\Lambda \equiv$ scale of New Physics.

⇒ expect excess of high-p_T jets compared to SM

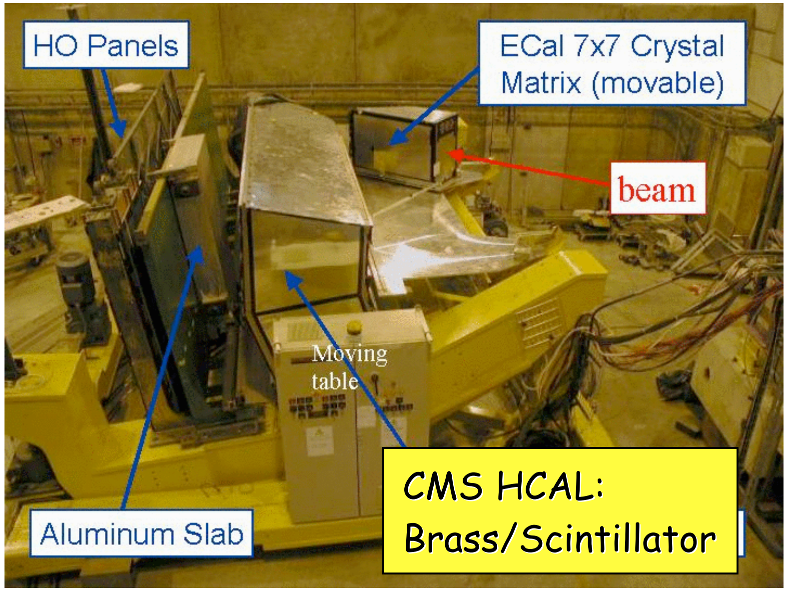
The higher Λ the smaller the excess.

LHC sensitivity up to $\Lambda \approx$ 40 TeV

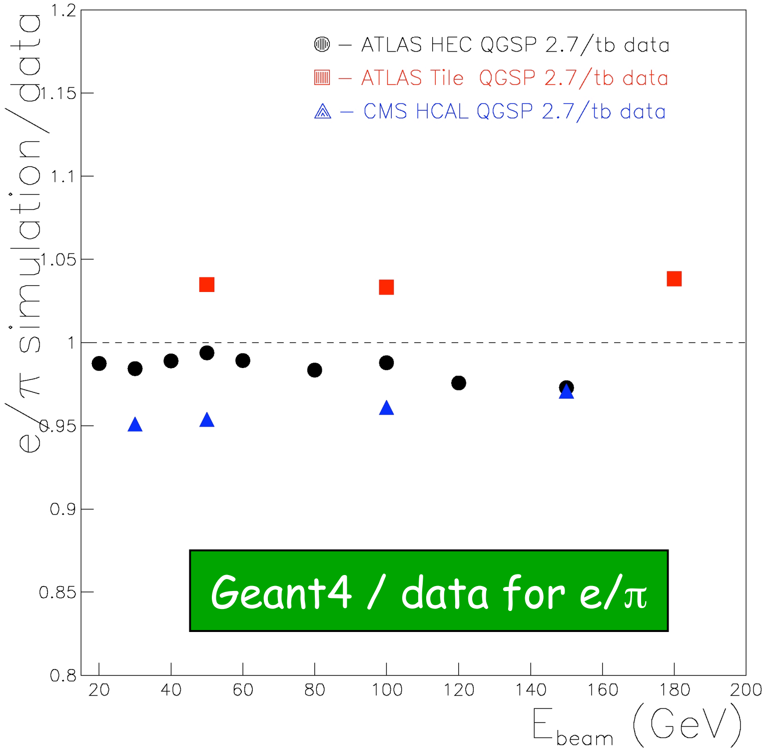
A hadron calorimeter non-linearity of 1.5 % at E_{jet} ~ 4 TeV, not reproduced by simulation, may fake a scale $\Lambda \approx$ 30 TeV ⇒ inadequacy of simulation would limit LHC physics reach

To avoid this : simulation must reproduce e/π response ratio (which governs response non-linearity to jets) to few percent

After extensive comparisons with test-beam data, iterations with the GEANT4 team, lot of efforts on experiment and simulation sides :

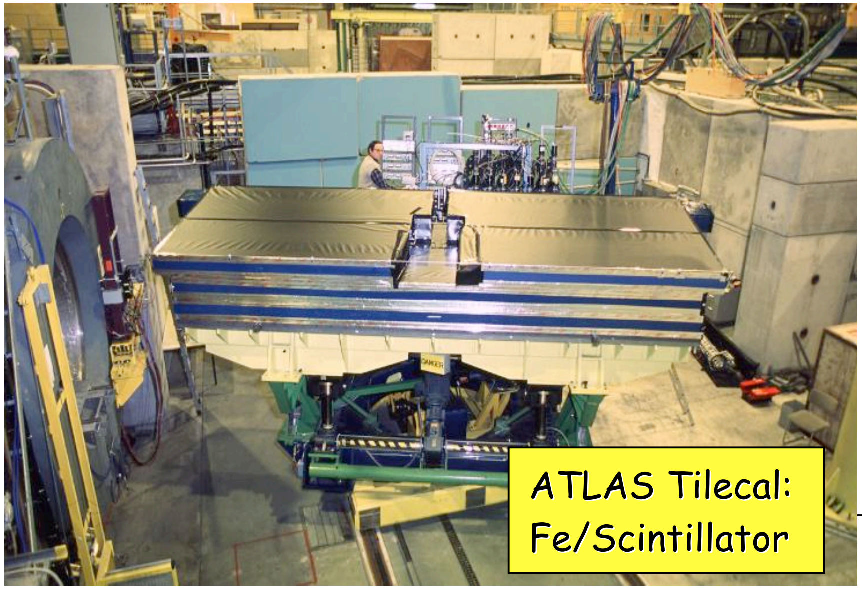


CMS HCAL: Brass/Scintillator

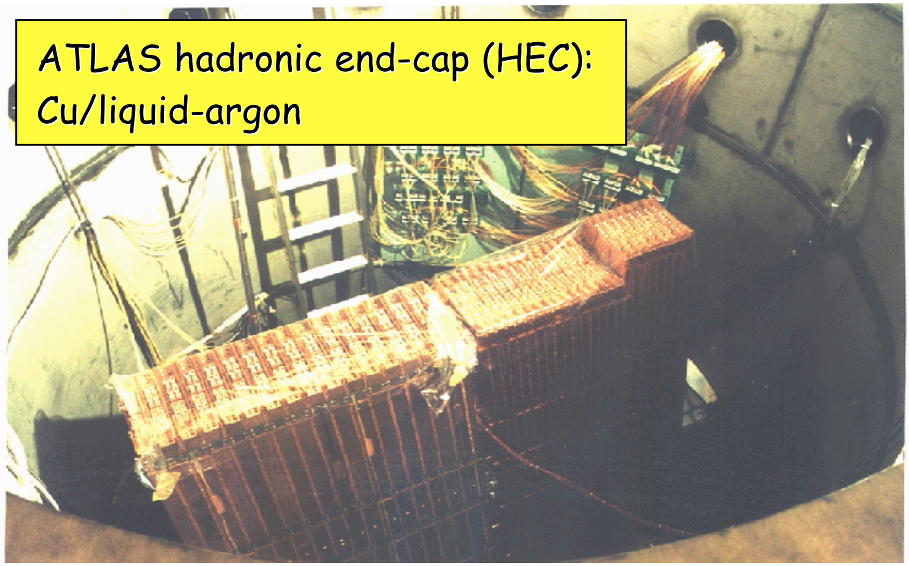


Geant4 / data for e/π

close to few % goal accuracy

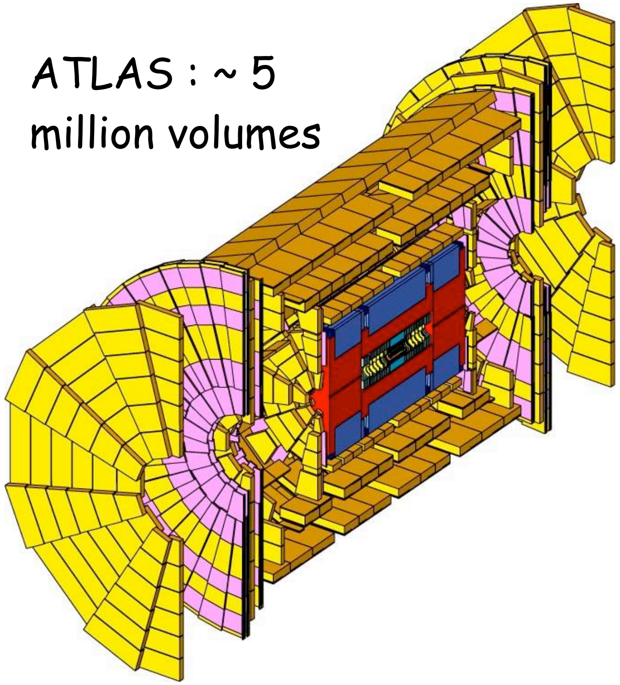


ATLAS Tilecal: Fe/Scintillator

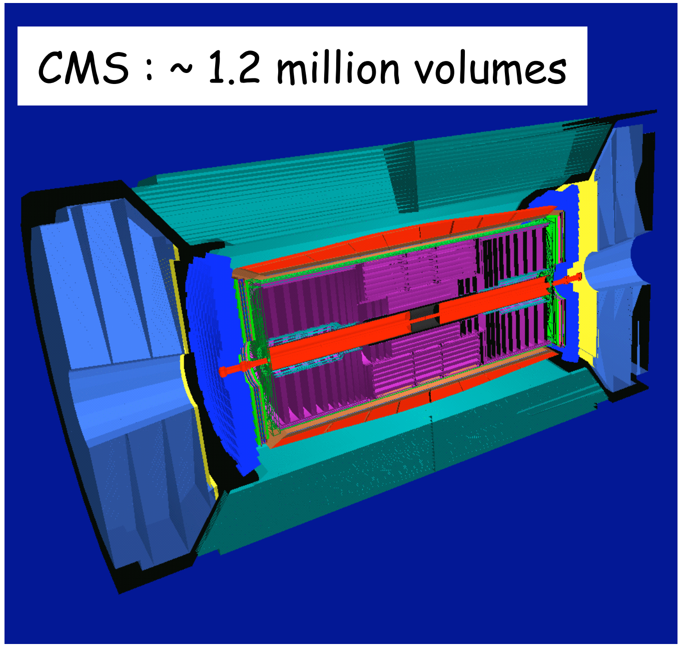


ATLAS hadronic end-cap (HEC): Cu/liquid-argon

ATLAS : ~ 5 million volumes

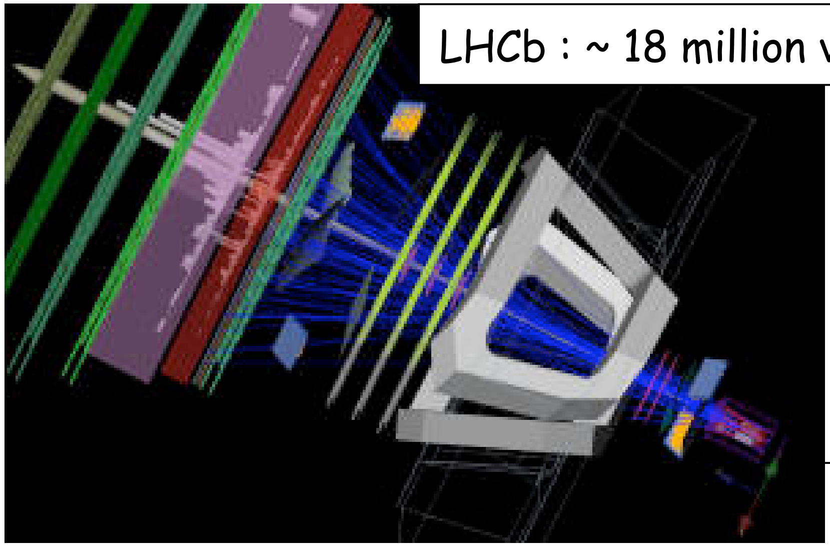


Huge numbers of physics processes, very low particle-tracking cuts, millions of volumes
→ robustness, CPU

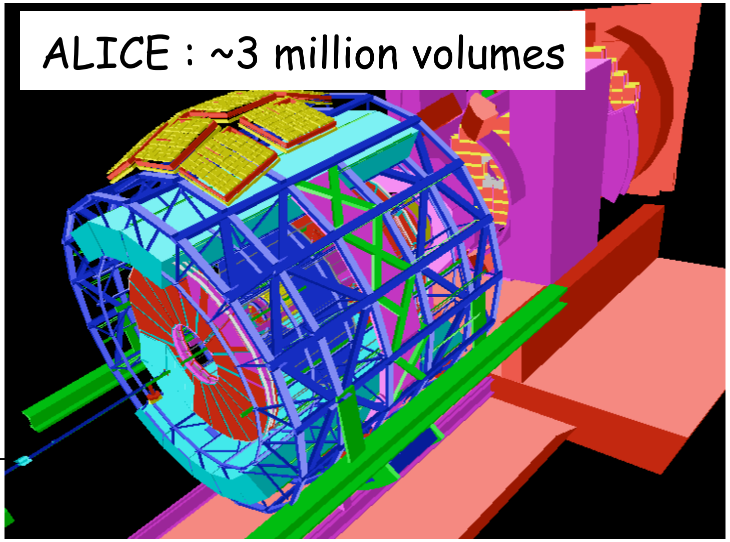


CMS : ~ 1.2 million volumes

Millions of events already fully-simulated in experiment Data Challenges
ALICE (G3) : 15 hours for one central event (1 GHz Pentium III)
ATLAS, CMS (G4) : ~ 20'-30' for one di-jet event with p_T (jet) ~ 1 TeV (1 GHz Pentium III)

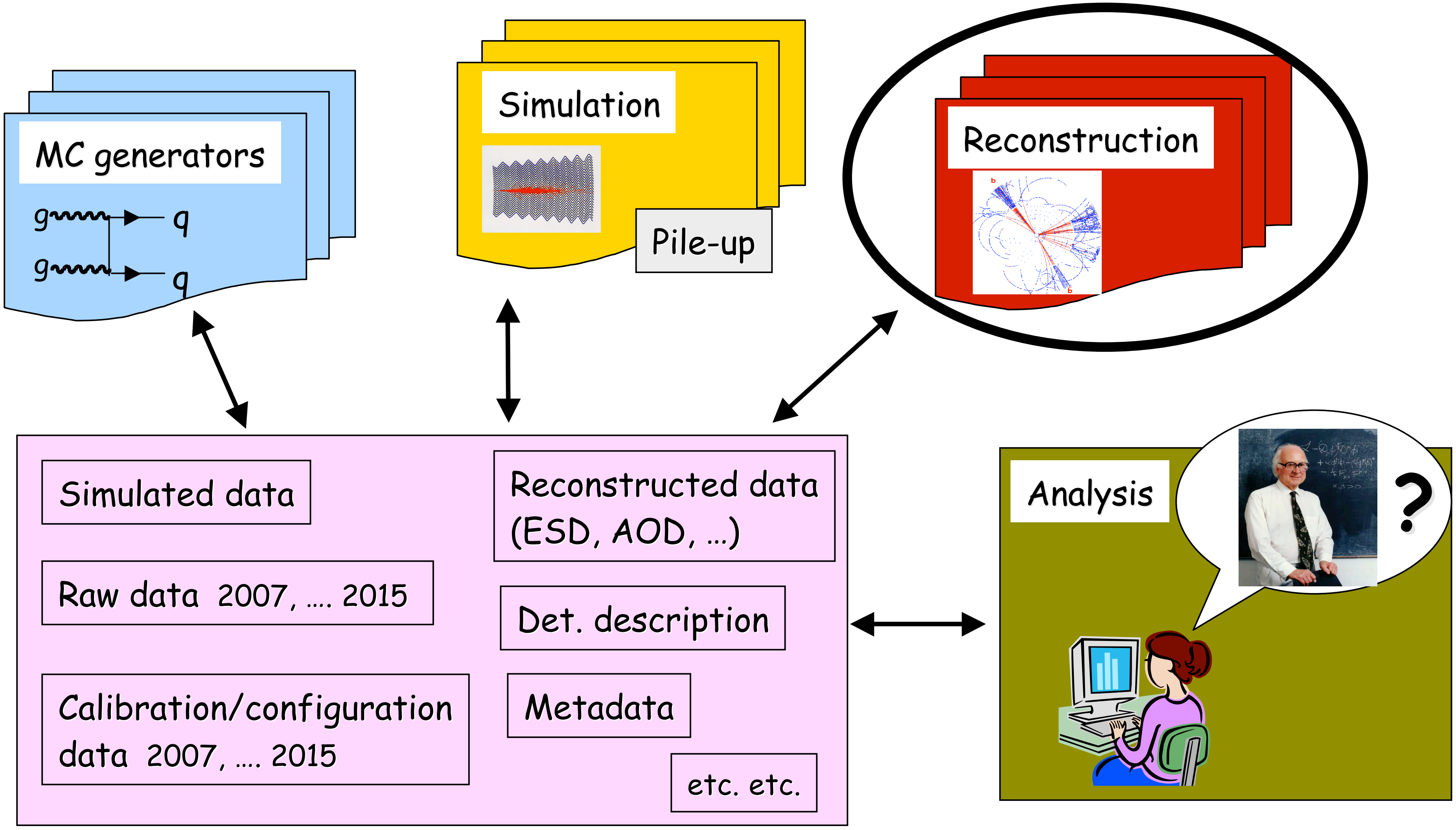


LHCb : ~ 18 million volumes



ALICE : ~3 million volumes

RECONSTRUCTION



MC generators

Simulation

Pile-up

Reconstruction

Simulated data

Raw data 2007, …. 2015

Calibration/configuration data 2007, …. 2015

Reconstructed data (ESD, AOD, …)

Det. description

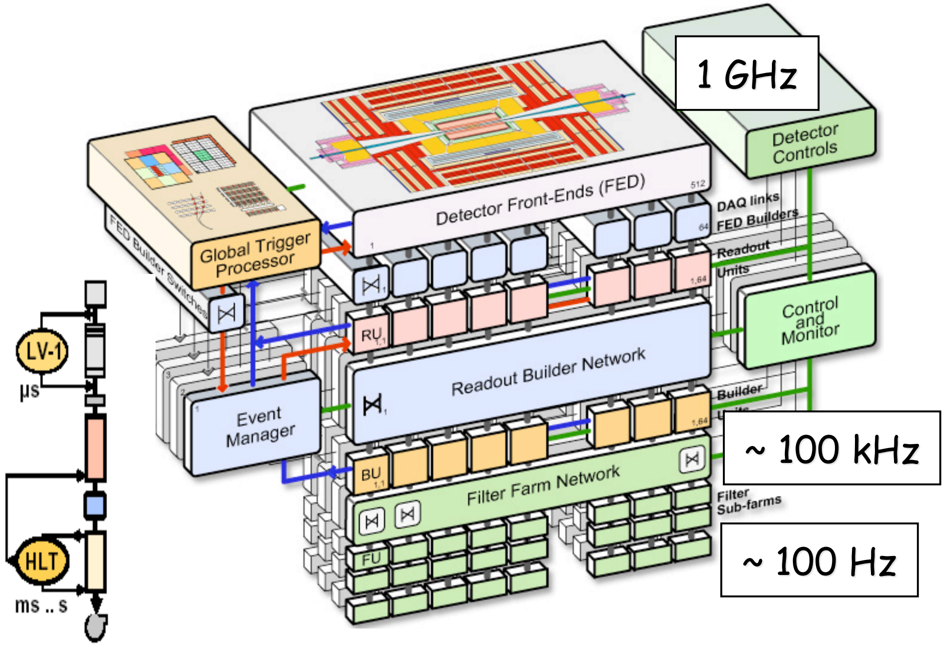
Metadata

etc. etc.

Analysis

Example 1 : CMS tracking at HLT (most demanding because no dedicated LVL2 in CMS)

- LVL1 (hardware) : 1 GHz → ~ 100 kHz
- HLT (software) : ~100 kHz → ~ 100 Hz



- HLT : farm of ~ O(10^3) CPU

→ on average ~ 40 ms/evt available

→ computing power : ~ 10^6 SI95

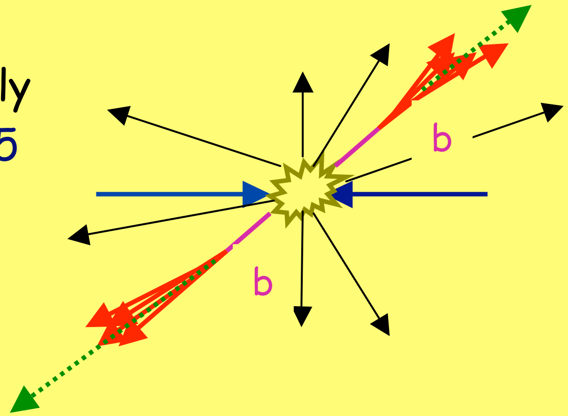
- Offline framework and code used

→ robustness, reliability

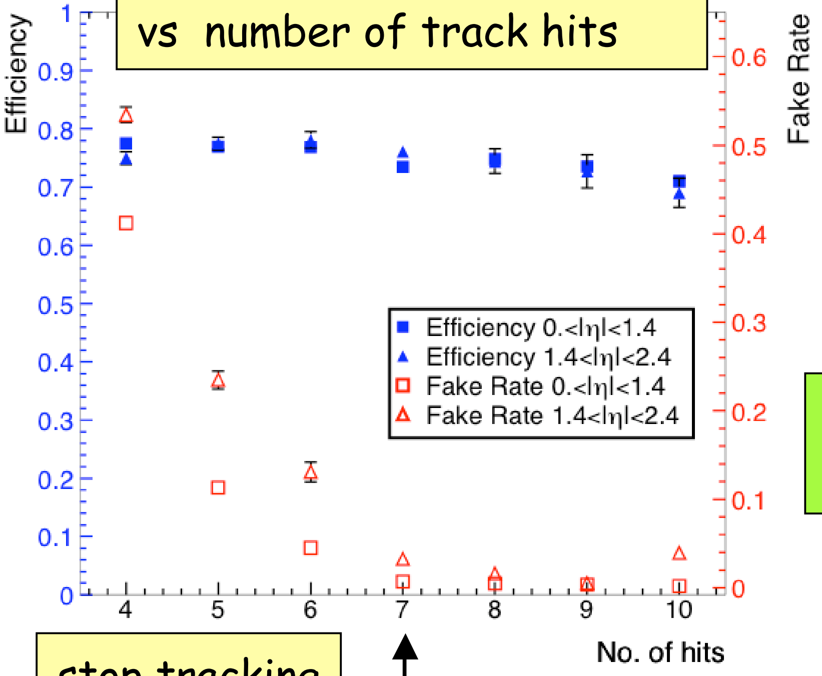
→ performance (CPU, memory) → data loading on demand, regional reconstruction, conditional reconstruction

Ex.: Select b-jets at HLT (3rd fermion family !) ?

- Fast (< 50 ms) determination of primary vertex from pixels only
- Regional reconstruction: tracks reconstructed inside a DR<0.25 cone around direction of LVL1 jet starting from pixel seeds
- Conditional reconstruction: tracking stopped when ~ 6-7 hits found on trajectory (don't need ultimate performance at HLT)



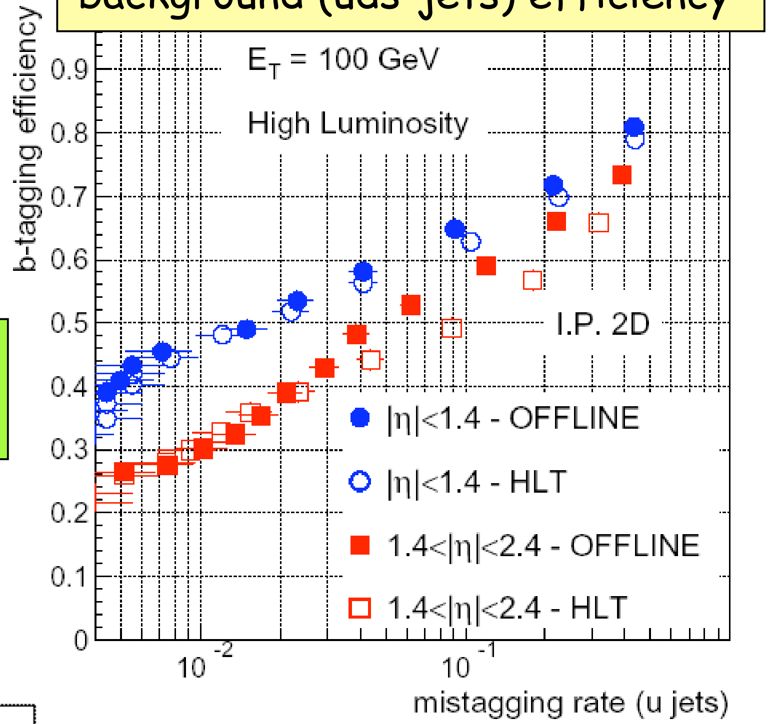
Track efficiency and fakes vs number of track hits



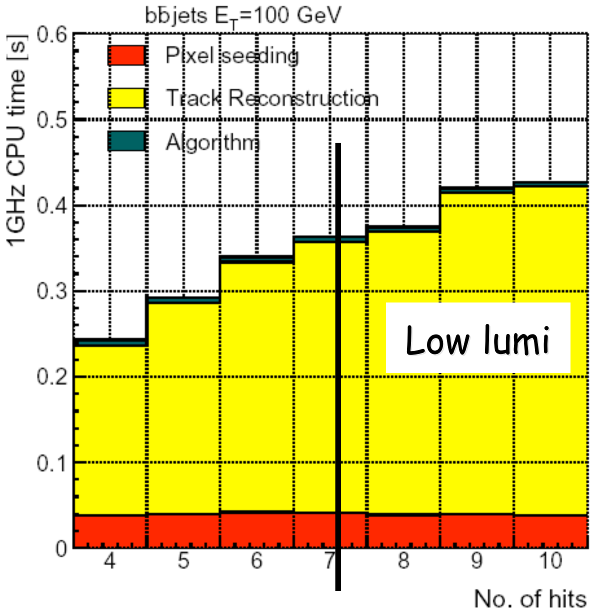
stop tracking here at HLT

CMS full simulation

b-tagging efficiency vs background (uds-jets) efficiency



CPU performance :
~ 300 ms/evt if
≤ 7 hits/track
Extrapolation to
2007 : ~ 40 ms/evt

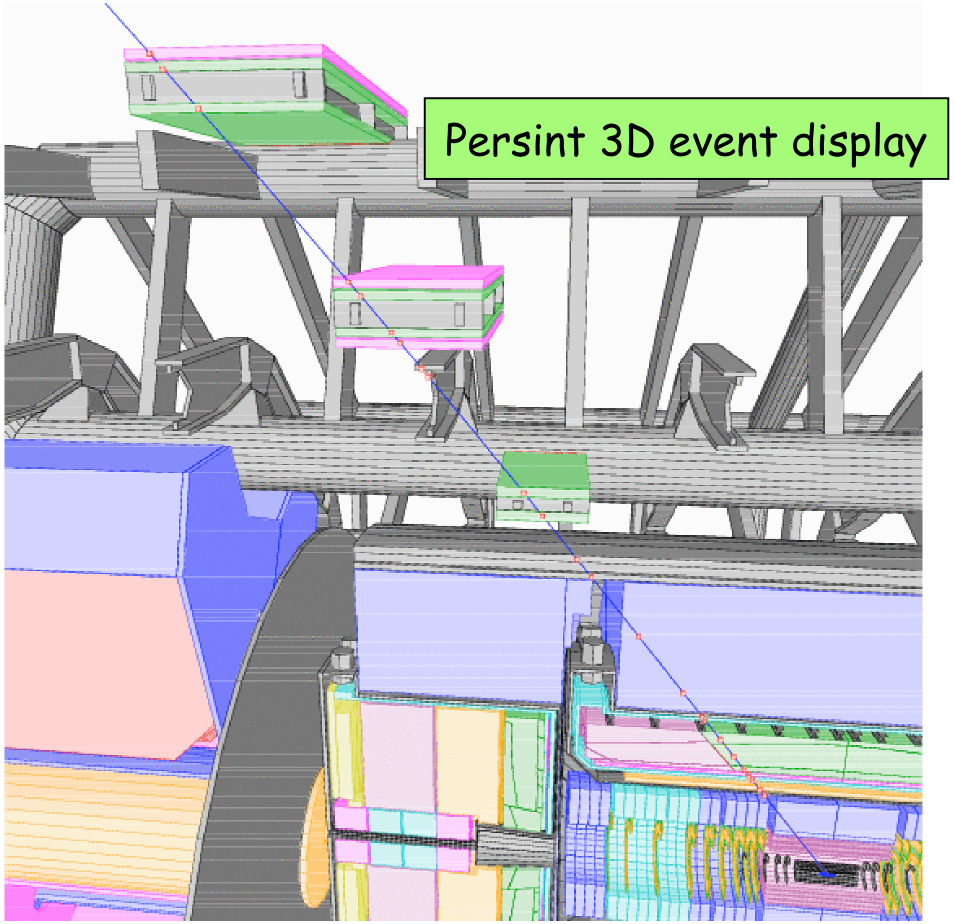


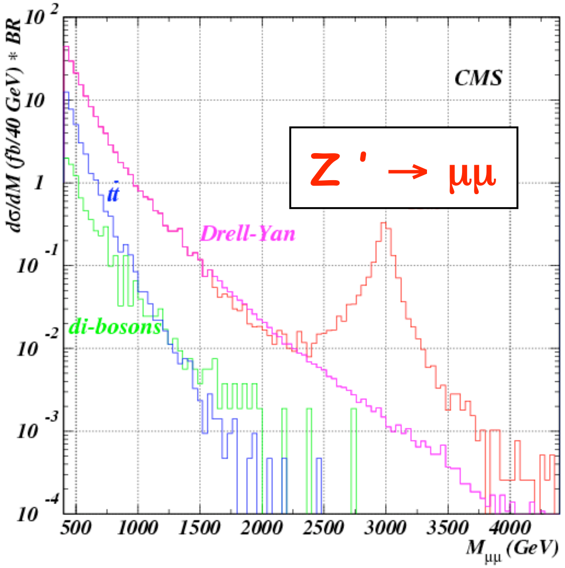
Note : this performance is for perfect detector.
Next step: introduce mis-alignments, etc.

Example 2 : Reconstruction of E ~ TeV muons in the ATLAS spectrometer

(most demanding because of very high energy)

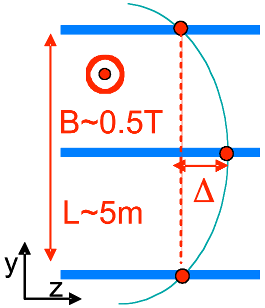
One of LHC goals : look for new resonances in the TeV region





Need :

- ε_μ (reconstruction) > 90% because < 10 evts expected for m (Z') ~ 5 TeV
- $\sigma/p < 10\%$ for $E_\mu \sim$ TeV to observe a "narrow" peak

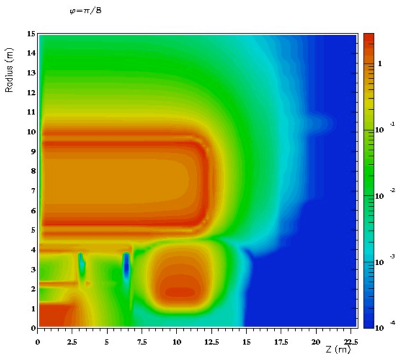


$E_\mu \sim 1$ TeV $\Rightarrow \Delta \sim 500\ \mu$m

$\sigma/p \sim 10\%$ $\Rightarrow \delta\Delta \sim 50\ \mu$m

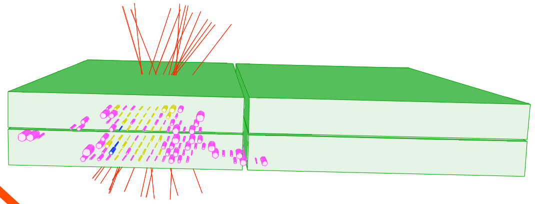
Accurate description of upstream material (to ~ 5%) and E-losses → detector description, simulation

Alignment to < 30 μm → calibration, Condition DB



Pattern recognition in highly non-uniform (air-core) toroidal field → access time to field-map ≤ 1 μs

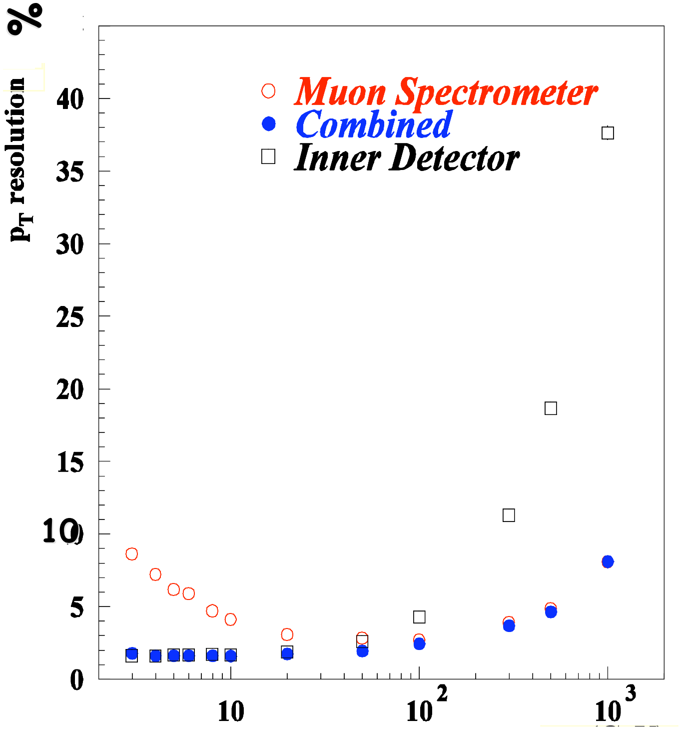
Catastrophic E-losses in calorimeters (probability increases with E) and cavern background → additional hits

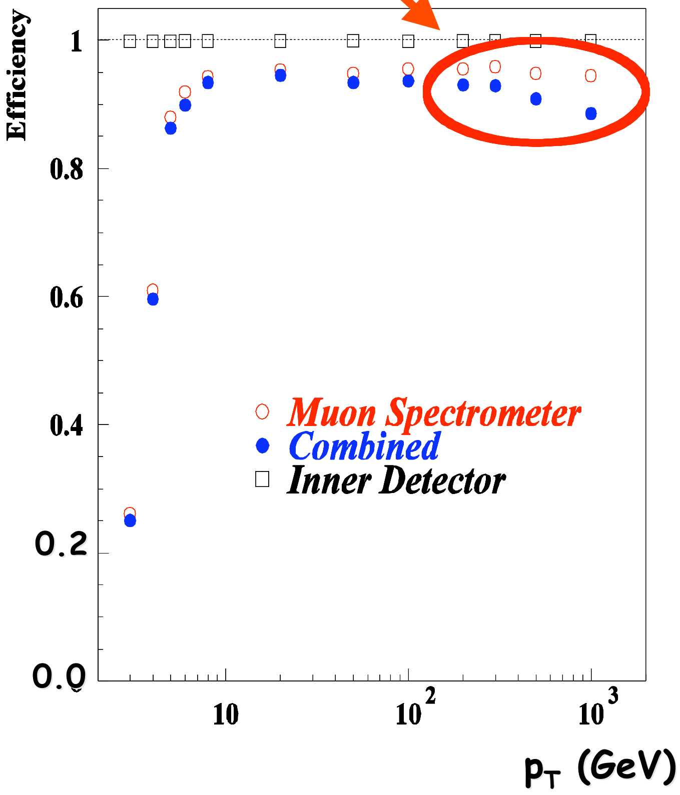


Examples of achieved performance (full simulation)

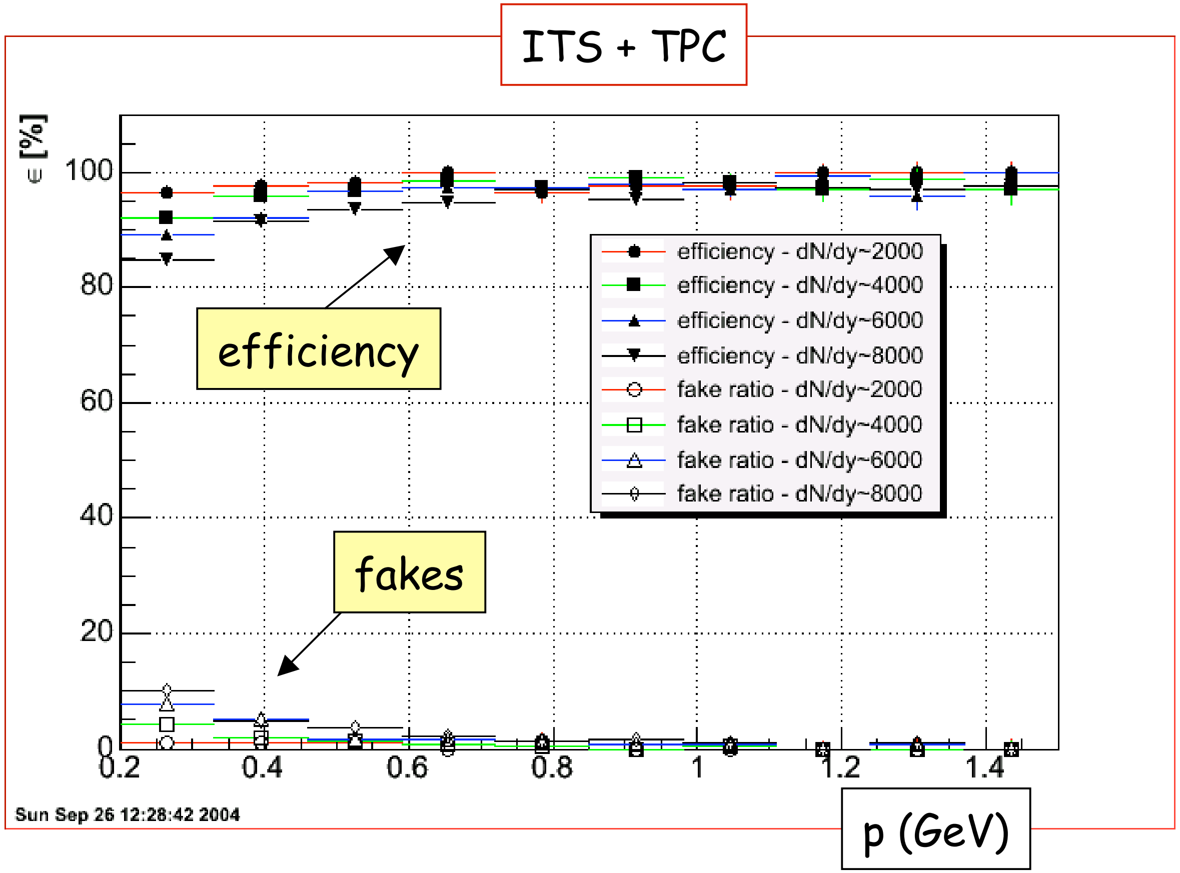
Muonboy

ATLAS
full simulation

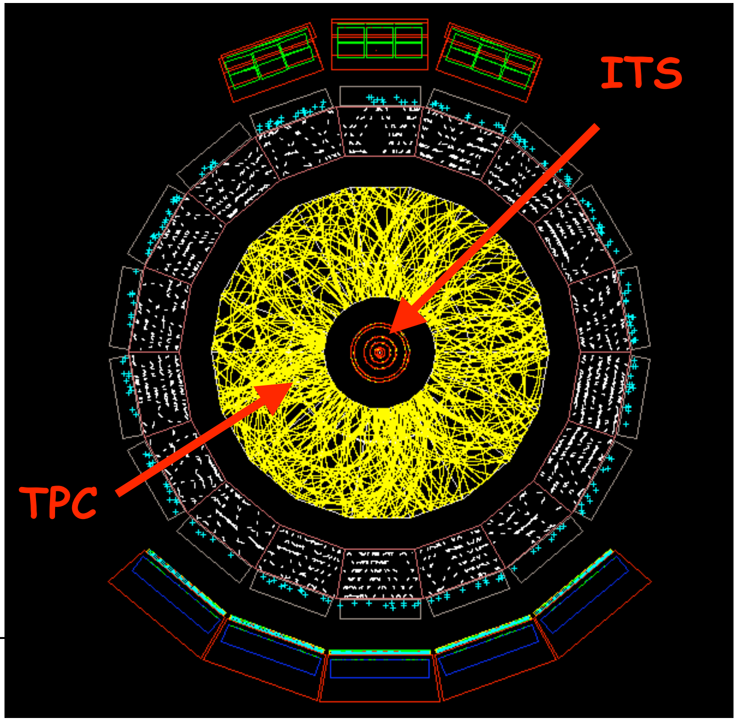




Example 3 : Tracking in ALICE (most demanding because of very-high particle multiplicity)



(see M. Ivanov's talk)

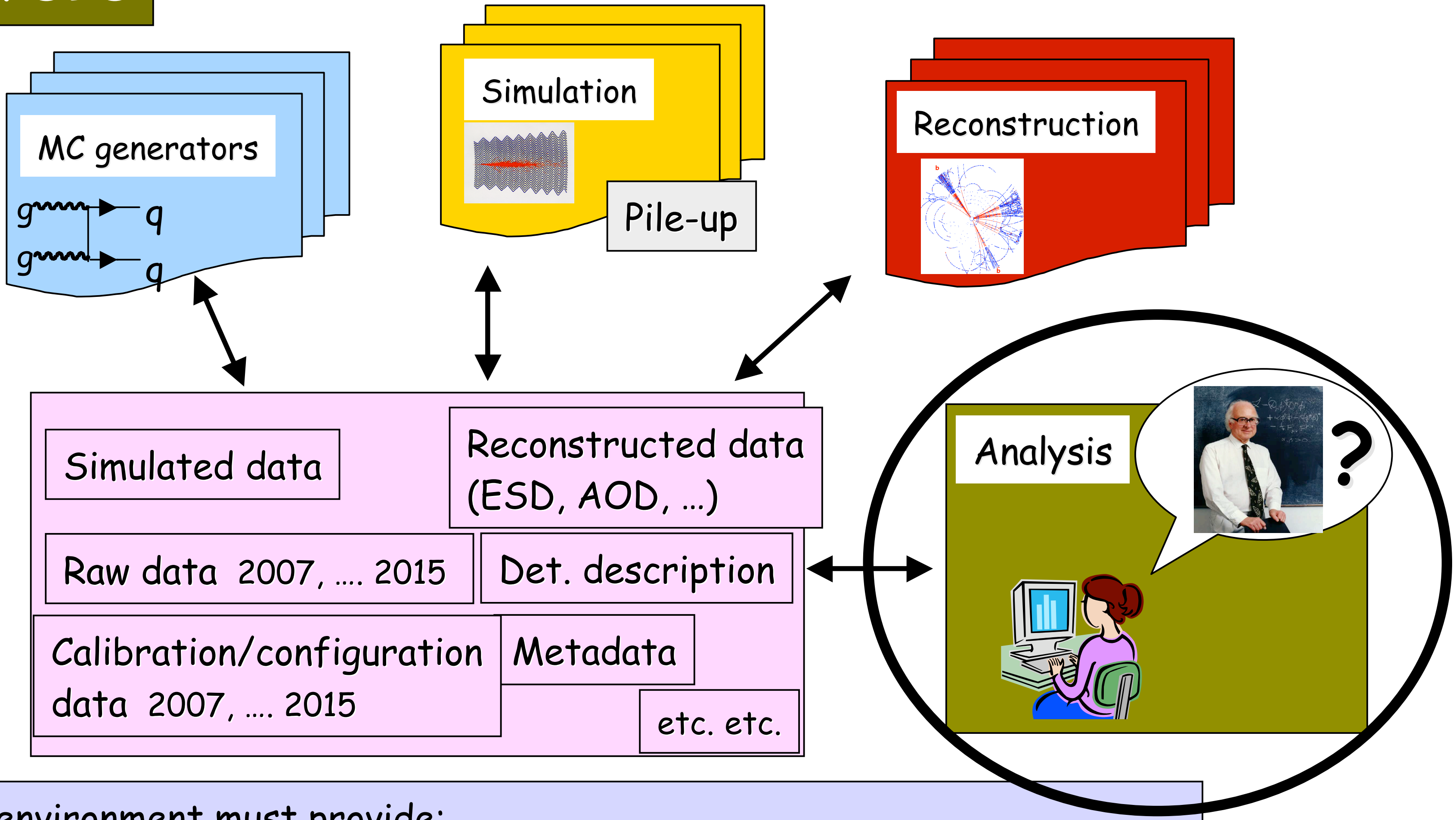


TPC redundancy (~160 points)

Special treatment of clusters with extended shapes to account for track overlaps

CPU : ~ 80 s for dN/dy ~ 6000 (3 GHz Pentium IV)

ANALYSIS



MC generators

Simulation

Pile-up

Reconstruction

Simulated data

Reconstructed data (ESD, AOD, …)

Raw data 2007, …. 2015

Det. description

Calibration/configuration data 2007, …. 2015

Metadata

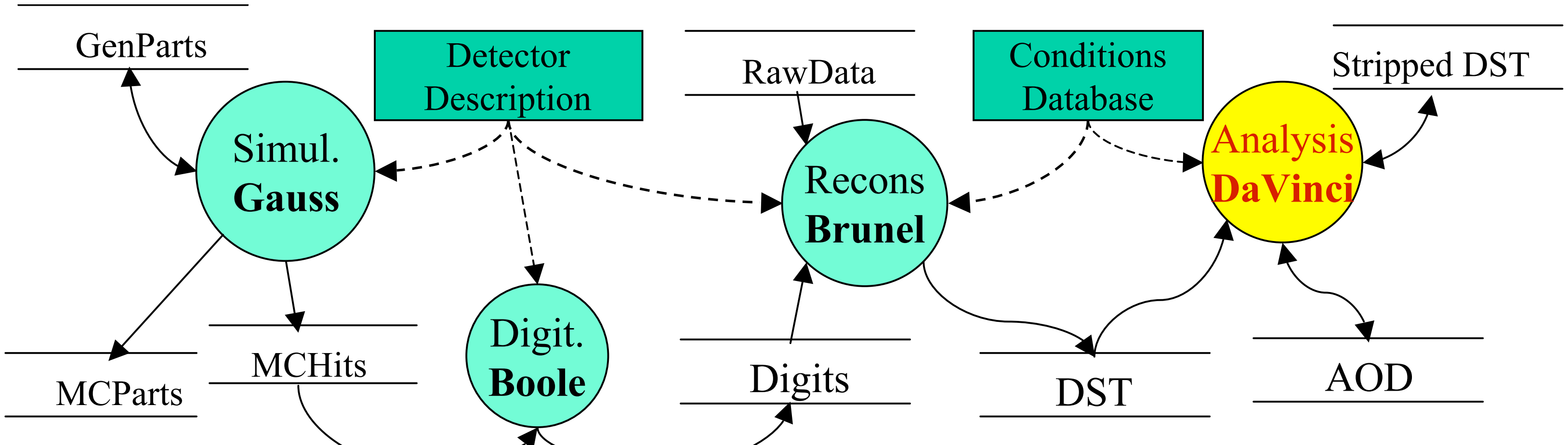
etc. etc.

Analysis

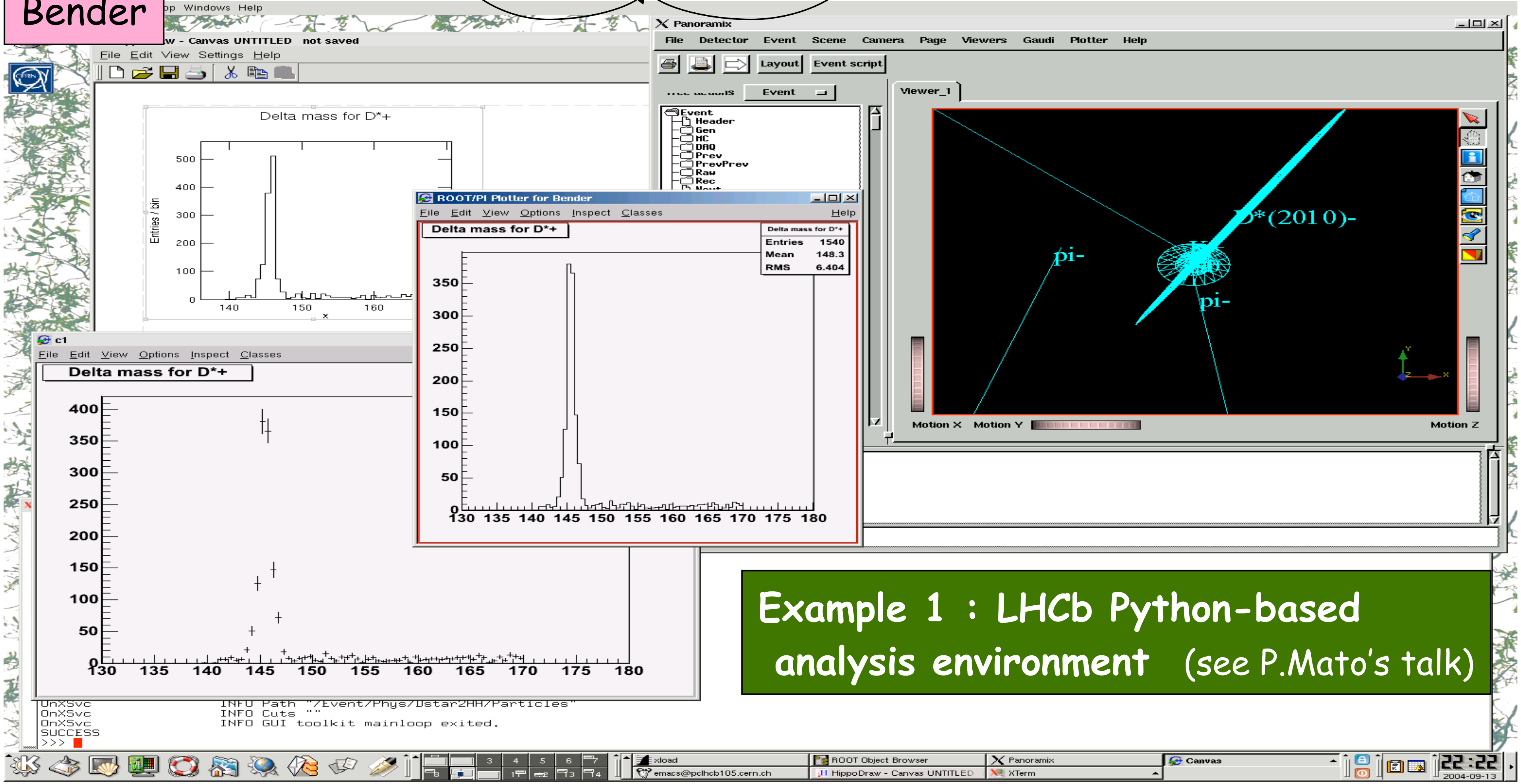
<u>Analysis environment must provide</u>:

- batch and interactive functionalities, transparent local ⇔ GRID transition
- access to all levels of data hierarchy (including Condition DB)
- access to event generation and simulation
- run reconstruction algorithms on raw and ESD data
- analysis at AOD and ntuple levels
- data, algorithm and task browsing, event display, visualization, …. etc. etc. ….

Note: " interactive" means:
[t (request) - t (answer)] ≈ sec



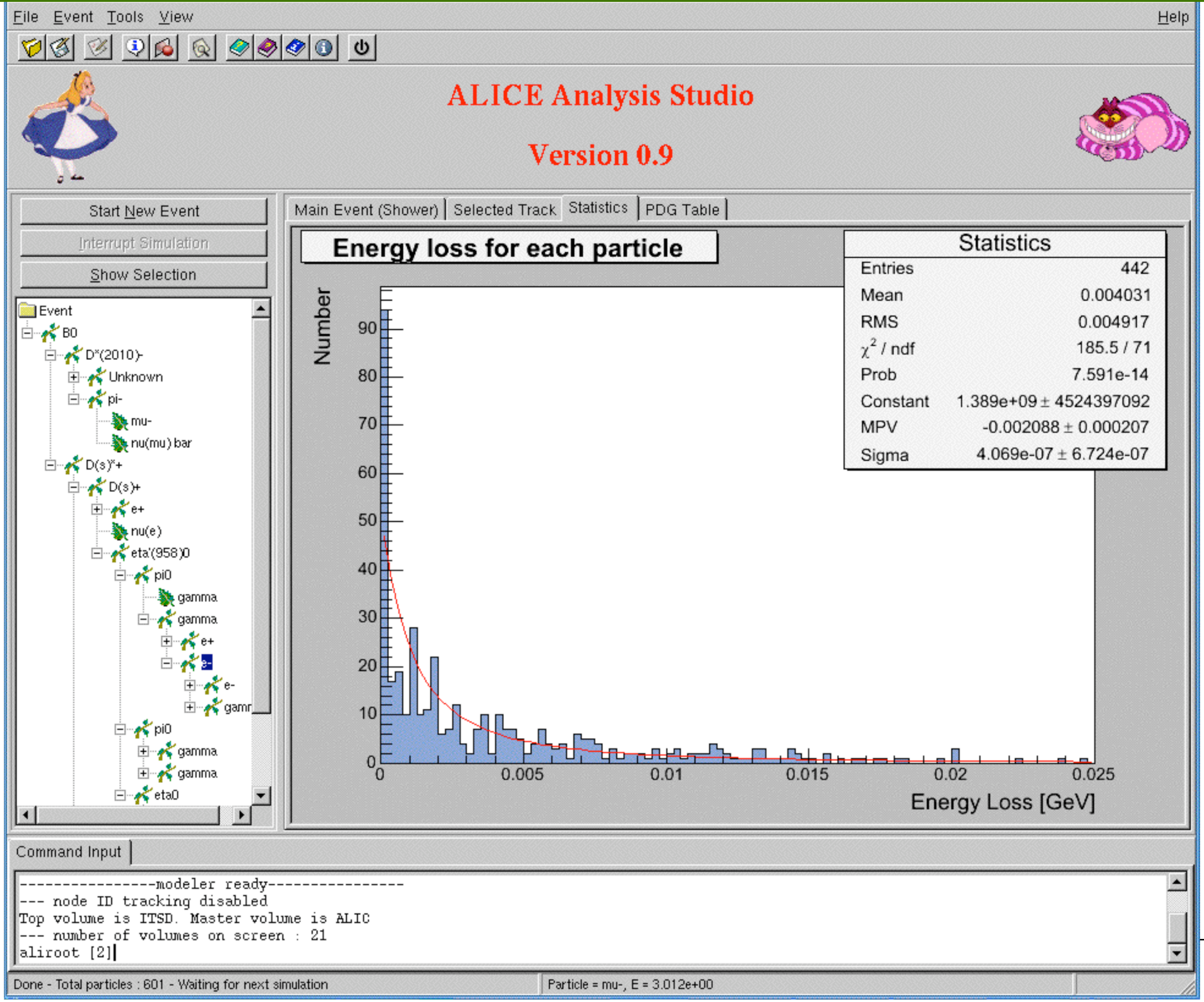
Bender



```
OnXSvc                 INFO Path "/Event/Phys/Dstar2HH/Particles"
OnXSvc                 INFO Cuts ""
OnXSvc                 INFO GUI toolkit mainloop exited.
SUCCESS
>>>
```

Example 1 : LHCb Python-based analysis environment (see P.Mato's talk)

Example 2 : ALICE ROOT-based analysis environment (see F. Carminati's talk)

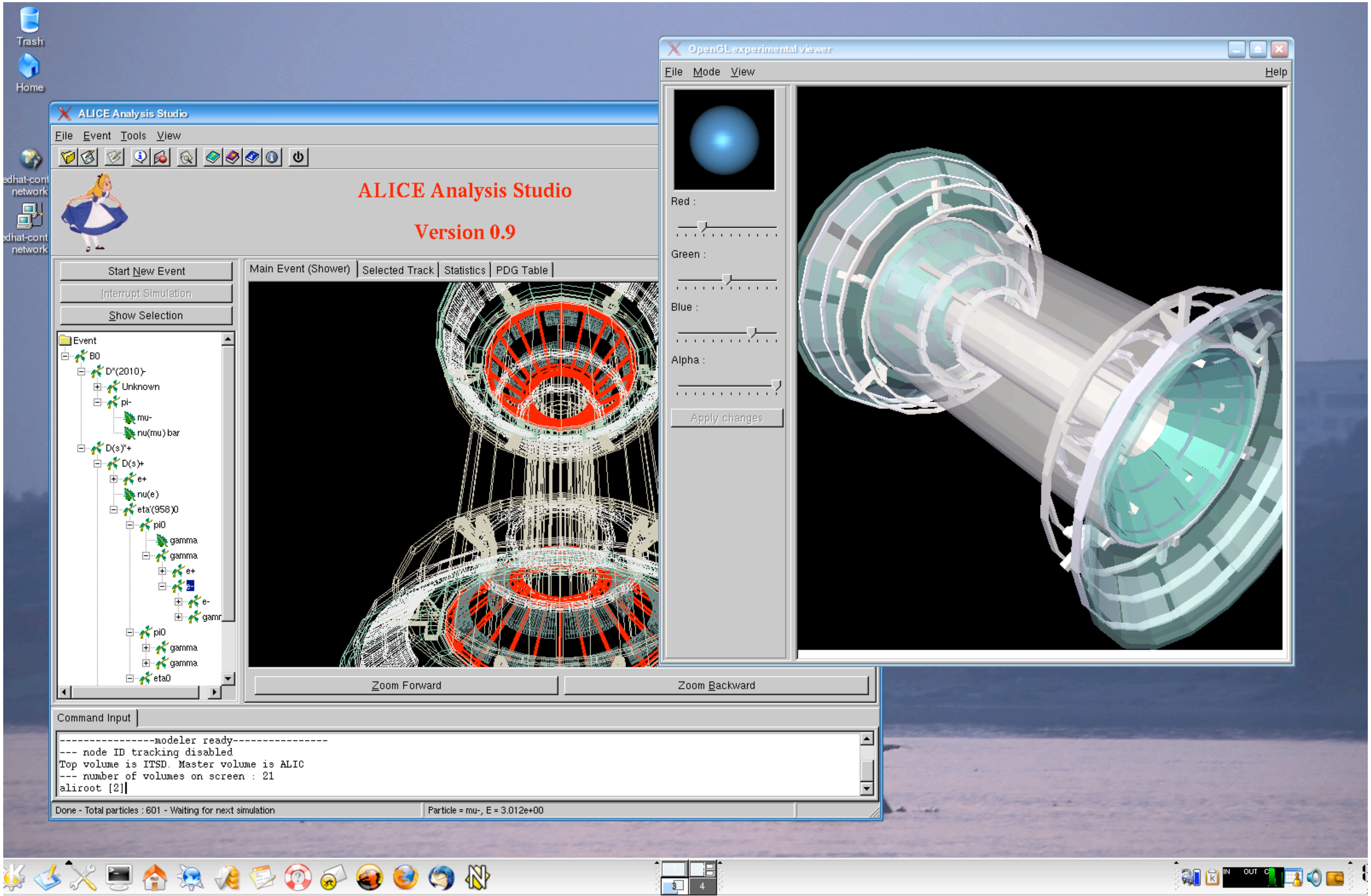


Statistics	
Entries	442
Mean	0.004031
RMS	0.004917
χ^2 / ndf	185.5 / 71
Prob	7.591e-14
Constant	1.389e+09 ± 4524397092
MPV	-0.002088 ± 0.000207
Sigma	4.069e-07 ± 6.724e-07

```
----------------modeler ready----------------
--- node ID tracking disabled
Top volume is ITSD. Master volume is ALIC
--- number of volumes on screen : 21
aliroot [2]
```

Done - Total particles : 601 - Waiting for next simulation

Particle = mu-, E = 3.012e+00

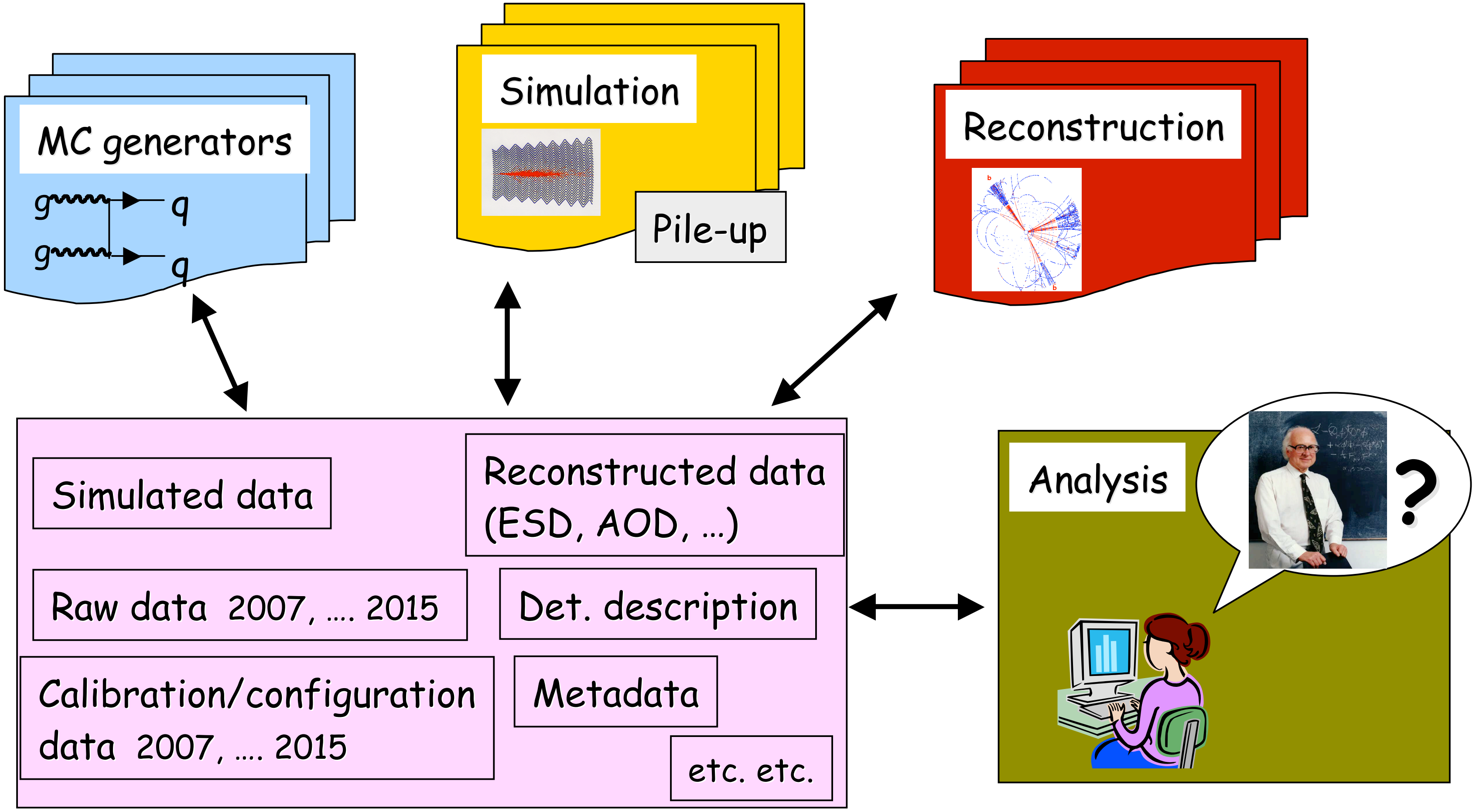


```
----------------modeler ready----------------
--- node ID tracking disabled
Top volume is ITSD. Master volume is ALIC
--- number of volumes on screen : 21
aliroot [2]
```

Done - Total particles : 601 - Waiting for next simulation

Particle = mu-, E = 3.012e+00

In addition ….



MC generators

Simulation

Pile-up

Reconstruction

Simulated data

Reconstructed data (ESD, AOD, …)

Raw data 2007, …. 2015

Det. description

Calibration/configuration data 2007, …. 2015

Metadata

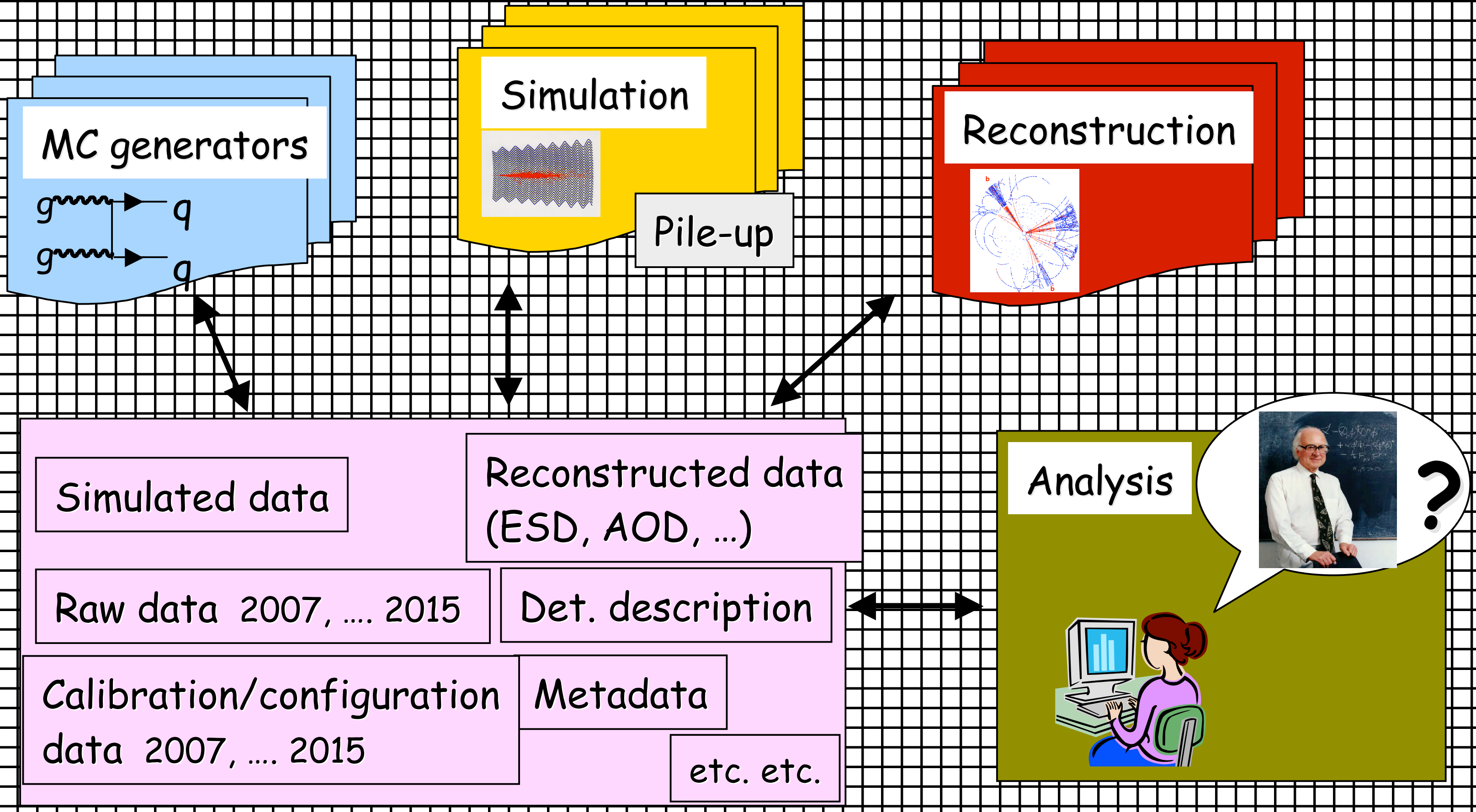
etc. etc.

Analysis

Evolution: with Software at time T I should be able to handle data (real, simulated, calibrations) produced at $T - N_{years}$ earlier $(N_{years} \le 15)$ in a transparent way

GRID:

- should allow everybody to do physics anywhere anytime
- should not slow down delivery of physics results
- must be transparent to the users
- user support must be available 24 hours x 365 days x 15 years
- basic tools (e.g. easy-to-understand diagnostic about failed jobs, job monitoring histos, status of job and resources) should be available to the end-users



MC generators

Simulation

Pile-up

Reconstruction

Simulated data

Reconstructed data
(ESD, AOD, …)

Raw data 2007, …. 2015

Det. description

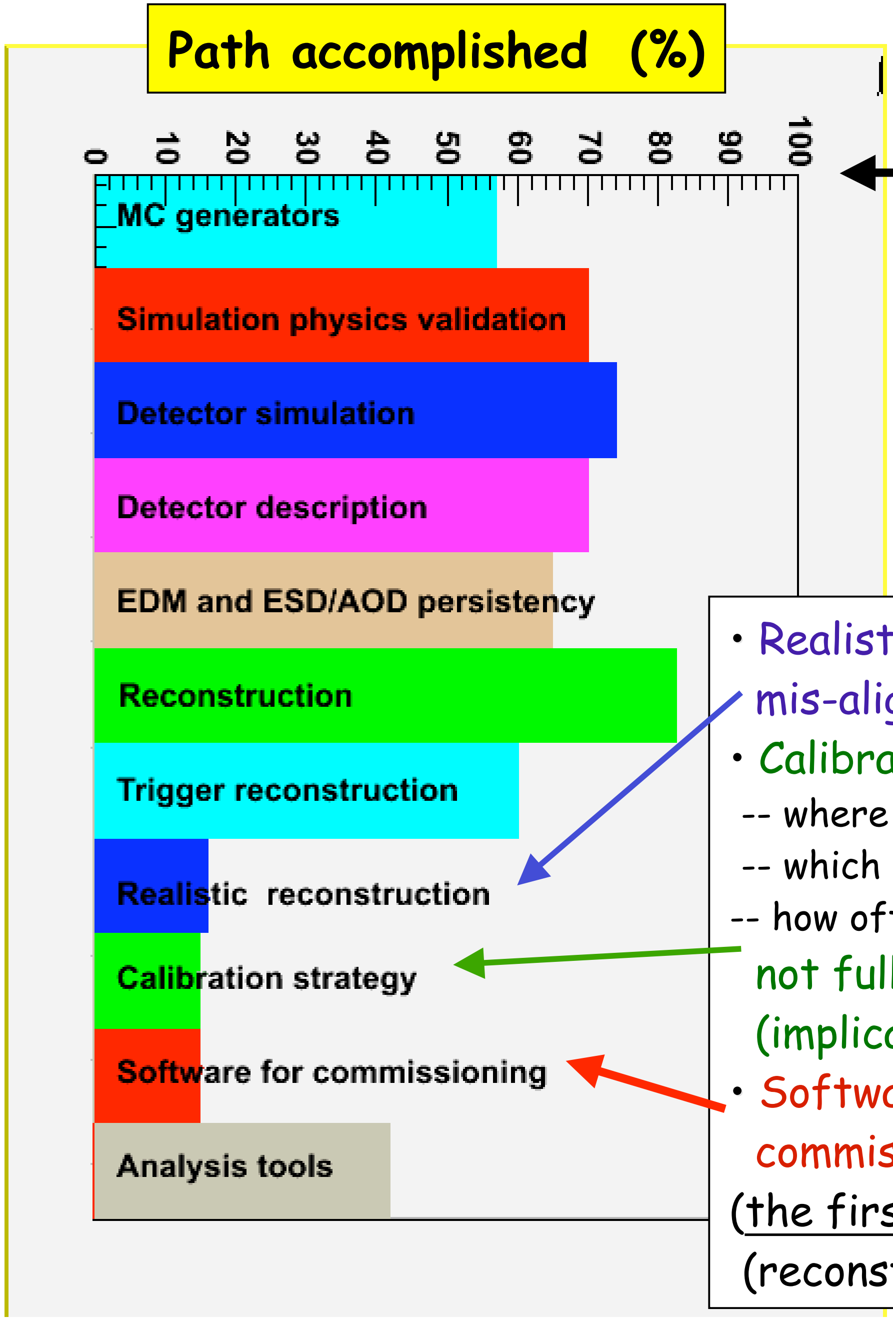
Calibration/configuration
data 2007, …. 2015

Metadata

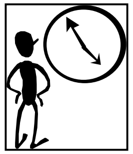
etc. etc.

Analysis

❸ Where do we stand today with the LHC (non-core) Software ?



2007



My very rough estimate …
average over 4 experiments

- Realistic detectors (HV problems, dead channels, mis-alignments, …) not yet implemented
- Calibration strategy :
  -- where (Event Filter, Tier0) ?
  -- which streams, which data size ?
  -- how often, how many reprocessings of part of raw data ? C PU ?
  not fully developed in most cases
  (implications for EDM and Computing Model ?)
- Software for experiment monitoring and for commissioning with cosmic and beam-halo muons
  (the first real data to be collected …) not developed yet
  (reconstruction must cope with atypical events …)

Conclusions

My 2 main messages (as an LHC physicist and end-user):

- LHC has unprecedented and highly compelling physics goals
  → Software/Computing should not limit the detector performance and LHC physics reach
- In spite of challenges and difficulties, the Software must be easy-to-use and stable

My 2 main worries today (as an LHC physicist and end-user):

- End-users not yet exposed to massive use/navigation of database and of GRID
  → what will happen when O(10^3) physicists will simultaneously access these systems ?
- Software and Computing Model developed for steady-state LHC operation (≥ 2009 ?)
  But : at the beginning they will be confronted with most atypical (and stressful) situations, for which a lot of flexibility will be needed:
  -- staged, non-perfect, non-calibrated, non-aligned detectors with all sorts of problems
  -- cosmic and beam-halo muons used to calibrate detectors during machine commissioning
  -- machine backgrounds ; higher-than-expected trigger rates
  -- fast/frequent reprocessing of part of data (e.g. special calibration streams)
  -- O(10^3) physicists in panic-mode using and modifying the Software and accessing the database, GRID …

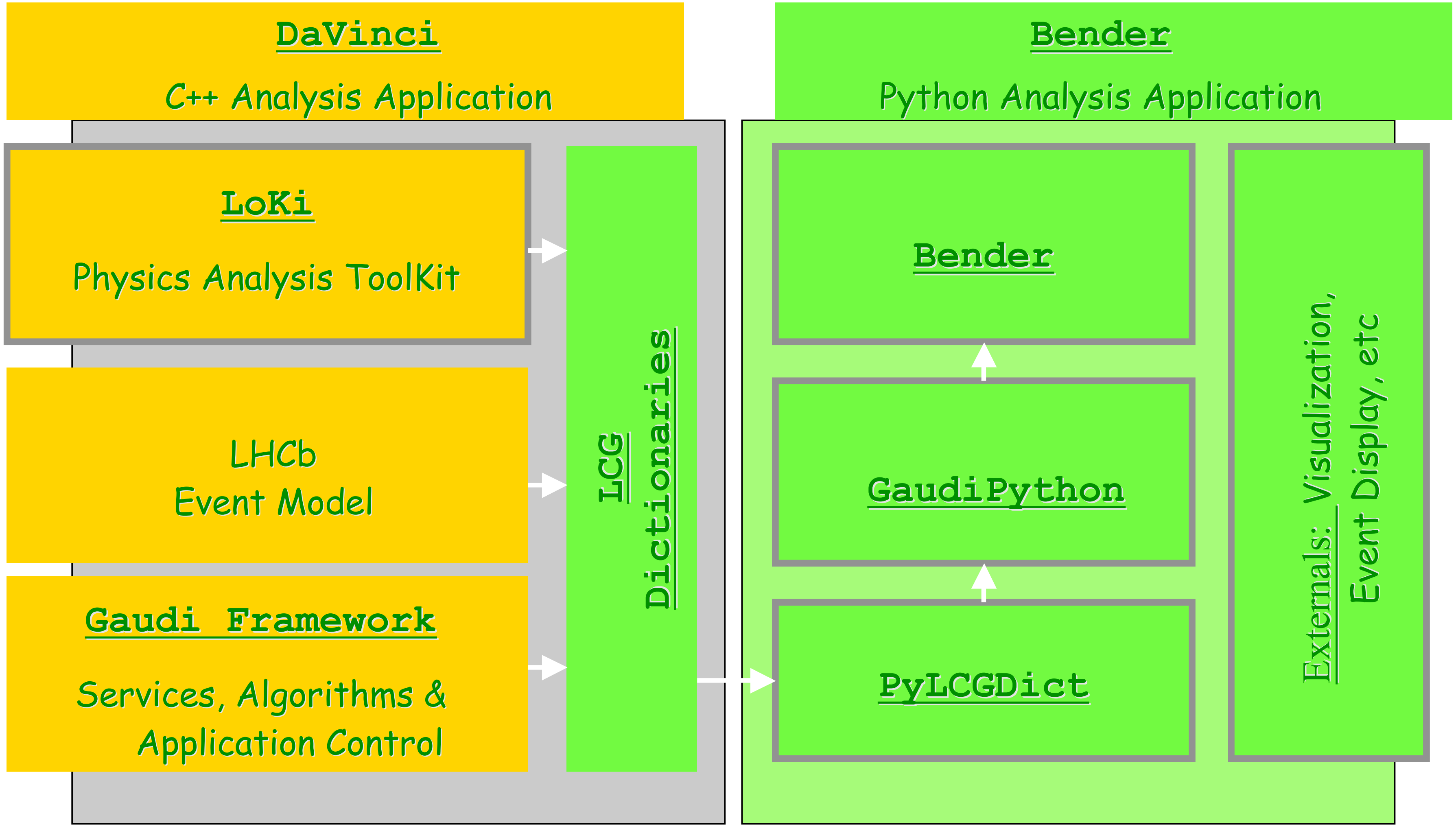
⇒ it is time for the Software/Computing to address the early phase of LHC operation, not to hinder the fast delivery of physics results (and a possible early discovery …)

Many thanks to :

D. Barberis, P. Bartalini, R. Brun, F. Carminati, L. Chevalier, G. Corti, A. Dell'Acqua, A. De Roeck, D. Froidevaux, R. Hawkings, V. Innocente, M. Ivanov, J.-F. Laporte, T. Lari, L. Mapelli, P. Mato, A. Nairz, W. Pokorski, D. Quarrie, F. Radermaker, A. Rimoldi, M. Stavrianakou, L. Silvestris, P. Sphicas, F. Teubert, S. Valuev, M. Virchaux

BACK-UP SLIDES

LHCb Analysis Applications



DaVinci

C++ Analysis Application

LoKi

Physics Analysis ToolKit

LHCb
Event Model

Gaudi Framework

Services, Algorithms & Application Control

LCG Dictionaries

Bender

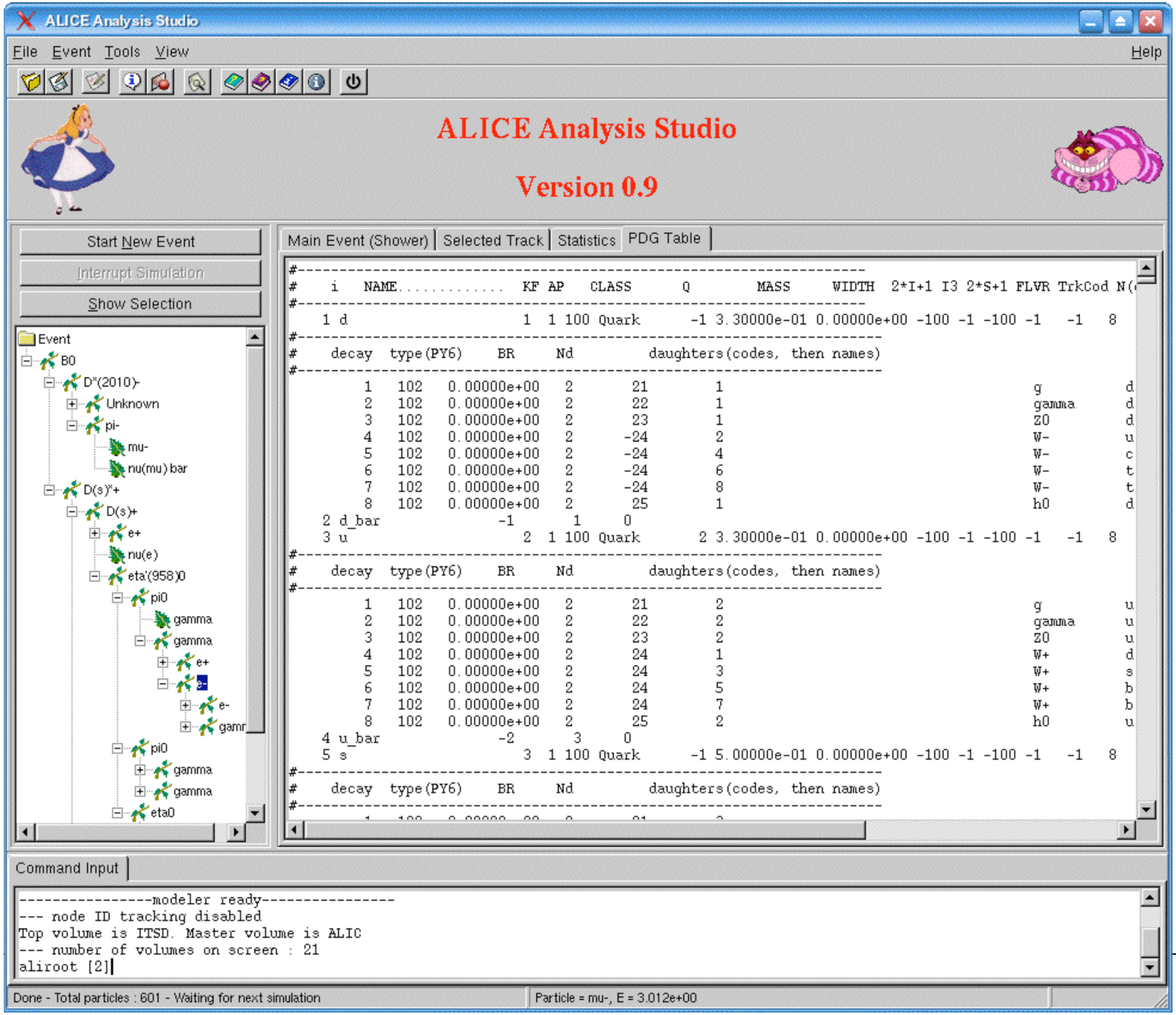
Python Analysis Application

Bender

GaudiPython

PyLCGDict

Externals: Visualization, Event Display, etc

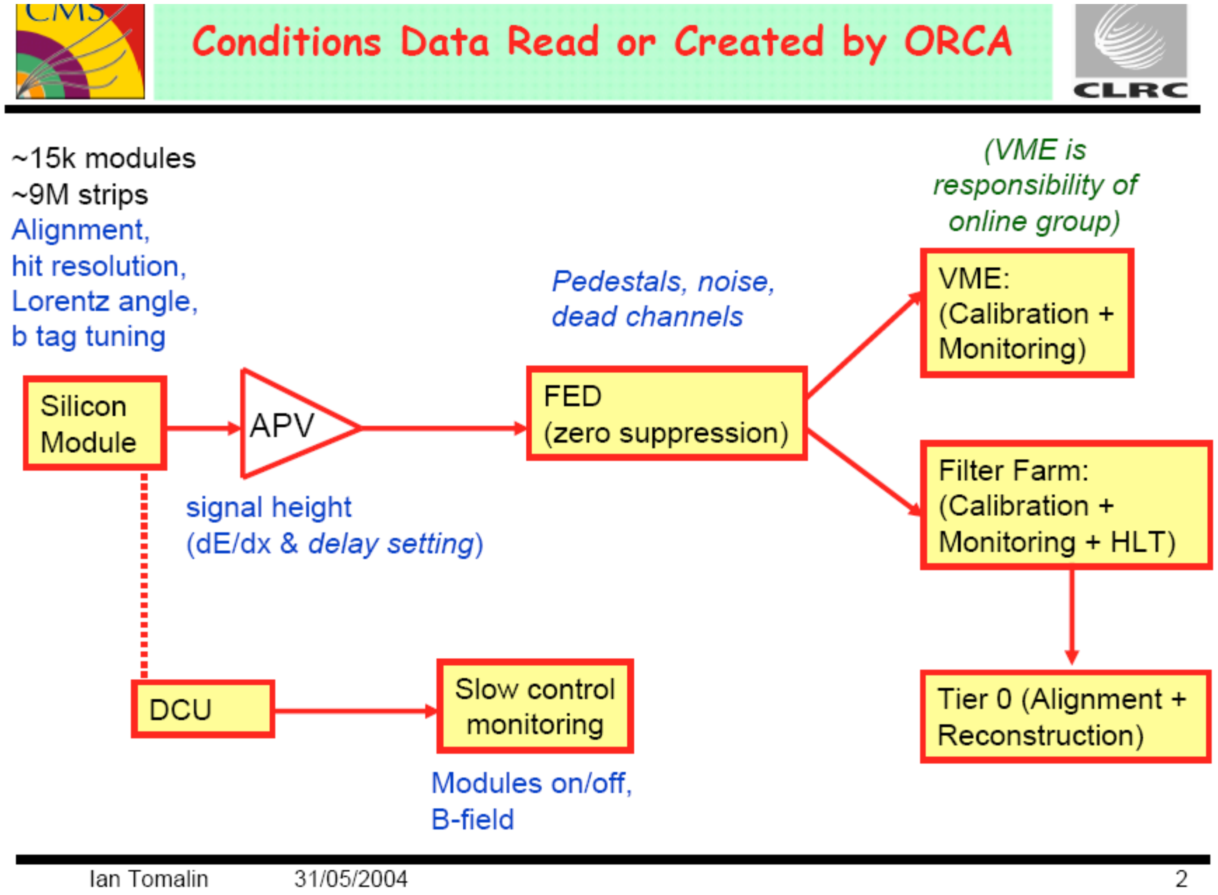


```
#-----------------------------------------------------------------------
#    i   NAME.............  KF AP   CLASS      Q        MASS     WIDTH  2*I+1 I3 2*S+1 FLVR TrkCod N(
#-----------------------------------------------------------------------
    1 d                     1  1 100 Quark      -1 3.30000e-01 0.00000e+00 -100 -1 -100 -1   -1   8
#-------------------------------------------------------------------------
#    decay  type(PY6)    BR      Nd         daughters(codes, then names)
#-------------------------------------------------------------------------
         1   102   0.00000e+00   2       21        1                                   g          d
         2   102   0.00000e+00   2       22        1                                   gamma      d
         3   102   0.00000e+00   2       23        1                                   Z0         d
         4   102   0.00000e+00   2      -24        2                                   W-         u
         5   102   0.00000e+00   2      -24        4                                   W-         c
         6   102   0.00000e+00   2      -24        6                                   W-         t
         7   102   0.00000e+00   2      -24        8                                   W-         t
         8   102   0.00000e+00   2       25        1                                   h0         d
    2 d_bar              -1        1      0
    3 u                     2  1 100 Quark       2 3.30000e-01 0.00000e+00 -100 -1 -100 -1   -1   8
#-------------------------------------------------------------------------
#    decay  type(PY6)    BR      Nd         daughters(codes, then names)
#-------------------------------------------------------------------------
         1   102   0.00000e+00   2       21        2                                   g          u
         2   102   0.00000e+00   2       22        2                                   gamma      u
         3   102   0.00000e+00   2       23        2                                   Z0         u
         4   102   0.00000e+00   2       24        1                                   W+         d
         5   102   0.00000e+00   2       24        3                                   W+         s
         6   102   0.00000e+00   2       24        5                                   W+         b
         7   102   0.00000e+00   2       24        7                                   W+         b
         8   102   0.00000e+00   2       25        2                                   h0         u
    4 u_bar              -2        3      0
    5 s                     3  1 100 Quark      -1 5.00000e-01 0.00000e+00 -100 -1 -100 -1   -1   8
#-------------------------------------------------------------------------
#    decay  type(PY6)    BR      Nd         daughters(codes, then names)
#-------------------------------------------------------------------------
```

```
----------------modeler ready----------------
--- node ID tracking disabled
Top volume is ITSD. Master volume is ALIC
--- number of volumes on screen : 21
aliroot [2]
```

Done - Total particles : 601 - Waiting for next simulation

Particle = mu-, E = 3.012e+00



Conditions Data Read or Created by ORCA

~15k modules
~9M strips
Alignment,
hit resolution,
Lorentz angle,
b tag tuning

(VME is responsibility of online group)

Pedestals, noise, dead channels

Silicon Module

APV

FED (zero suppression)

VME: (Calibration + Monitoring)

Filter Farm: (Calibration + Monitoring + HLT)

Tier 0 (Alignment + Reconstruction)

signal height (dE/dx & *delay setting*)

DCU

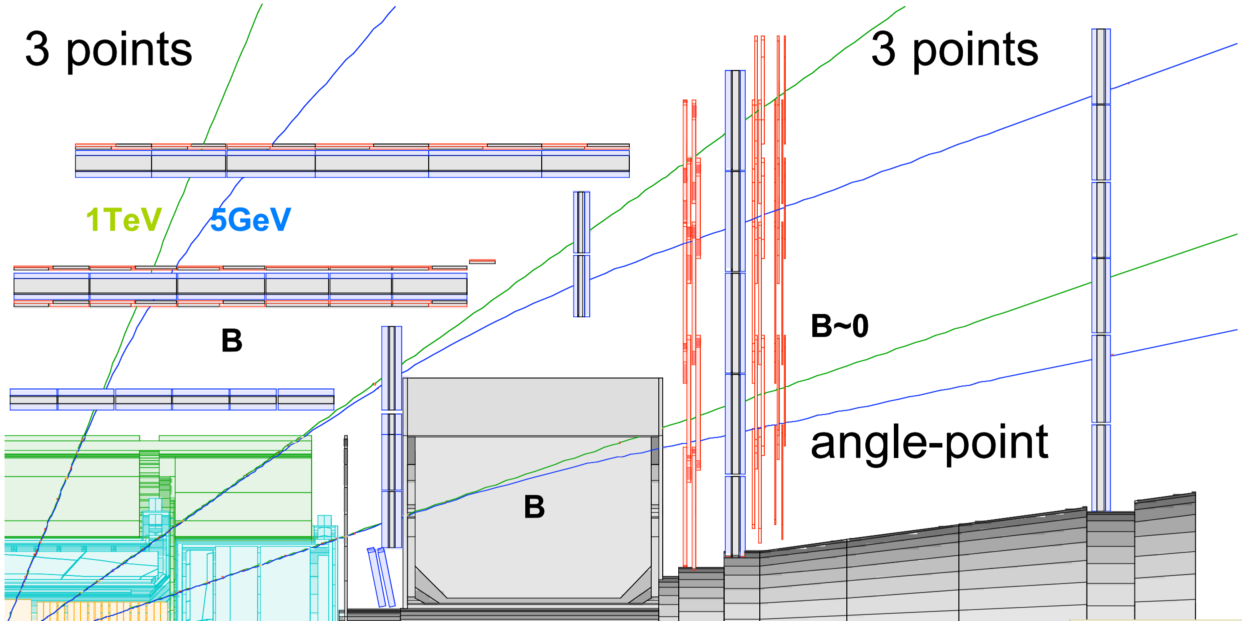
Slow control monitoring

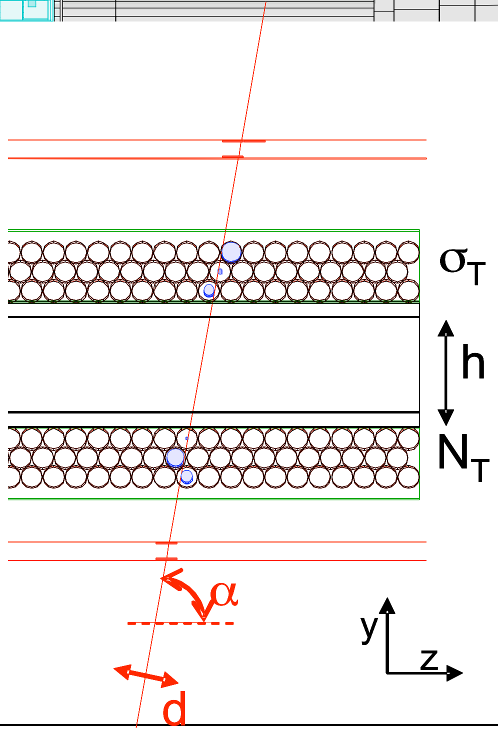
Modules on/off,
B-field

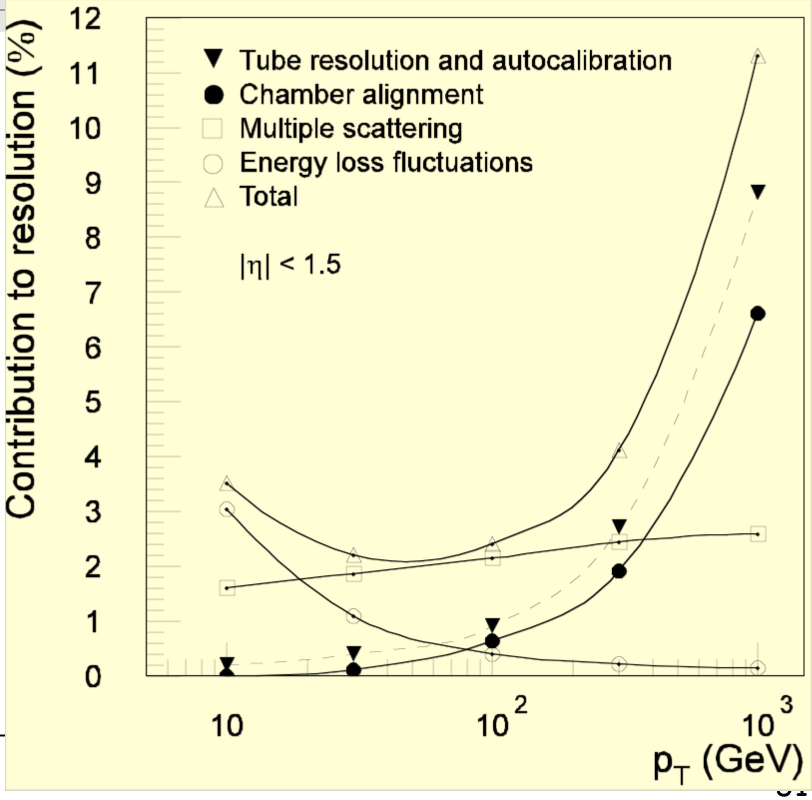
Ian Tomalin 31/05/2004 2

- Data stored in condition database : few TB/year
- Update frequency of calibration/alignment constants : once/hour --> once/run depending on detector
- Slow-control data updated every few minutes

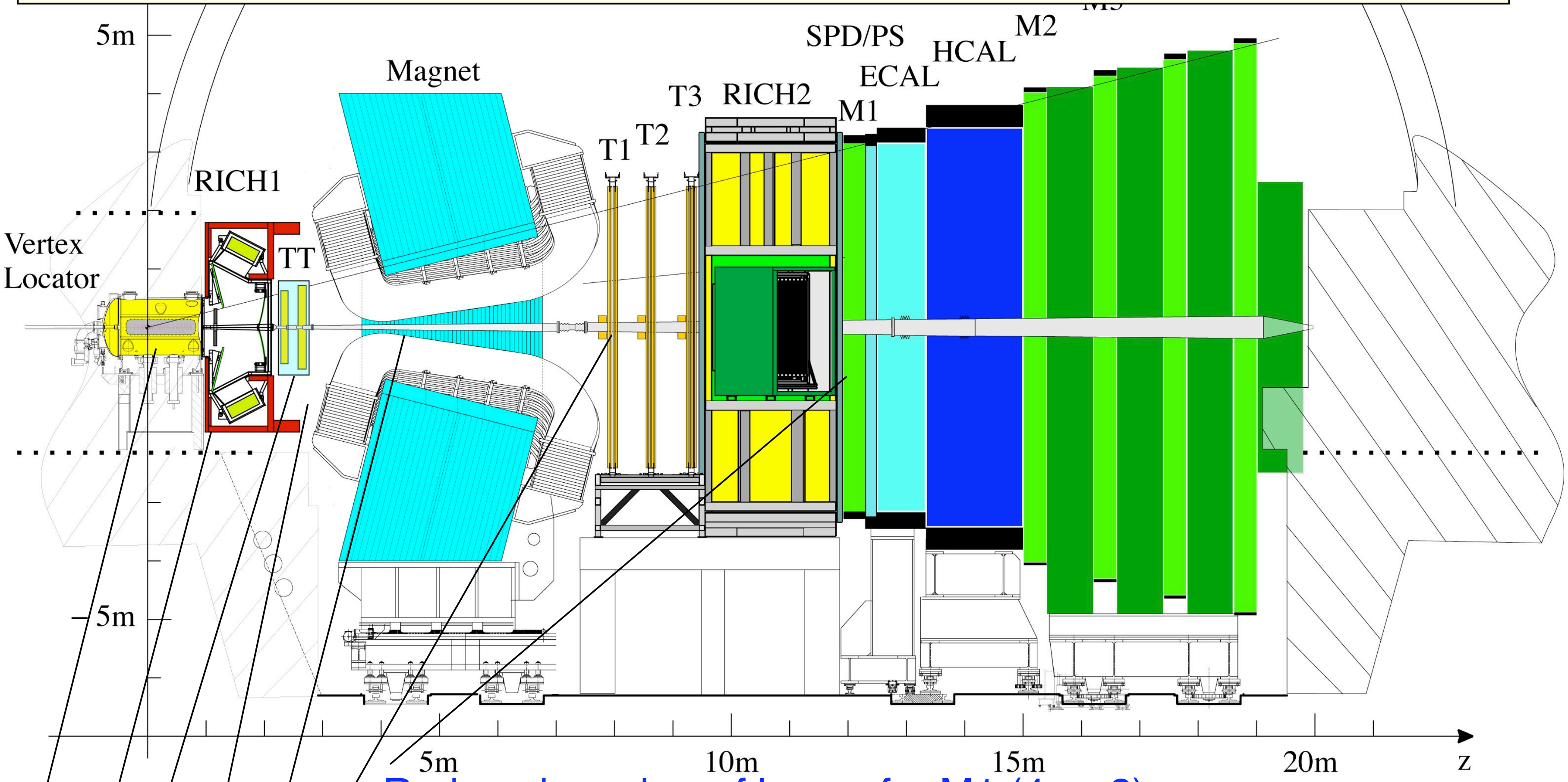
Note : need to update detector geometry with time automatically using latest constants


3 points
3 points
1TeV
5GeV
B
B~0
B
angle-point


σ_T
h
N_T
α
d
y
z


Contribution to resolution (%)
0
1
2
3
4
5
6
7
8
9
10
11
12
▼ Tube resolution and autocalibration
● Chamber alignment
☐ Multiple scattering
○ Energy loss fluctuations
△ Total
|η| < 1.5
10
10^2
10^3
p_T (GeV)

The LHCb Detector



Reduced number of layers for M1 (4 → 2)

Reduced number of tracking stations behind the magnet (4 → 3)

No tracking chambers in the magnet

No B field shielding plate

Full Si station

Reoptimized RICH-1 design

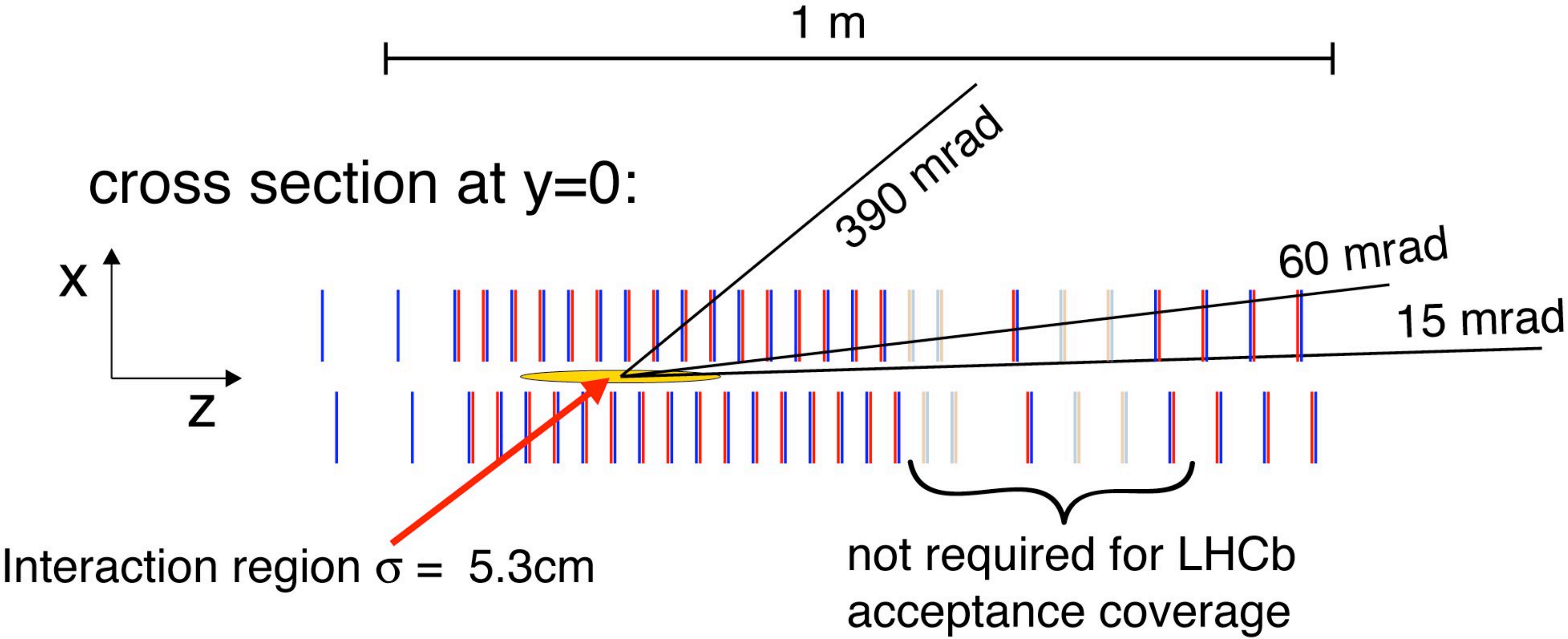
Reduced number of VELO stations (25 → 21)

Changes were made for **material reduction** and **L1 trigger improvement**

VELO

–

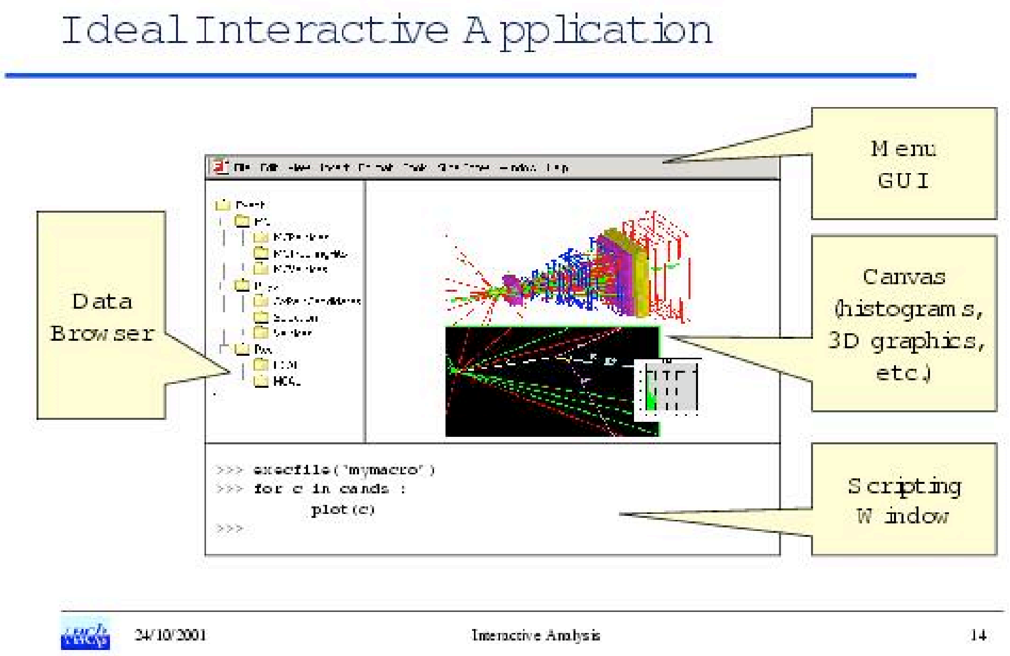
- 19 stations needed to cover the LHCb



Track loss for 25→21= 0.1%, 21→20 = 0.5%

(X_0 is dominated by the RF foil)

Pere Mato's dream



Ideal Interactive Application

Menu GUI

Data Browser

Canvas (histograms, 3D graphics, etc.)

Scripting Window

```
>>> execfile('mymacro')
>>> for c in cands :
        plot(c)
>>>
```

24/10/2001 Interactive Analysis 14