

# Οι Υπολογιστές στη Φυσική Υψηλών Ενέργειων

I. Αποστολακης  
**CERN**

v0.7, 16:05 26η Ιουνίου 2008

# Πλano της ομιλιας

## ⌘ Η χρήση των Υπολογιστων

- ☑ Ανακατασκευη (reconstruction)

  - ☒ Αμεσως (online) ή αργοτερα (off-line)

- ☑ Προσομοιωση (simulation)

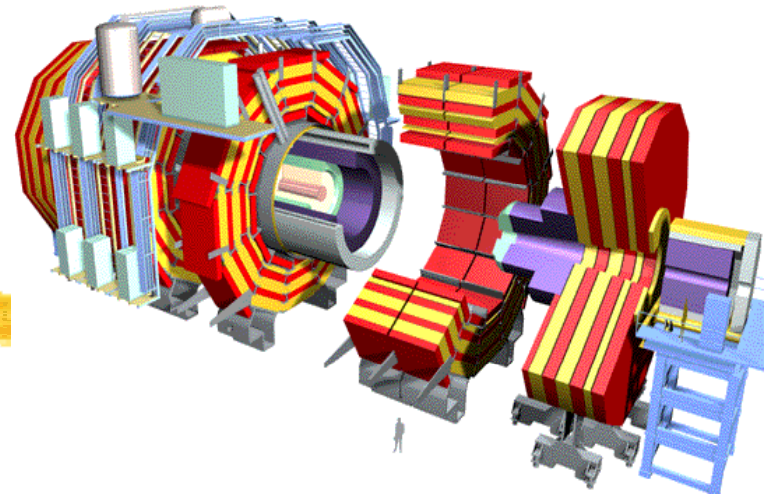
- ☑ Αναλυση δεδομενων (data analysis)

## ⌘ Μεγεθη και το GRID

- ☑ Υπολογιστικες αναγκες και ... GRID

## ⌘ Ομοιωτητες με αλλες “εφαρμογες”

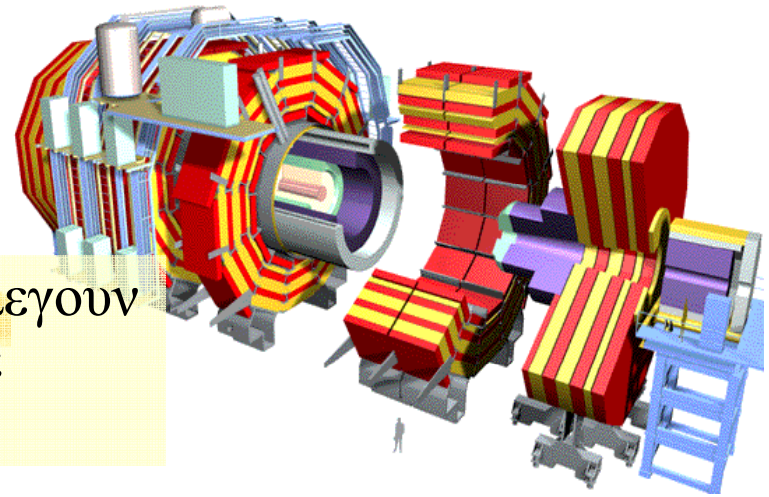
# Ο Επιταχυντής LHC



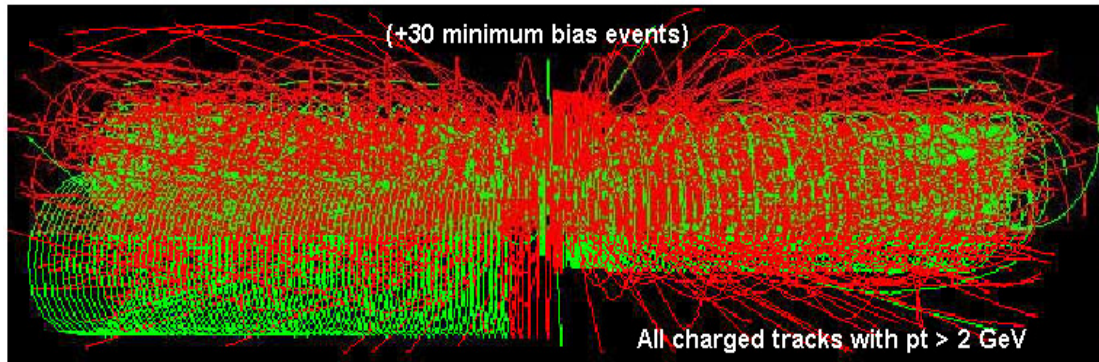
Ο επιταχυντής θα παραγει 40 εκατομυρια συγκρουσεις σωματιδιων (events) καθε δευτερολεπτο στο κεντρο καθε του ανιχνευτη καθε πειραματος



# Τα δεδομένα (*data*) του **LHC**



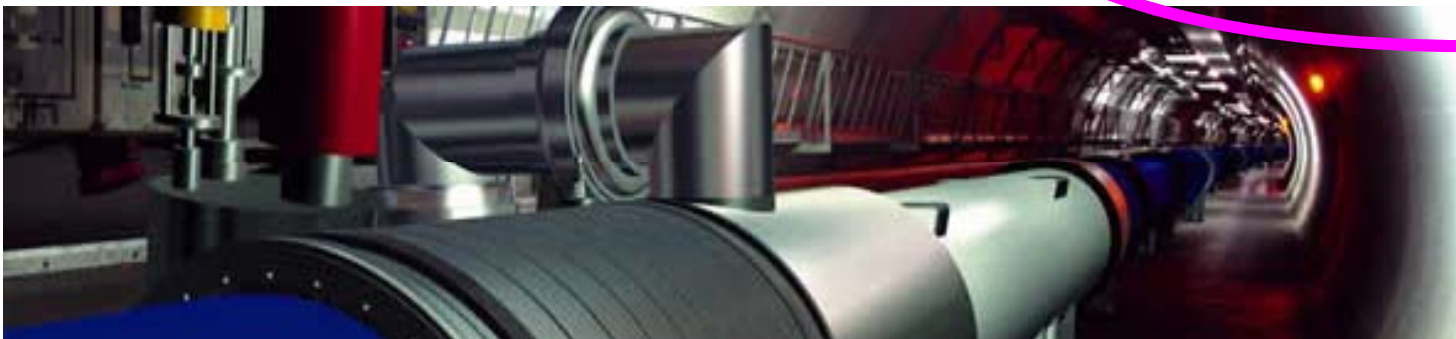
Τα ηλεκτρονικά και οι online υπολογιστές διαλέγουν μερικές εκατονταδες «καλά» events καθε δευτερολεπτο.



1 Peta = 1,000 Tera  
= 1,000,000 Giga

Καταγραφονται σε δισκους και μαγνητικες ταινιες με ρυθμο 100-1,000  
MegaBytes/sec =

→ 15 PetaBytes το χρονο  
για τα τεσσερα πειραματα μαζι



# Ανακατασκευή



Μια γρήγορη εισαγωγή

# Atlas : Physics Signatures and Event Rates

- Οι δεσμες διασταυρονονται με ρυθμο 40 MHz
- $\sigma_{inelastic} = 80 \text{ mb}$ 
  - Σε καθε περασμα πολles συγκρουσεις
  - $10^9$  συγκρουσεις το δευτερολεπτο
- Διαφορετικοι στοχοι, ο καθενας με τη δικια του «υπογραφη»
  - Το Χιγκς (Higgs) μεσονιο
  - Υπερσυμμετρια (Supersymmetry)
  - Το αγνωστο
  - Οι συμμετριες στα B μεσονια
- Τα ενδιαφεροντα συμβαντα ειναι καρφιτσες στα αχυρα σε ενα χωριο γιοματο σταβλους ( $\sim 1$  in  $10^5 - 10^9$ )

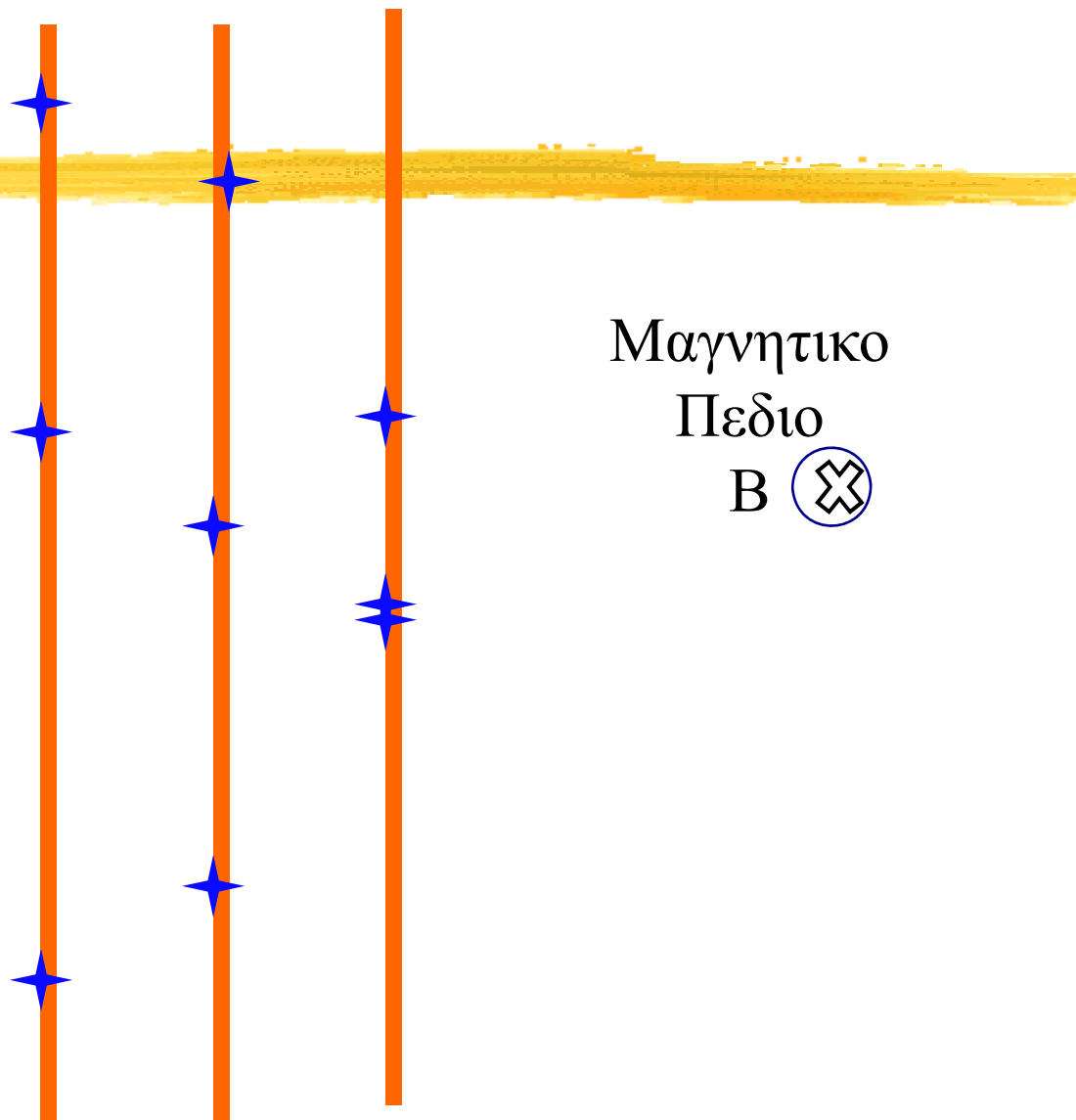
QuickTime™ and a decompressor are needed to see this picture.

# Τι είναι η ανακατασκευή?

- ⌘ Οι μετρησεις είναι σαν ένας γρίφος
  - ☒ Τι τροχιές τις προκάλεσαν?
- ⌘ Καθε μετρηση θεσης βοηθαι
  - ☒ Υπαρχουν ομως 100-αδες ως χιλιαδες μετρησεις
- ⌘ Η ανακατασκευη πρεπει να βρει τη λυση!
  - ☒ Ξεροντας καλα το μαγνητικο πεδιο
    - ☒ Βρισκουμε ποιες μετρησεις ανοικουν σε ποιες τροχιες

# Ανακατασκευή στην πράξη

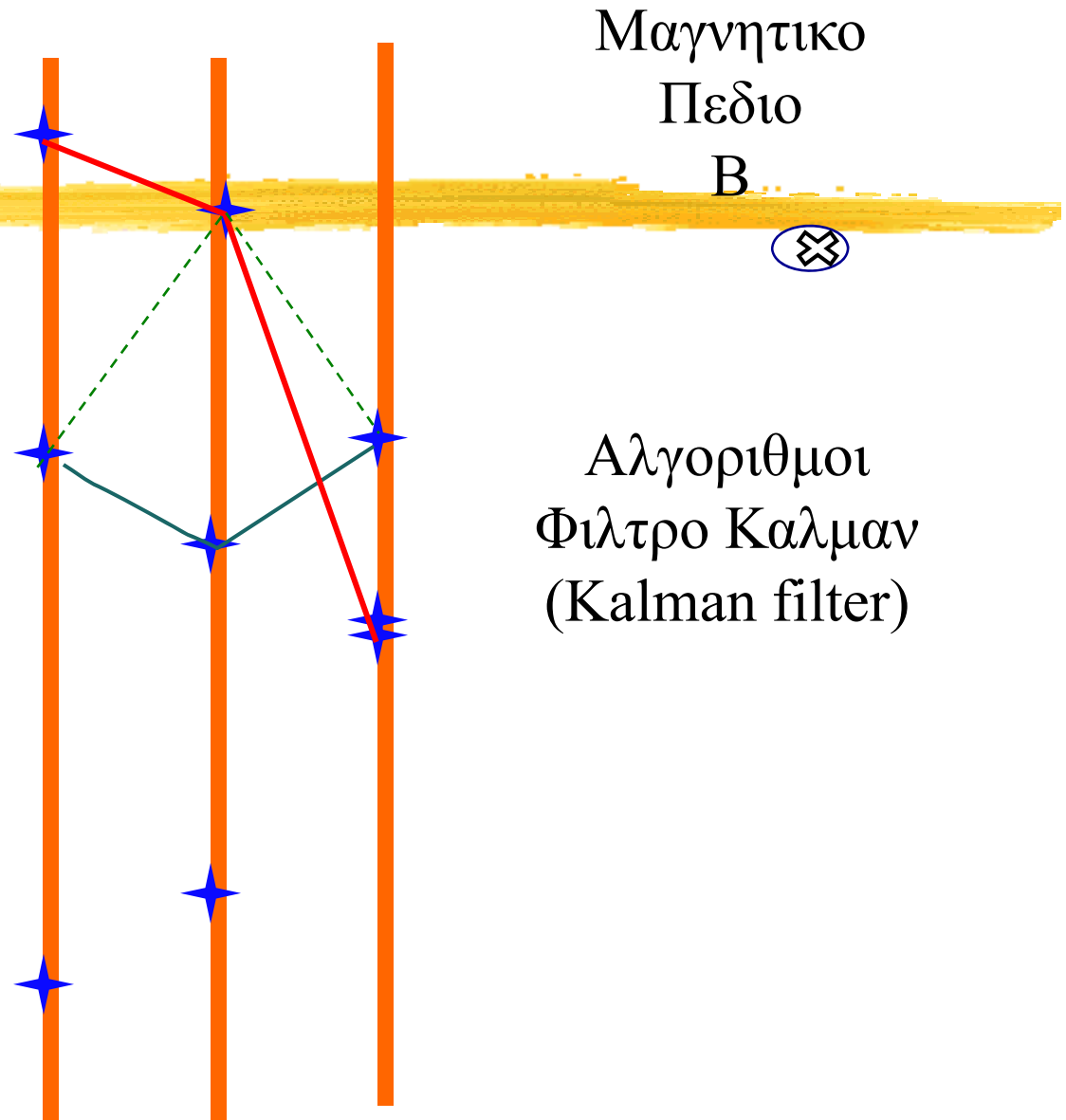
- Αρχίζει με τις θέσεις διαβάσης των σωματιδίων





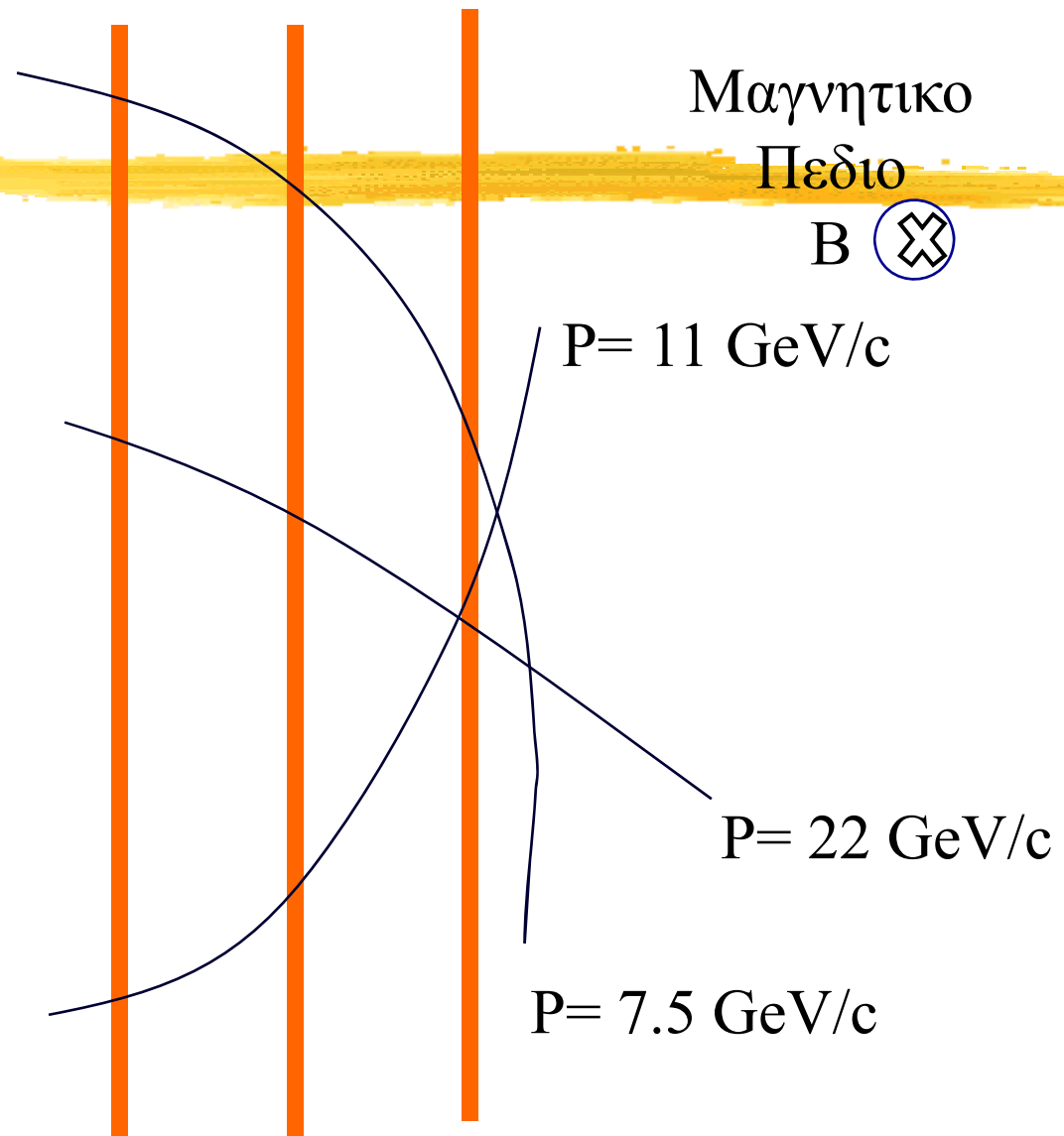
# Ανακατασκευή στην πράξη

- Αρχίζει με τις θέσεις διαβάσης των σωματιδίων
- Δωκιμάζονται διαφοροί συνδιασμοί
  - και υπολογίζεται η διαφορά μετρήσης-προβλεψής
  - Και έτσι πιθανότητα του καθε συνδιασμου
- 



# Ανακατασκευή: αποτέλεσμα

- Αρχίζει με τις θέσεις διαβάσης των σωματιδίων
- Δωκιμαζονται διαφοροι συνδιασμοι
  - και υπολογιζεται η διαφορα μετρησης-προβλεψης
  - Και ετσι πιθανοτητα του καθε συνδιασμου
- Τελικα εχουν βρεθει ολες οι τροχιες
  - ή «στα γρηγορα» αυτες με μεγαλη ορμη- οι κυριες τροχιες



# Προσομοίωση και Ανιχνευτές



Τι είναι η προσομοίωση

Γιατί υπάρχει

Πώς γίνεται

# Οι σημερινοι ανιχνευτες

## ⌘ Πολλα τμηματα

- ☒ Διαφορετικες αναγκες

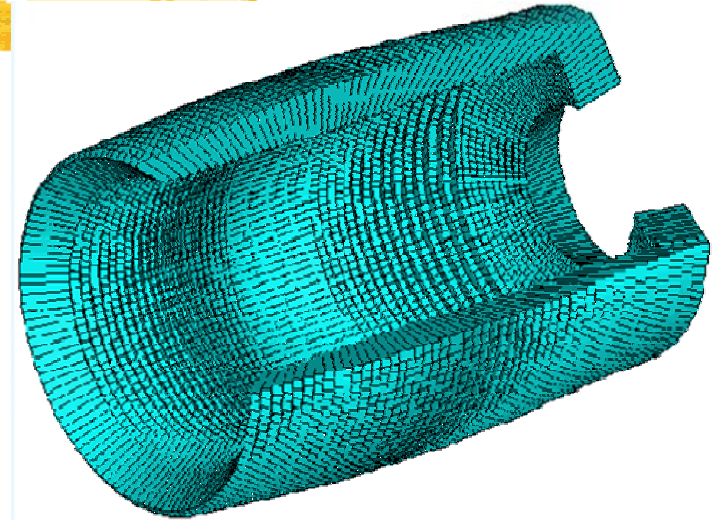
  - ☒ Μετρηση θεσης (τρακερ - trackers)

  - ☒ Μετρηση ενεργειας (θερμιδομετρα)

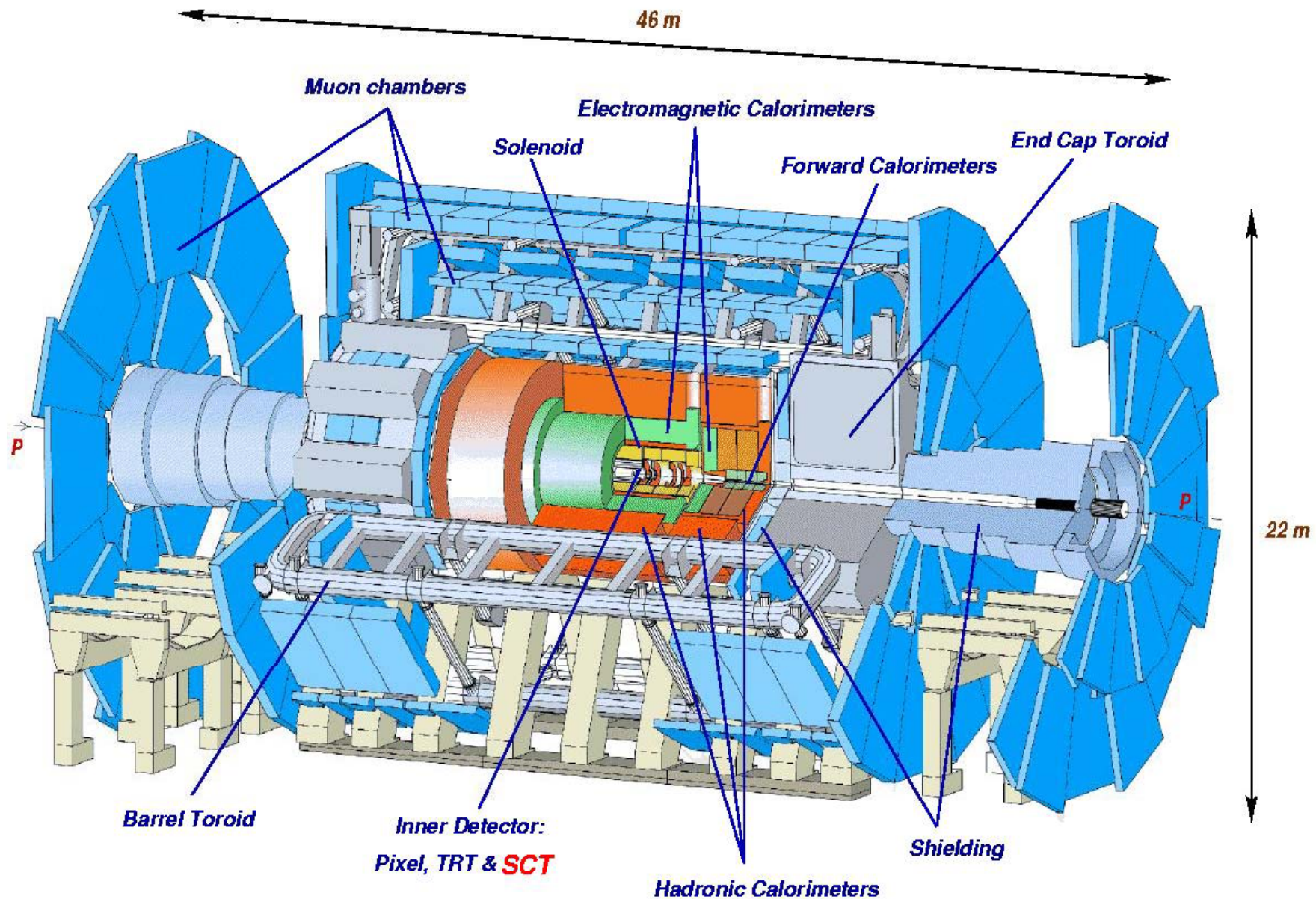
## ⌘ Λογω της πολυ-πλοκοτητας

- ☒ οι πιο πολλες μελετες

  - χρειαζονται πολλα υπολογιστικα εργαλεια



# Οι ανιχνευτές του ΑΤΛΑΣ



# Τι είναι προσομοίωση?

⌘ Φτιαχνουμε μοντελα

☑ Του ανιχνευτη

☑ Γεωμετρια

☑ Υλικα

☑ Των αλληλεπιδρασεων

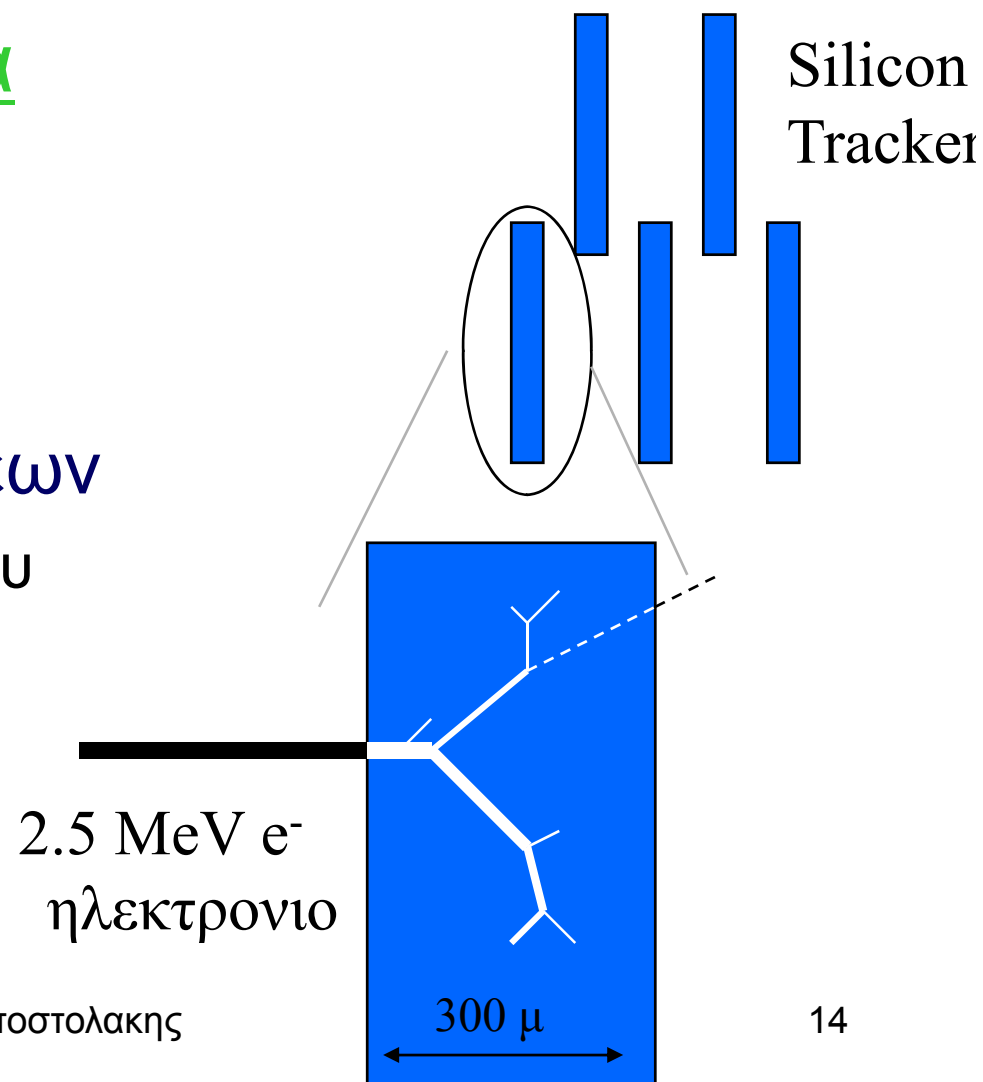
☑ Καθε γνωστου τυπου

- Ηλεκτρομαγνητικου
- Υσχηρου πυρηνικου

$$\sigma_{\text{συνολο}} = \sum \sigma_{\text{φαινομενου}}$$

27 Ιουνιου 2008

Ι. Αποστολακης



14

# Γεωμετρία ενός ανιχνευτή

⌘ EM  
Calorimeter

⊠ Crystal

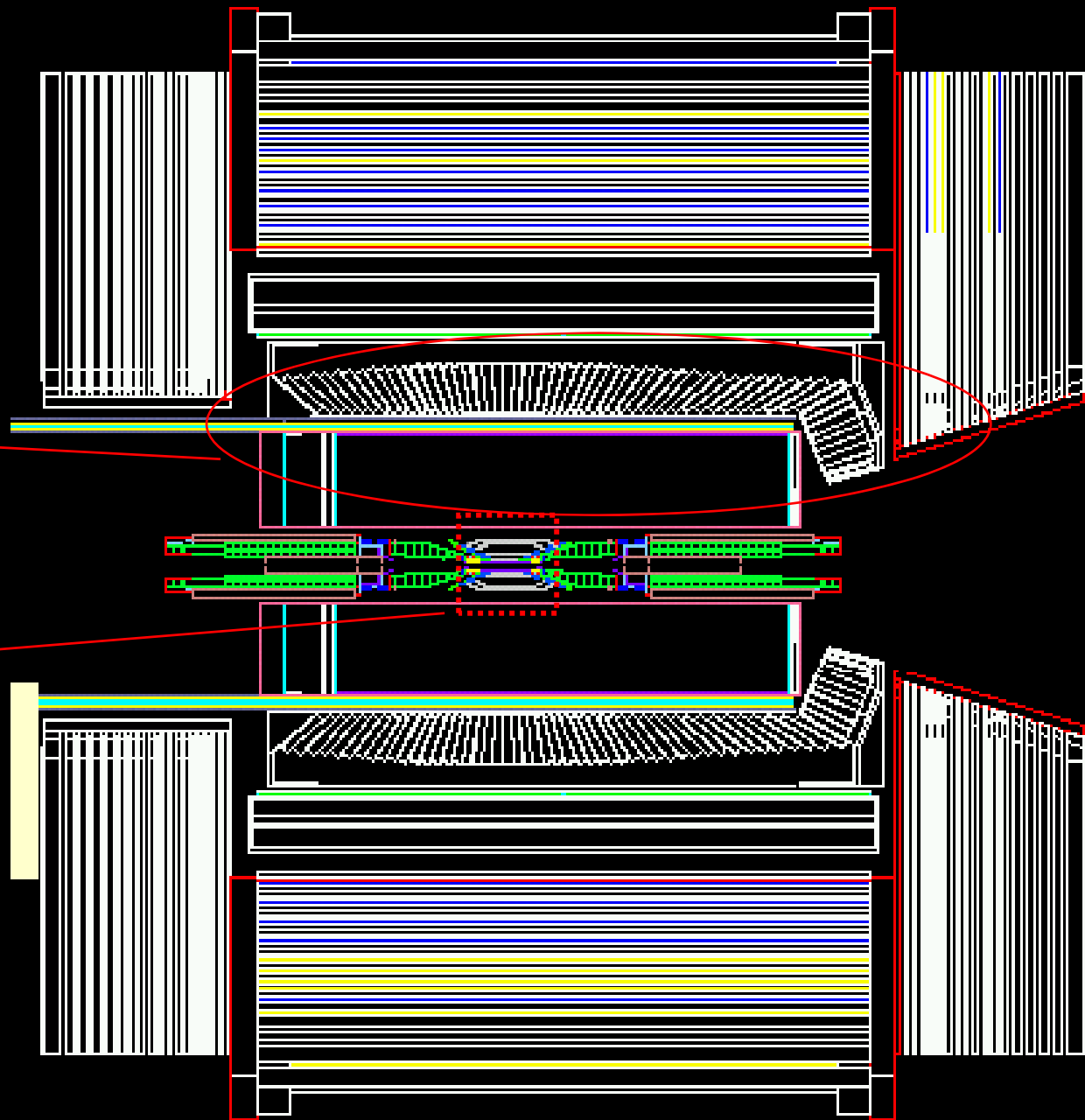
⌘ Tracker

⊠ Precision

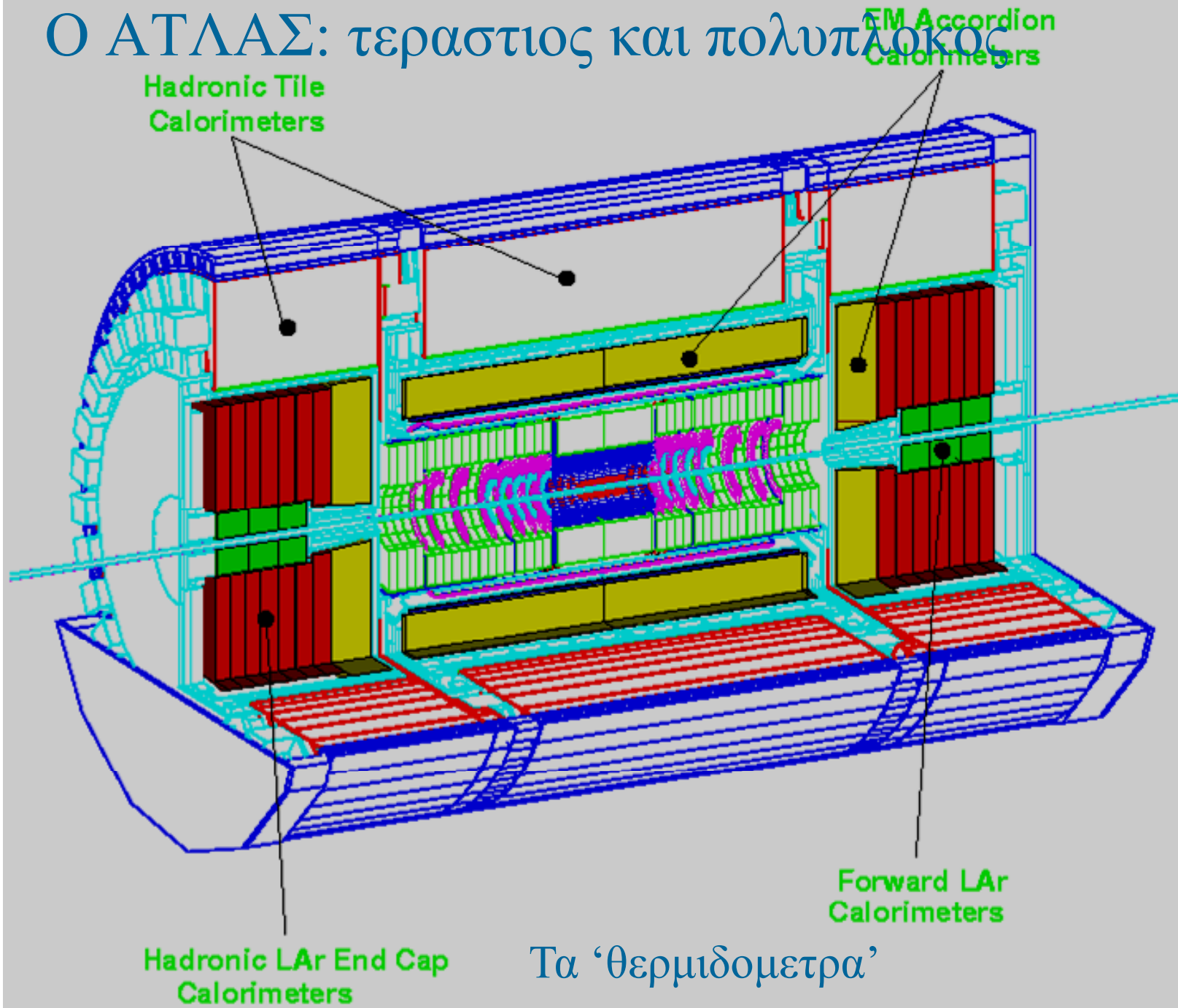
⊠ Drift  
Chamber

BABAR  
(SLAC, US)

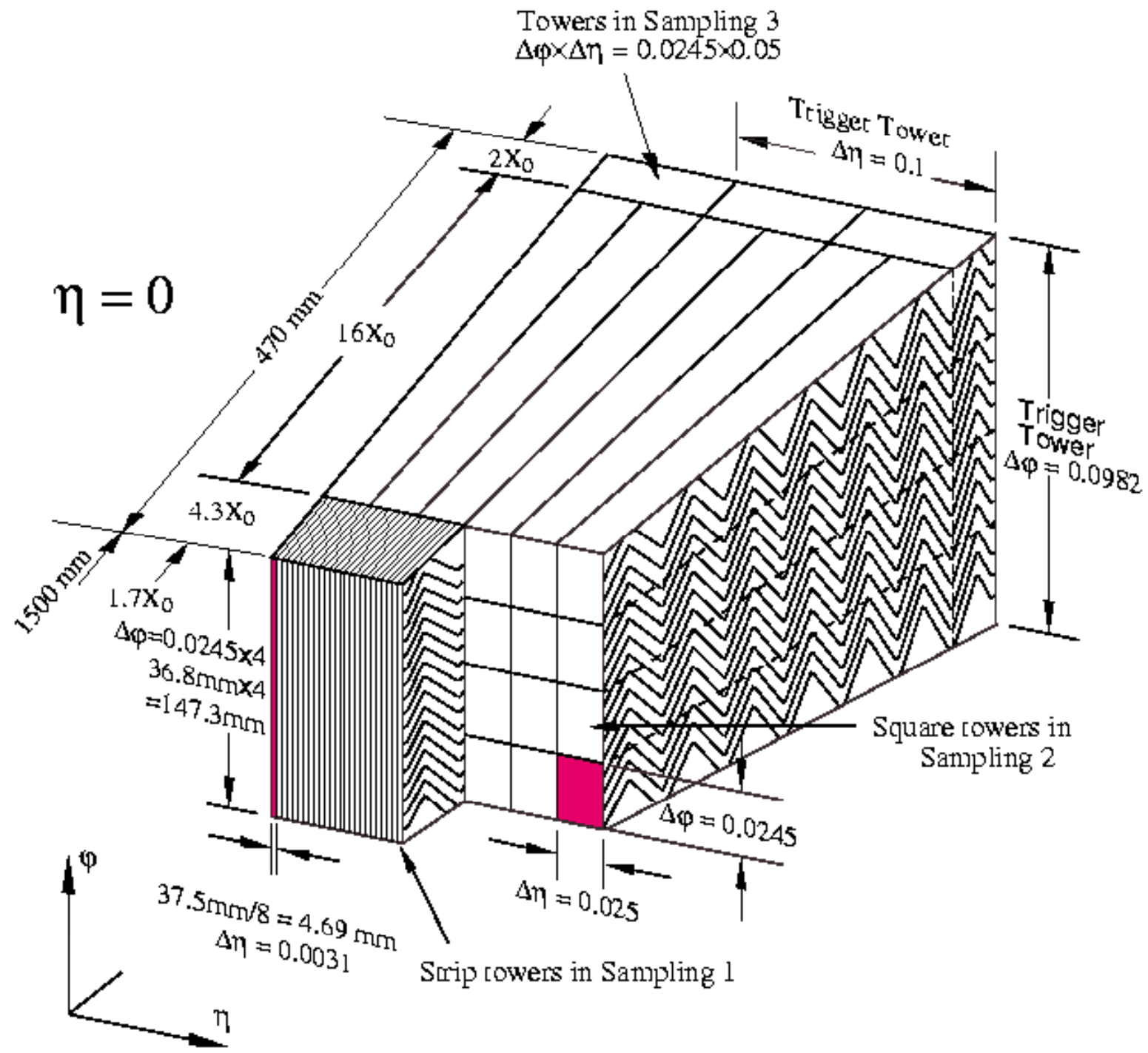
27 louviou



# Ο ΑΤΛΑΣ: τεραστιος και πολυπλοκος



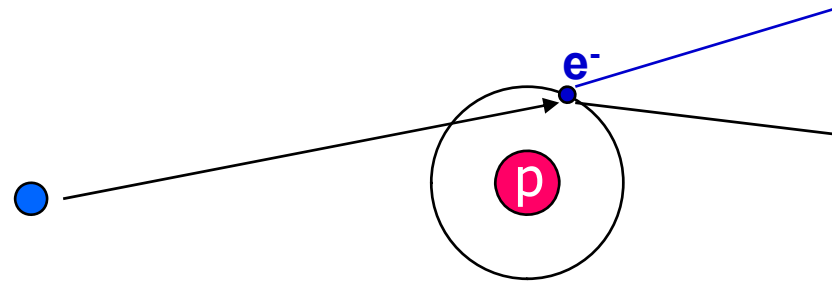




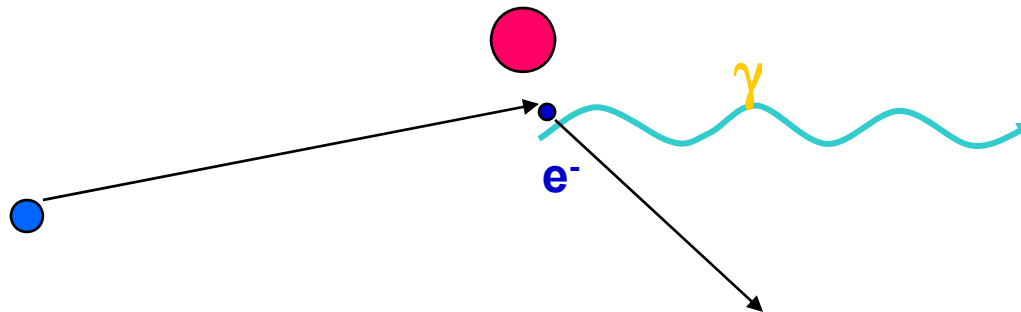
# Βασικές Αλληλεπιδρασεις

- ⌘ Οι διαφορες αλληλεπίδρασης του σωματιδίου με το υλικό (τμήμα του ανιχνευτή η αλλο)
  - ☑ παραγωγή δευτερευοντος σωματιδιου

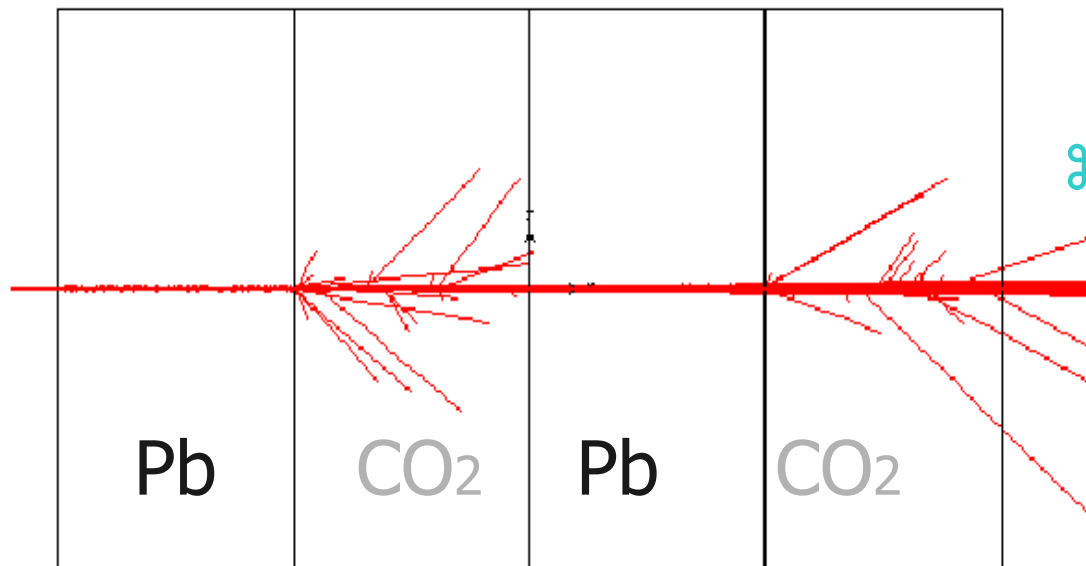
☑ Ιοντισμός



☑ Bremsstrahlung

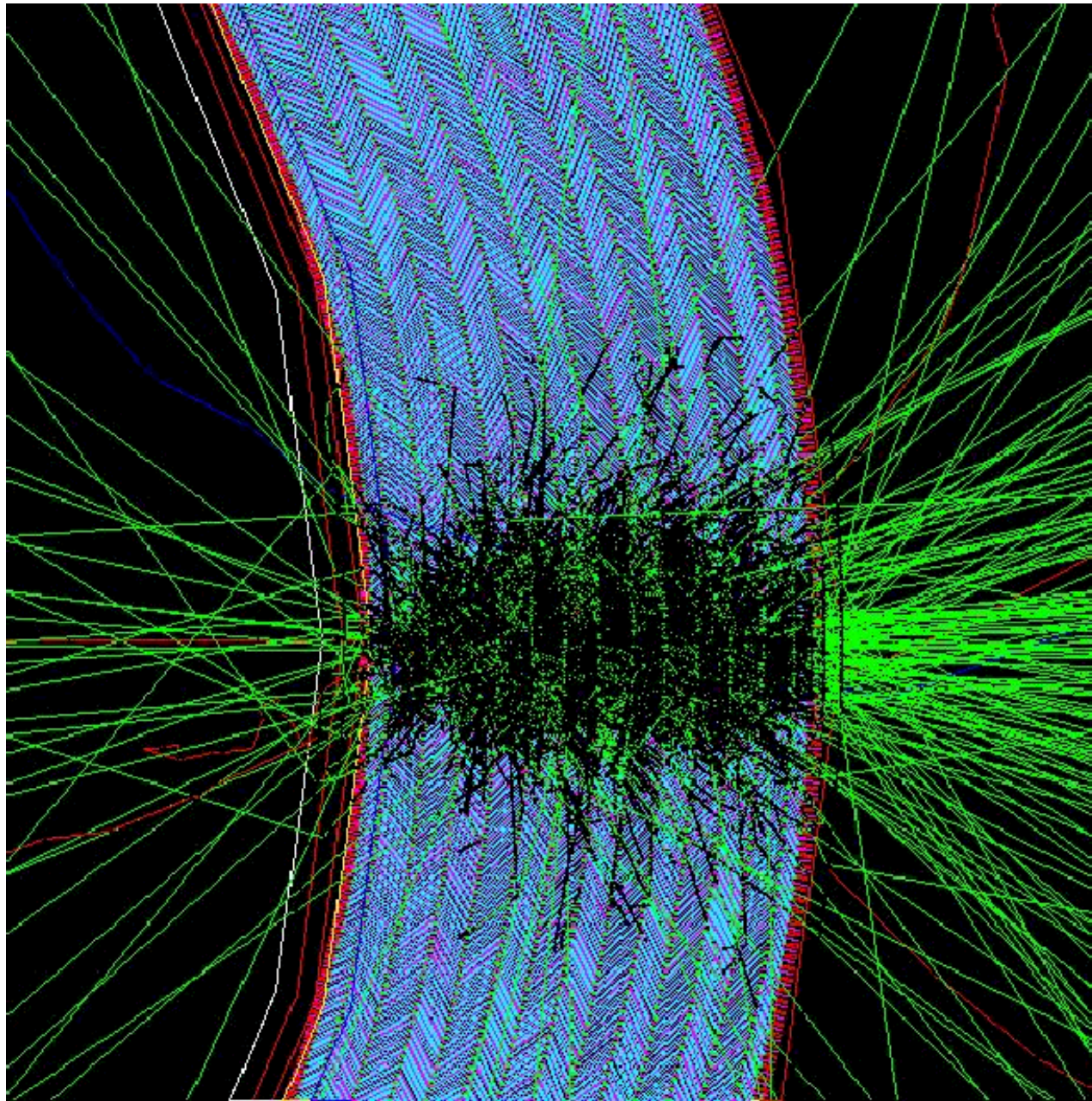


# Ενα απλο παραδειγμα



- ⌘ Στο μολυβδο παραγονται πιο πολλα δευτερευοντα σωματιδια,
  - ☒ αλλα δεν πανε μακρια
- ⌘ Το διοξειδιο του ανθρακα, σαν αεριο, εχει μικρη πυκνοτητα
  - ☒ Οσα σωματιδια φτανουν η παραγωνται, πανε μακρια
  - ☒ Παραγονται λιγοτερα

GEANT 3



# Γιατι προσομοιωση ?



⌘ Για να σχεδιασουμε  
τους ανιχνευτες

⌘ Για να ετοιμασουμε  
τις μεθοδους  
ανακατασκευης

⌘ Για να καταλαβου-με  
τον ανιχνευτη

# Υπαρχει απλη λυση ?

⌘ Αρκει η μεση τιμη της αποθεσης ενεργειας (π.χ.)  
?

☑ Για μερικες απλες ερευνες, Ναι

☑ Για πολλες (τις περισσοτερες) χρειαζεται ολη η εικονα

☒ Υπαρχει κατανομή τιμων

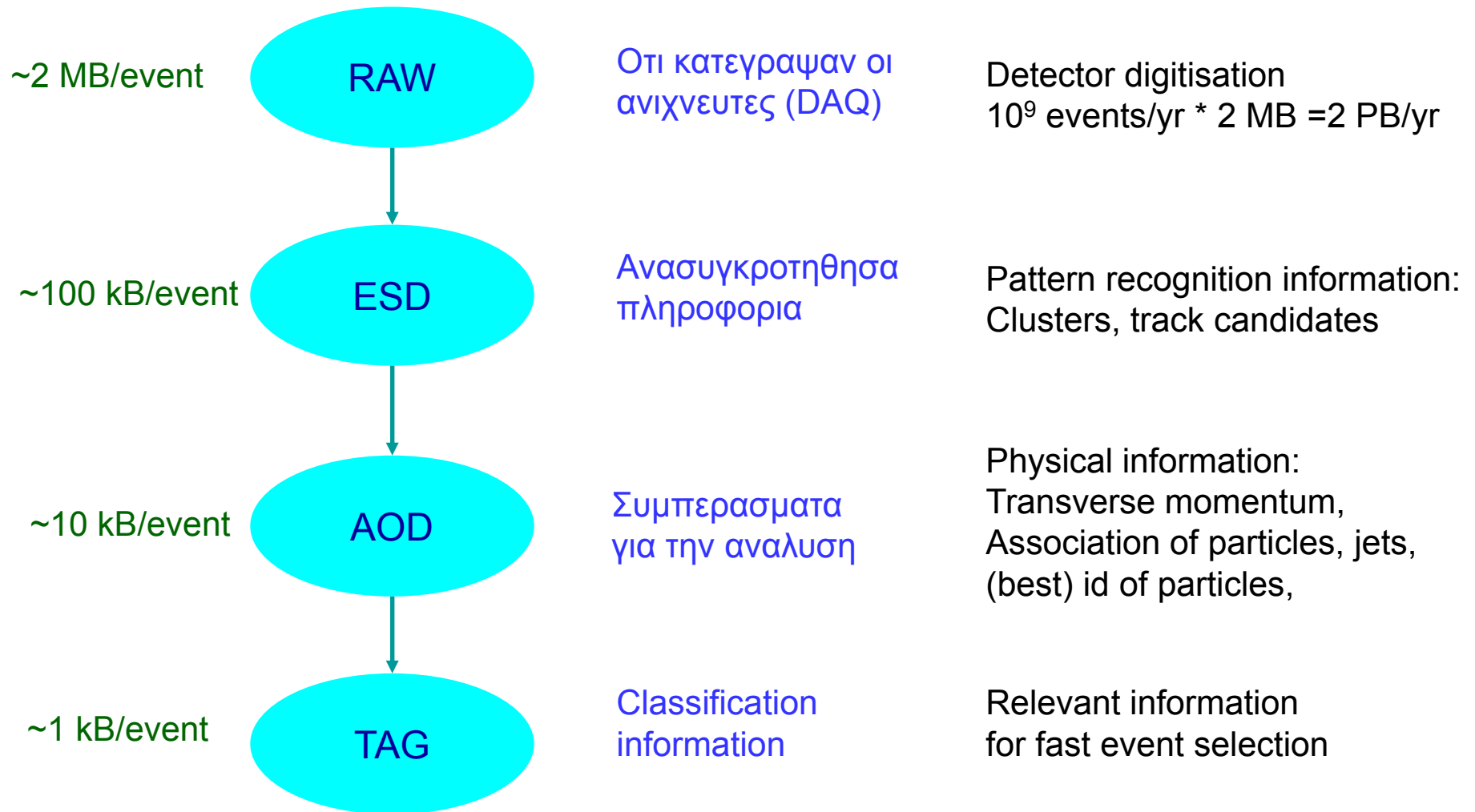
- Που δεν ειναι παντα συμμετρικη η απλη
- Οι ουρες των κατανομων μπορουν να παιξουν μεγαλο ρολο

⌘ Με ποιο σφαλμα ξερετε την ταδε ενεργεια?

# Αναλυση δεδομενων



# Ιεραρχία Δεδομενων (Data)



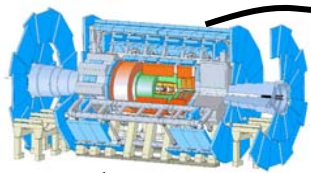


# Οι φυσικοί προγραμματίζουν

- ⌘ Όλη η ανάλυση μετρήσεων γίνεται με υπολογιστές
- ⌘ Οι φυσικοί ΥΕ χρησιμοποιούν ειδικευμένα προγράμματα
  - ☑ Μερικοί γράφουν μεγάλες ρουτίνες (routines/Fortran, methods/C++)
  - ☑ Οι πιο πολλοί κάνουν μικρές ρουτίνες, για τις δικές τους ανάγκες
- ⌘ Όλοι θα χρησιμοποιούσα τα ‘εργαλεία’ να δουν τις περιληψεις των μετρησεων



# Data Handling and Computation for Physics Analysis



detector

Φίλτρο γεγονότων  
(διασυλογη & ανακατασκευή)

Ανακατασκευή  
(reconstruction)

δεδομένα  
(Raw data)

event  
reprocessing

event  
summary  
data

processed  
data

Ανάλυση  
'Batch'

analysis

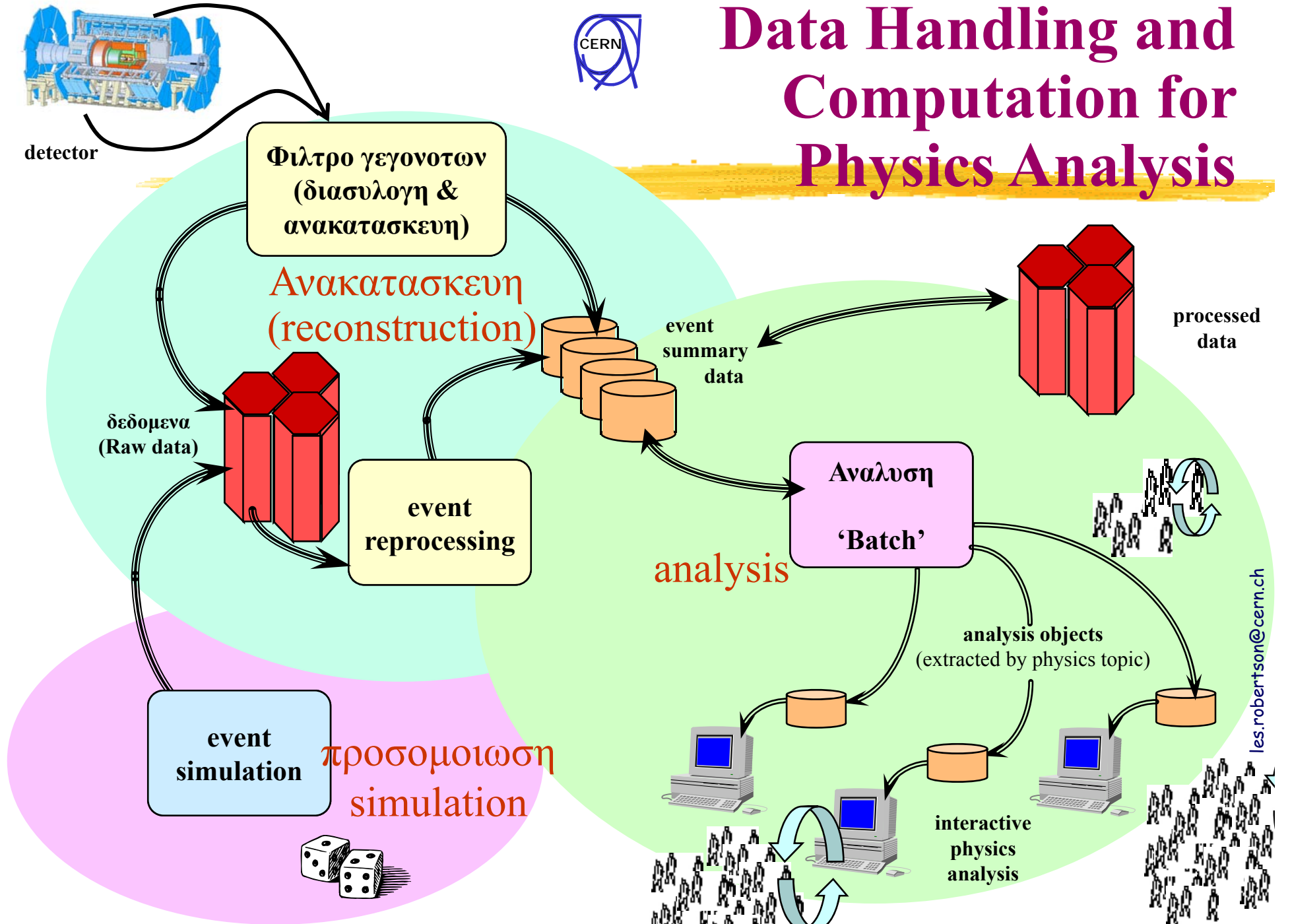
analysis objects  
(extracted by physics topic)

event  
simulation

προσομοίωση  
simulation

interactive  
physics  
analysis

les.robertson@cern.ch



# Computing Characteristics

⌘ Independent events (collisions of particles)

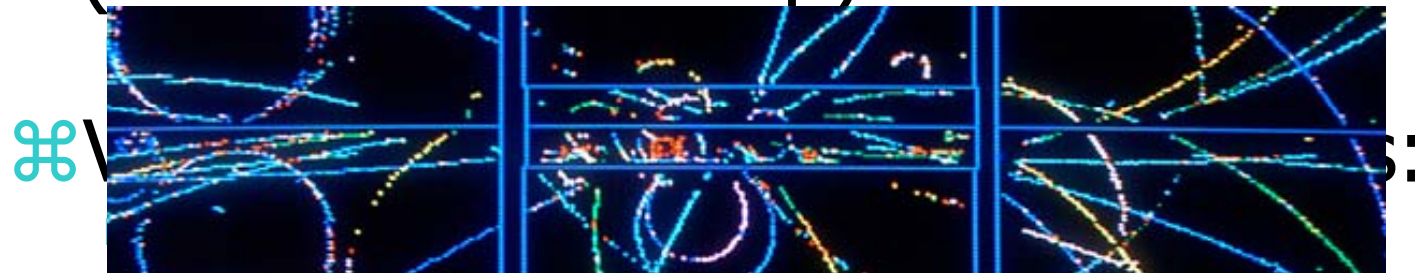
⊞ trivial (read: pleasant) parallel processing

⌘ Bulk of the data is read-only

⊞ versions rather than updates

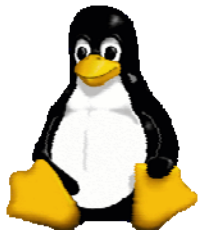
⌘ Meta-data in databases, but physics data in “flat” files

⌘ Compute power measured in SPECint (rather than SPECfp)

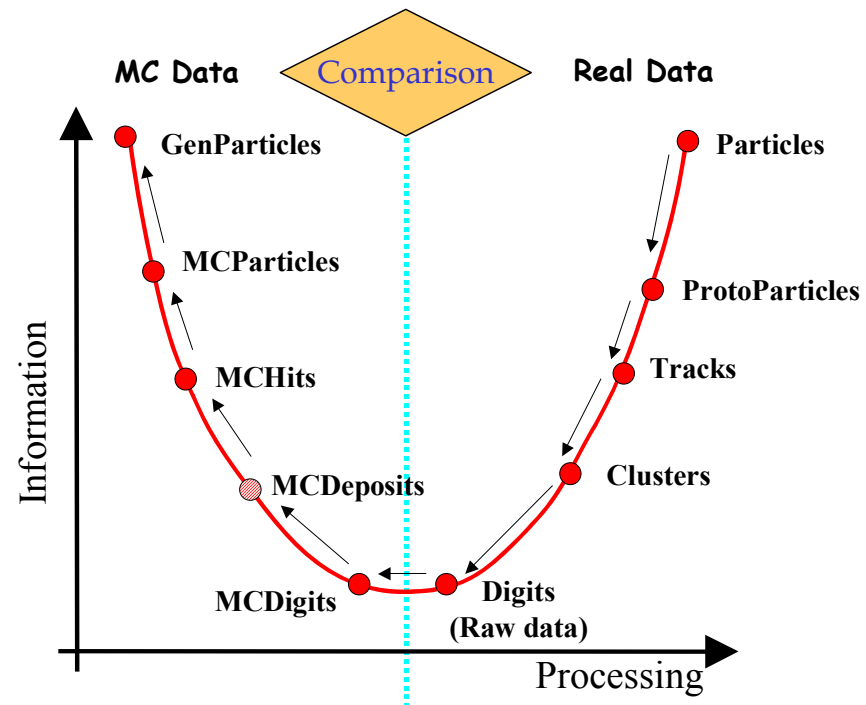


# The Computing Environment

- ⊞ Χρησιμοποιούμε τους υπολογιστές για το συνολικό δυναμικό τους
  - ⊗ Οχι τους πιο γρηγορους
  - ⊗ Συνηθεις – `της αγορας`
  
- ⊞ High-throughput computing
- ⊞ (based on reliable “commodity” technology)
  - ⊗ Around 7500 (dual-socket Xeon) PCs with “Scientific Linux”
    - Now typically also “dual-core”
    - Quad-core expected for next acquisition



# Data Organisation



***Backup***

A thick, horizontal yellow brushstroke underline that spans across the width of the slide, positioned below the word 'Backup'.

***More on simulation***

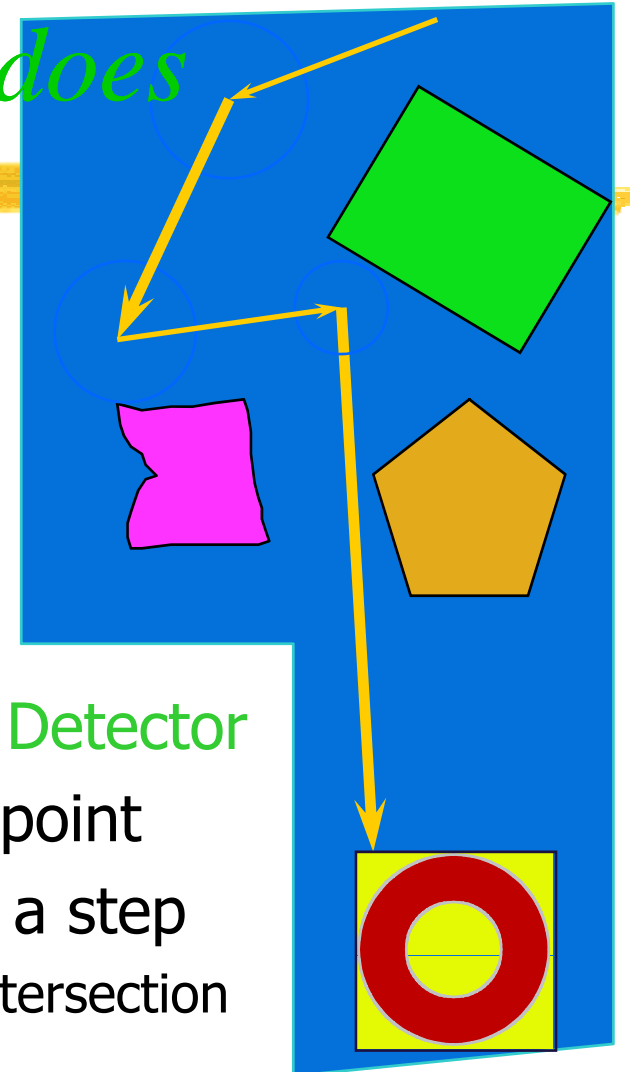
# Geant4 geometry: what it does

## Describes a Detector

- ⌘ Hierarchy of volumes
- ⌘ Many volumes repeat
  - ⌘ Volume & sub-tree
- ⌘ Up to millions of volumes for LHC era
- ⌘ Import detectors from CAD systems

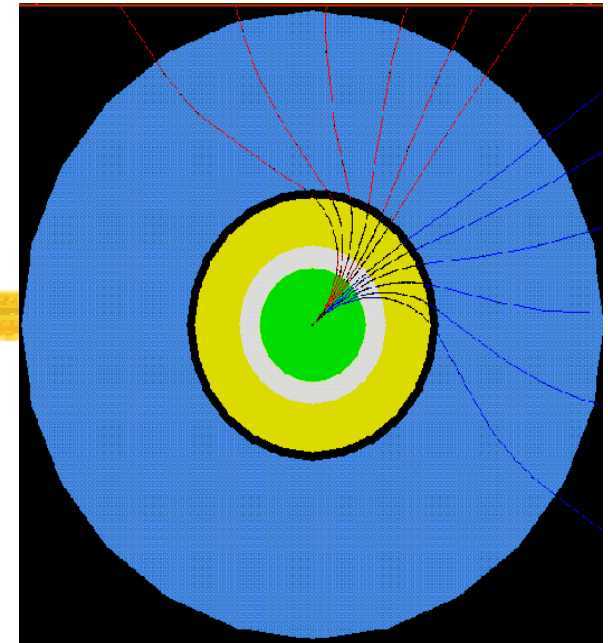
## Navigates in Detector

- ⌘ Locates a point
- ⌘ Computes a step
  - ⌘ Linear intersection



# ***Propagating in a field***

Charged particles follow paths that approximate their curved trajectories in an electromagnetic field.



⌘ It is possible to tailor

- ⌘ the accuracy of the splitting of the curve into linear segments,
- ⌘ the accuracy in intersecting each volume boundaries.

⌘ These can be set now to different values for a single volume or for a hierarchy.



# *Electromagnetic physics*



## ⌘ Gammas:

- ⊞ Gamma-conversion, Compton scattering, Photo-electric effect

## ⌘ Leptons( $e$ , $\mu$ ), charged hadrons, ions

- ⊞ Energy loss (Ionisation, Bremsstrahlung) or PAI model energy loss, Multiple scattering, Transition radiation, Synchrotron radiation,

## ⌘ Photons:

- ⊞ Cerenkov, Rayleigh, Reflection, Refraction, Absorption, Scintillation

## ⌘ High energy muons and lepton-hadron interactions

## ⌘ Alternative implementation (“low energy”)

- ⊞ for applications that need to go below 1 KeV

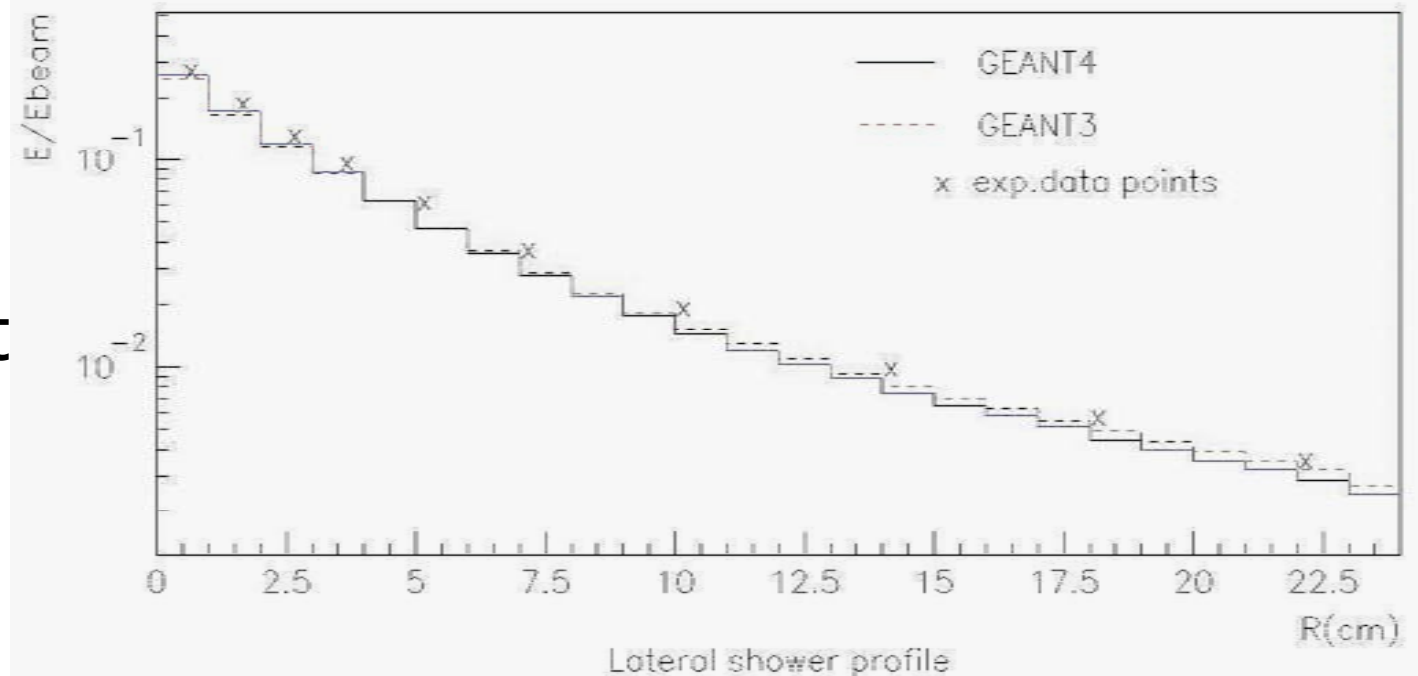
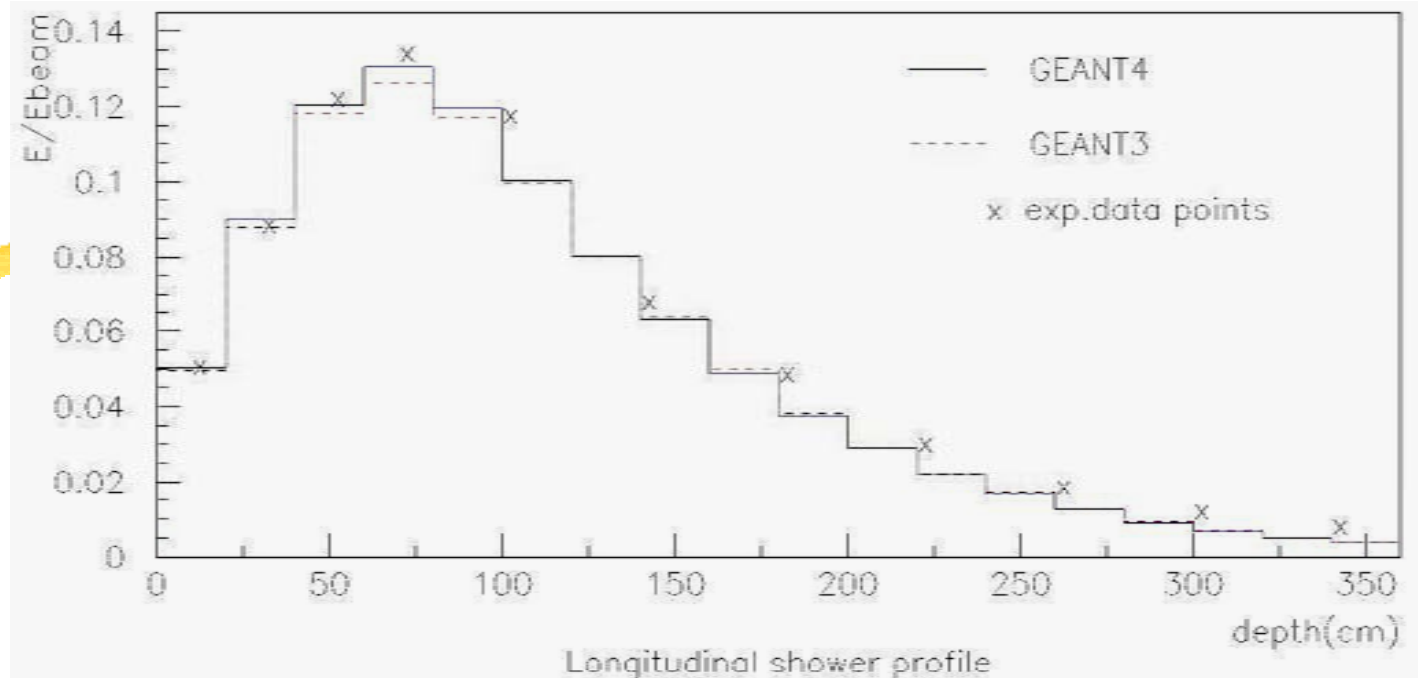
# Shower profile

1 GeV  
electron  
in H<sub>2</sub>O

G4,  
Data  
G3

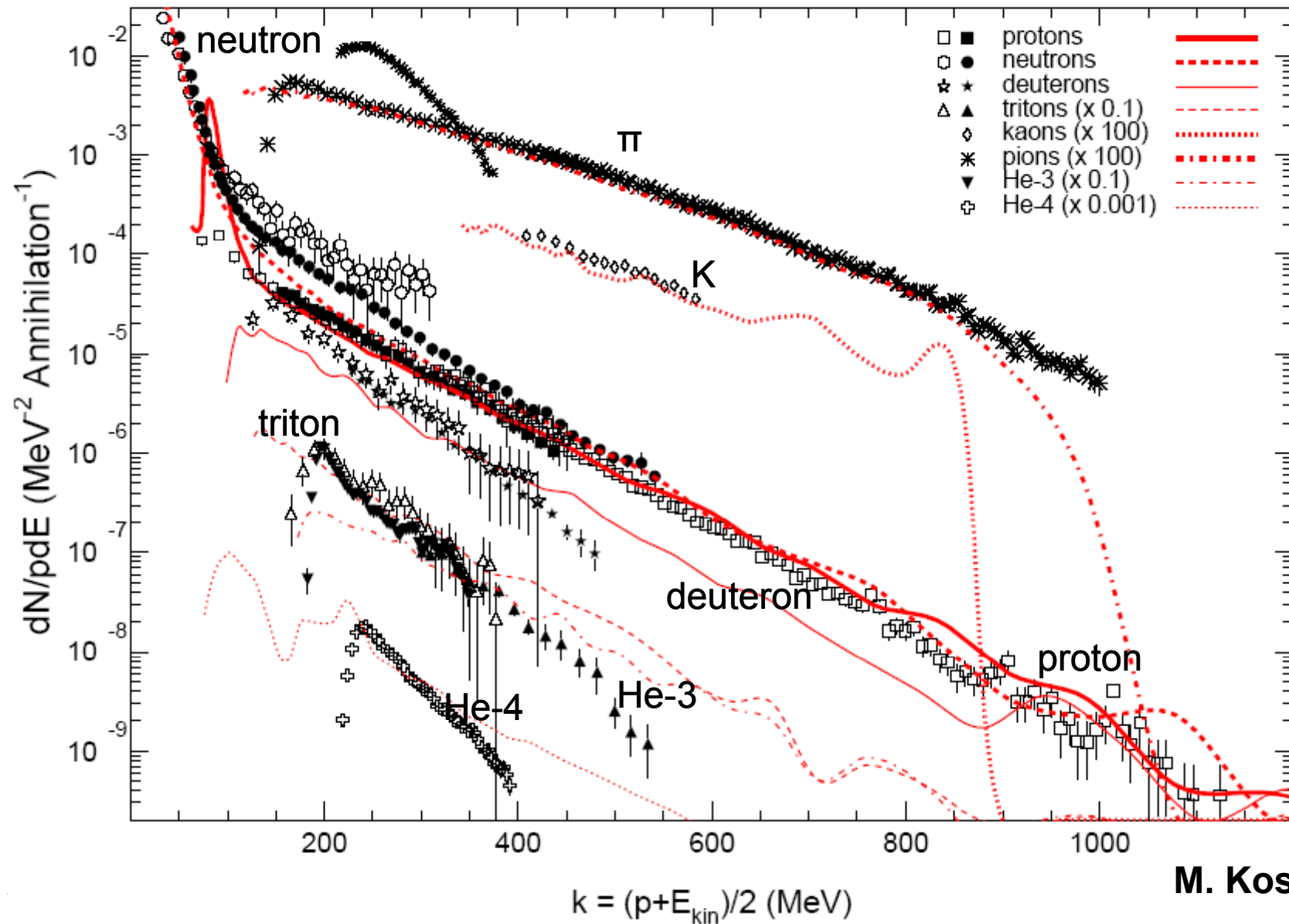
⌘ Good  
agreement  
seen with  
the data

27 louviou 2008



# Antiproton annihilation - CHIPS Model

Antiproton annihilation on  $^{238}\text{U}$  nucleus



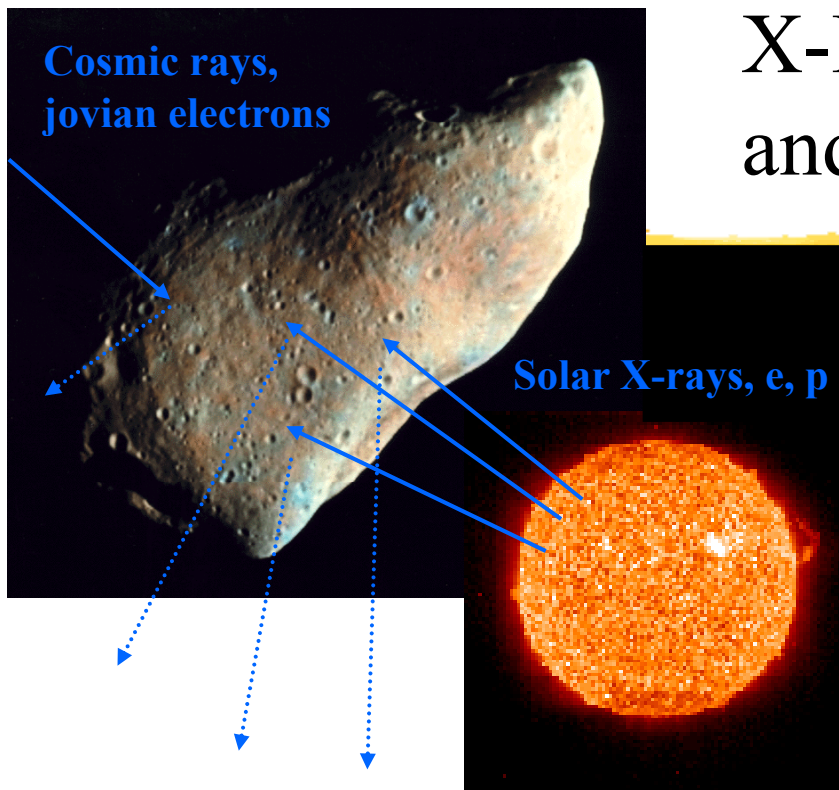
M. Kossov

# *Simulation 'packages'*



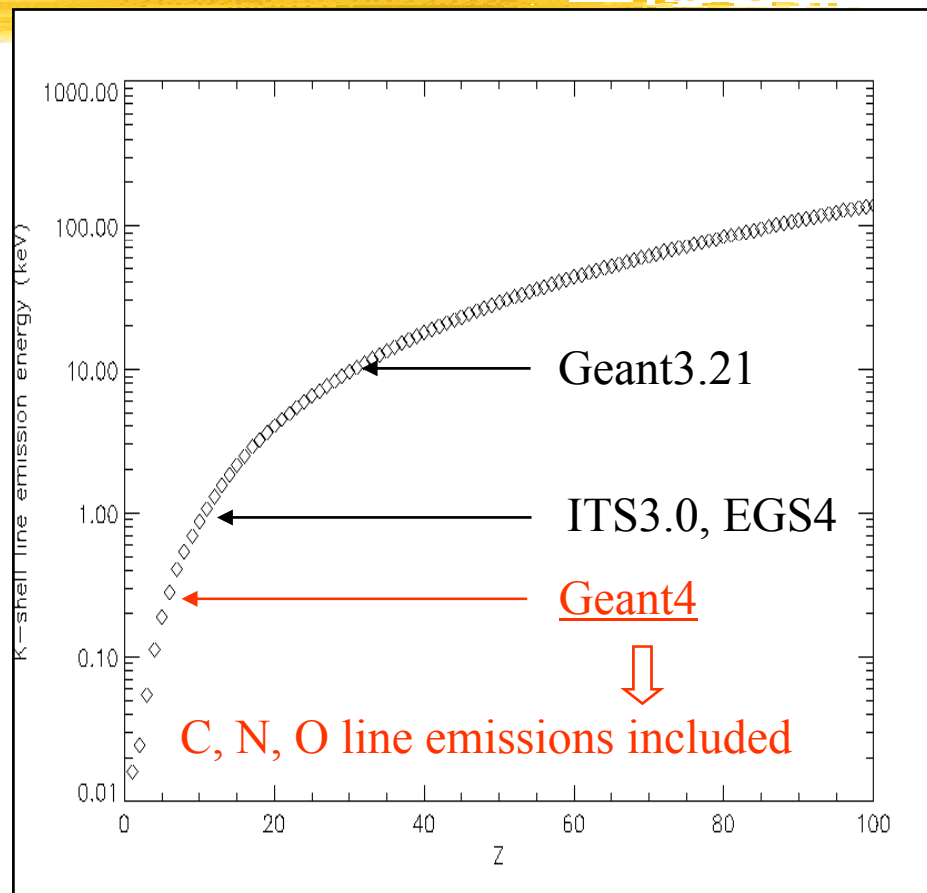
- ⌘ Provides the means to simulate
  - ☒ the **physical processes** and
  - ☒ **detector response** of an experiment.
- ⌘ As was realised by many in the past,
  - ☒ **most of the parts** needed can be **common** between experiments (eg physics, geometry blocks) .
- ⌘ So it makes eminent sense to create and use a **general purpose package**
  - ☒ That includes the common parts,
  - ☒ And enables an experiment to describe those parts with are specific to it.

# X-Ray Surveys of Asteroids and Moons

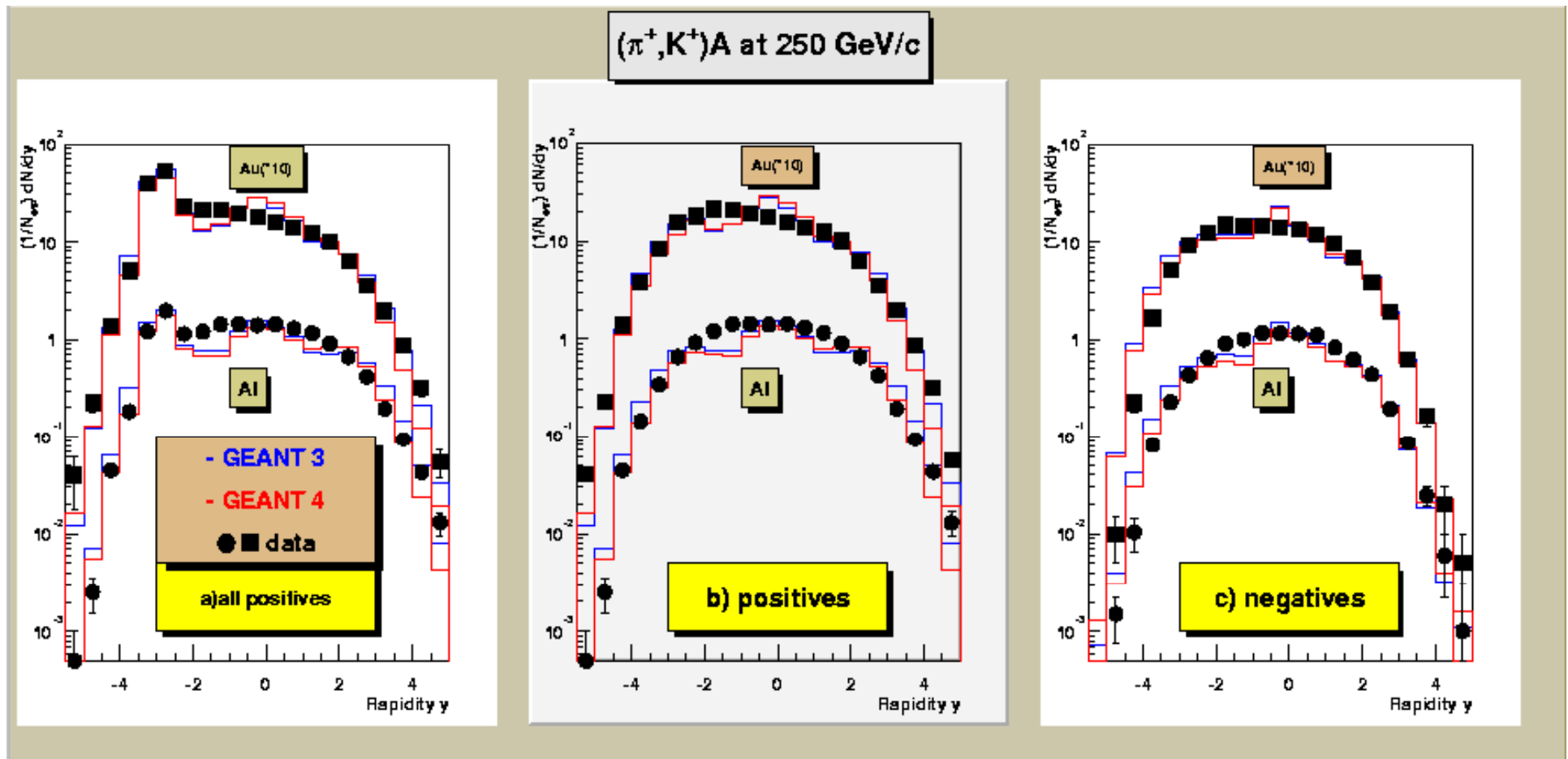


Courtesy SOHO EIT

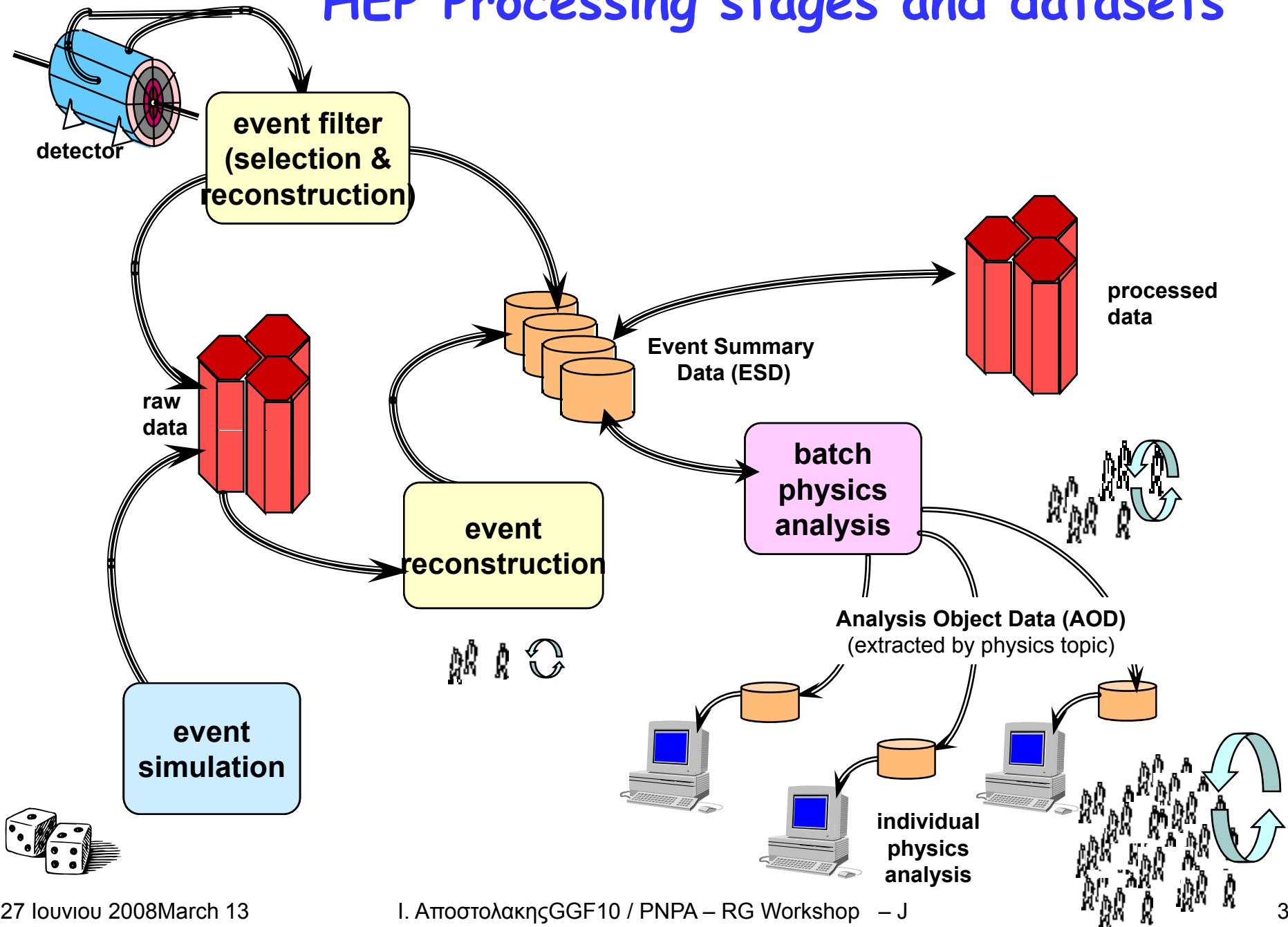
Induced X-ray line emission:  
indicator of target composition  
(~100  $\mu\text{m}$  surface layer)



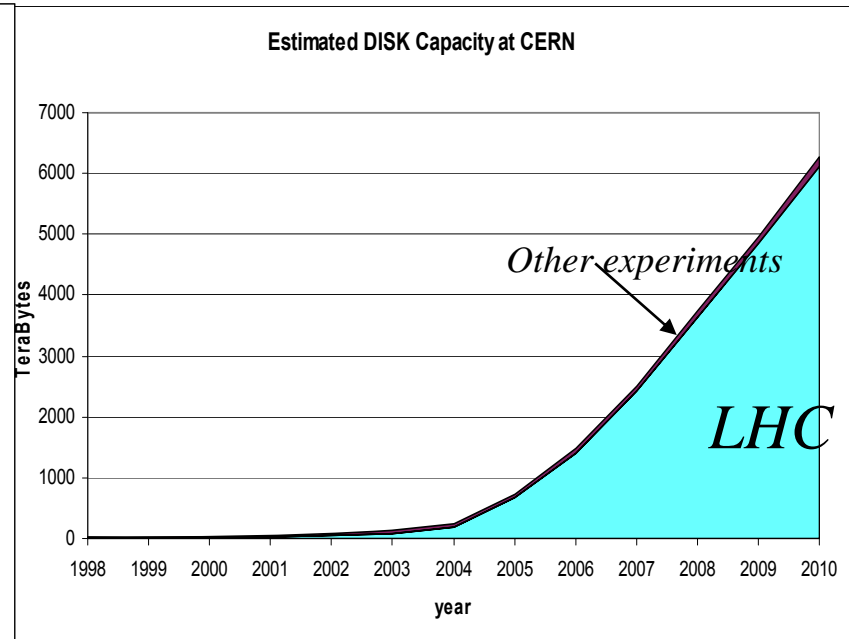
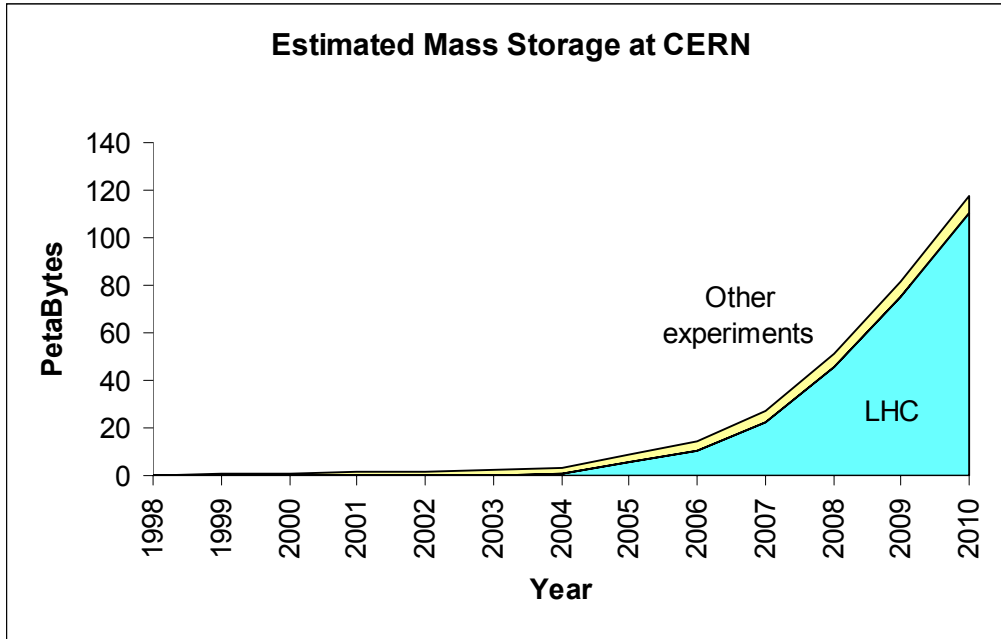
# High Energy $K, \pi$ on Al, Au



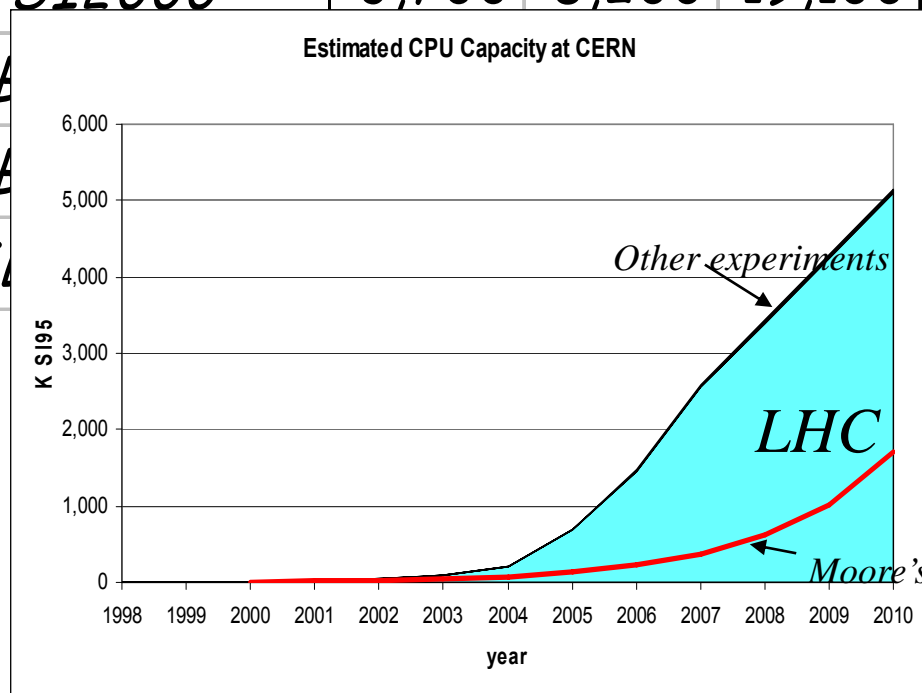
# HEP Processing stages and datasets



# CERN Centre Capacity Requirements for all

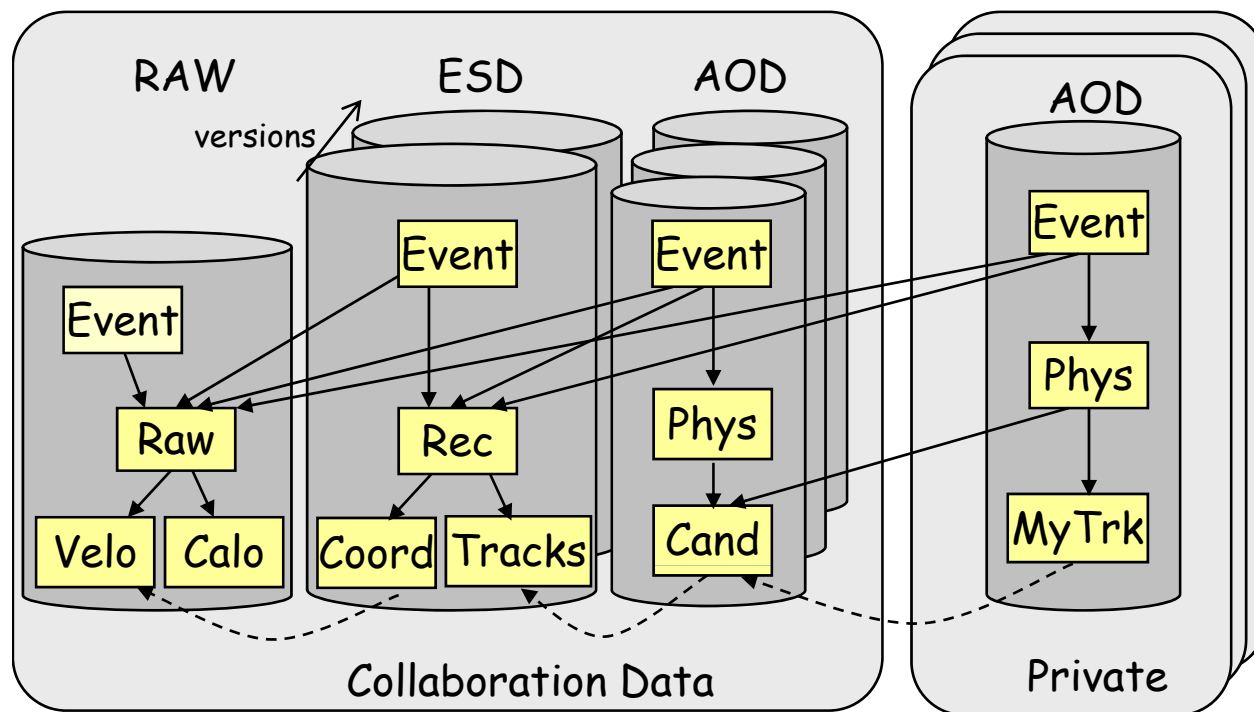


processing	K \$12000	3,700	8,200	19,100	29,000	34,000
disk	PA				5.0	6.7
tape media	PA				36	48
tape I/O	G				3.9	3.9



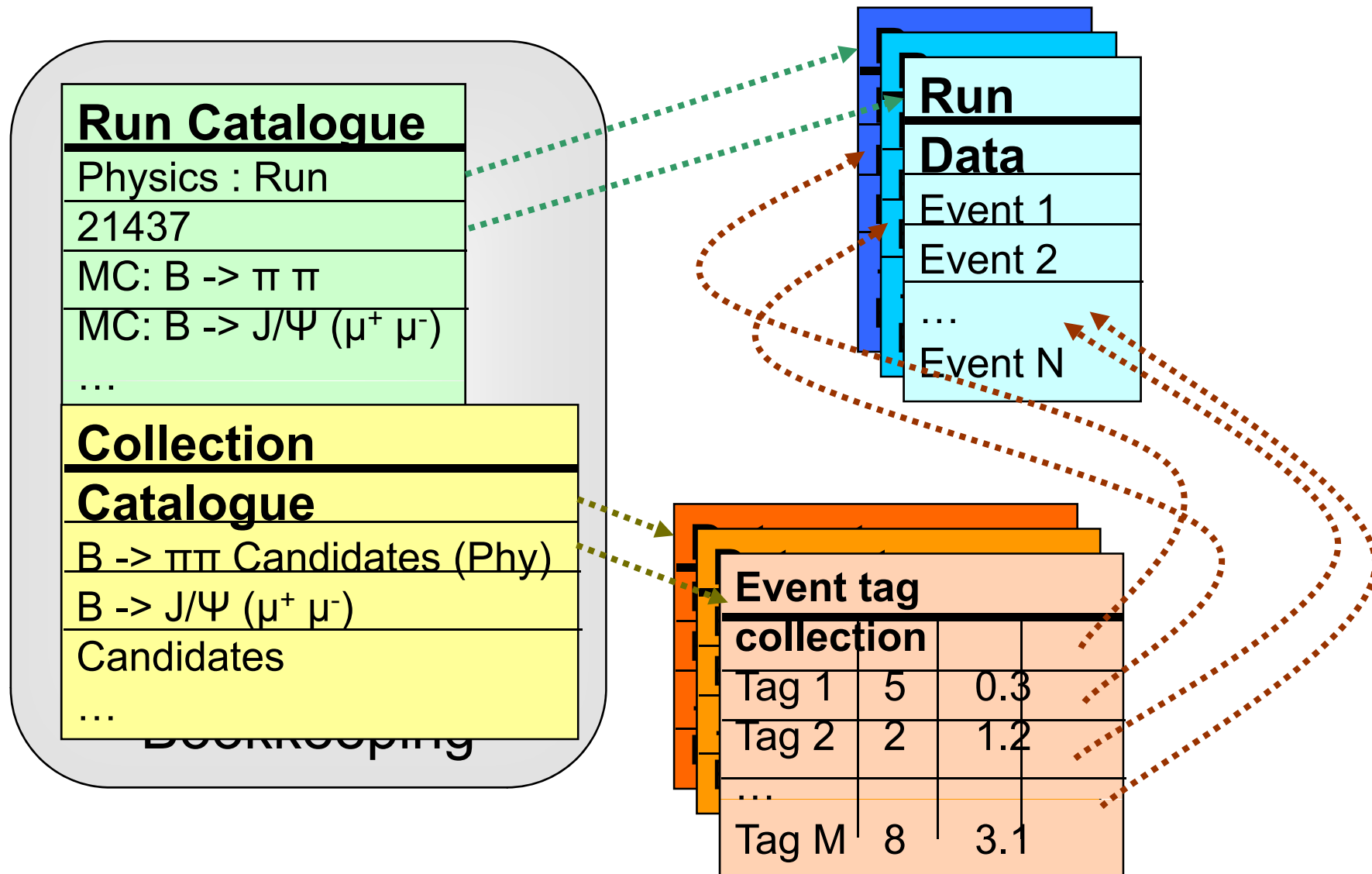


# Event Data



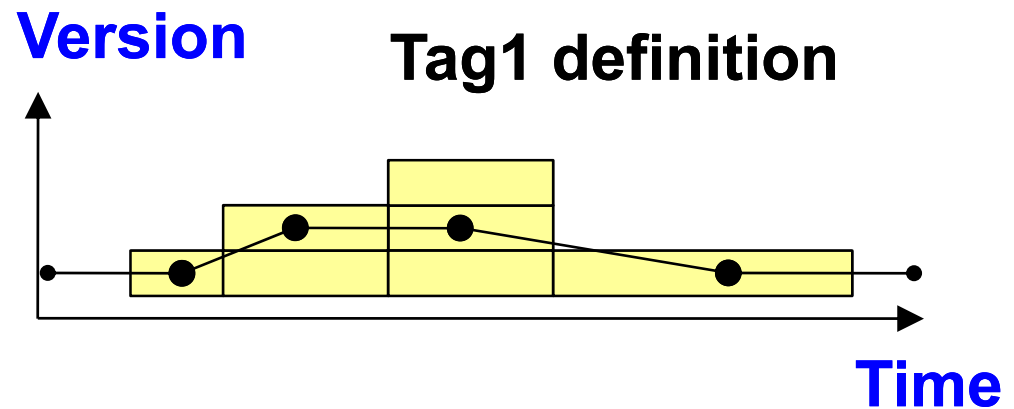
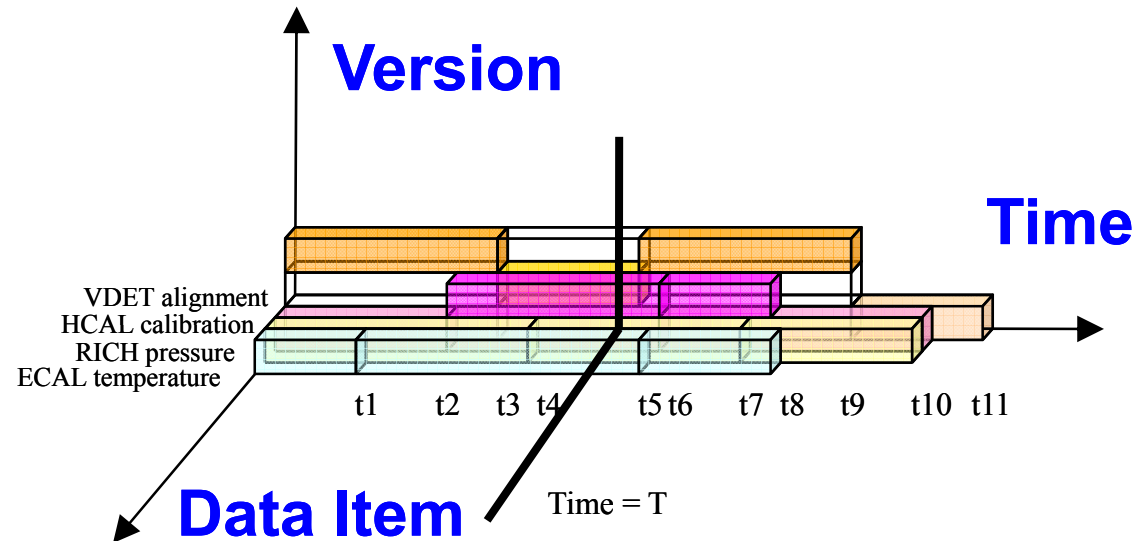
- ❑ Complex data models
  - ~500 structure types
- ❑ References to describe relationships between event objects
  - unidirectional
- ❑ Need to support transparent navigation
- ❑ Need ultimate resolution on selected events
  - need to run specialised algorithms
  - work interactively
- ❑ Not affordable if uncontrolled

# HEP Metadata - Event Collections

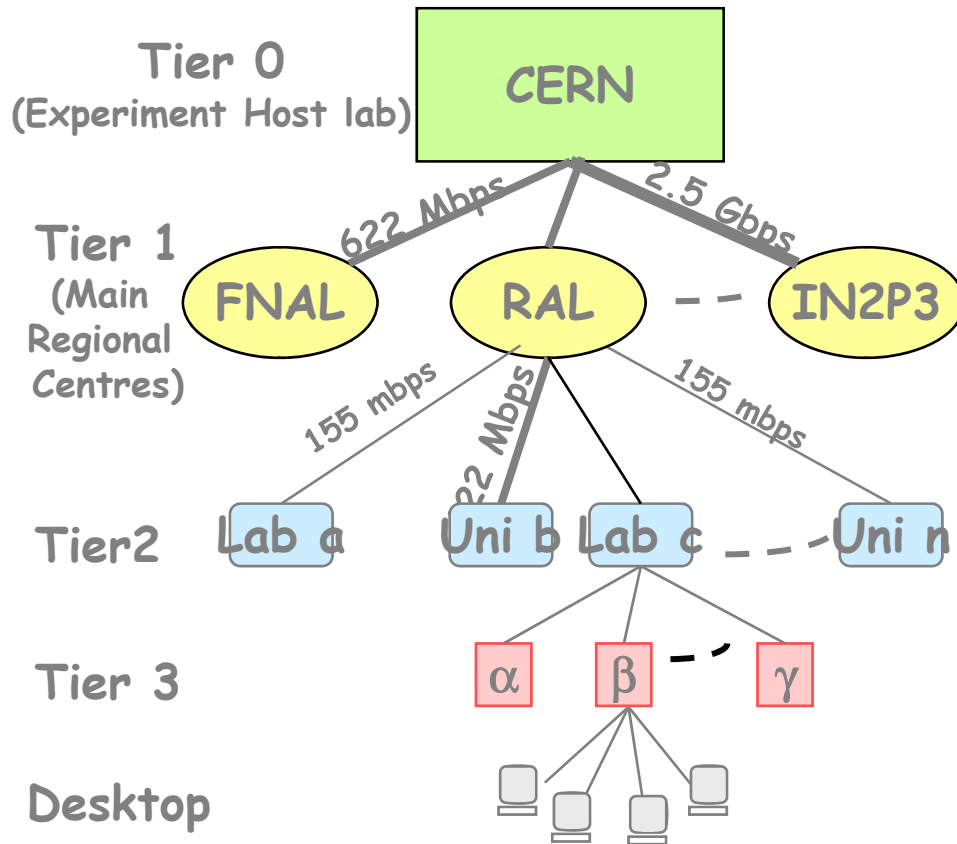


# Detector Conditions Data

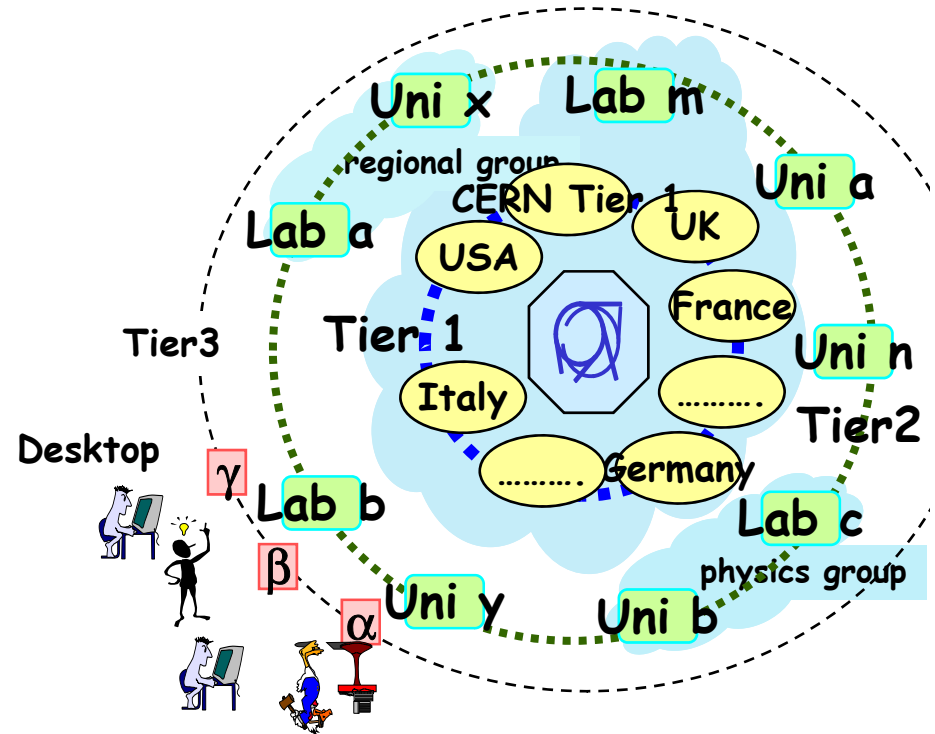
- ❑ Reflects changes in state of the detector with time
- ❑ Event Data cannot be reconstructed or analyzed without it
- ❑ Versioning
- ❑ Tagging
- ❑ Ability to extract slices of data required to run with job
- ❑ Long life-time



# A Multi-Tier Computing Model



Manager View



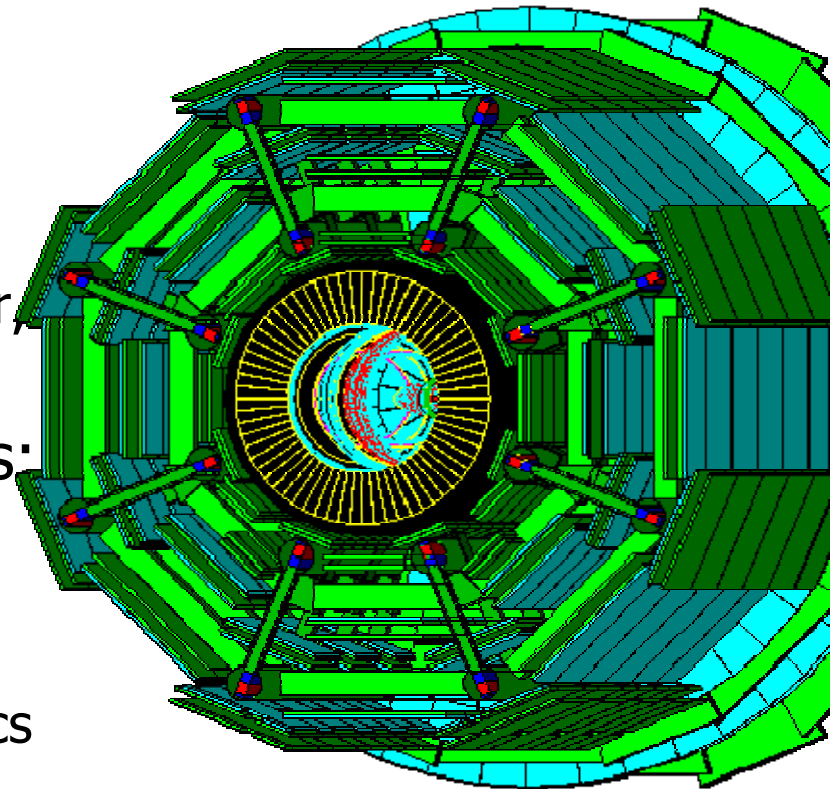
User View

# Distributed Analysis - the real challenge

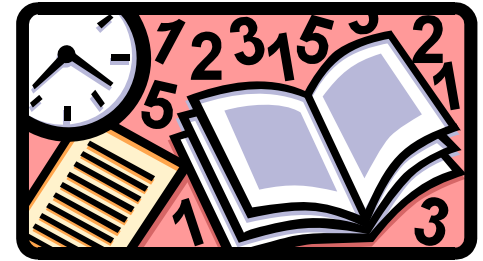
- ❑ Analysis will be performed with a mix of "official" experiment software and private user code
  - How can we make sure that the user code can execute and provide a correct result wherever it "lands"?
- ❑ Input datasets not necessarily known a-priori
- ❑ Possibly very sparse data access pattern when only a very few events match the query
- ❑ Large number of people submitting jobs concurrently and in an uncoordinated fashion resulting into a chaotic workload
- ❑ Wide range of user expertise
- ❑ Need for interactivity - requirements on system response time rather than throughput
- ❑ Ability to "suspend" an interactive session and resume it later, in a different location
- ❑ Need a continuous dialogue between developers and users

# Visualization

- ⌘ Much functionality is implemented
- ⌘ Several drivers:
  - ☑ OpenGL, *VRML*, Open Inventor, Opacs, *DAWN renderer (G4)*
- ⌘ Also choice of User Interfaces:
  - ☑ Terminal (text) or
  - ☑ GUI: Momo (G4), OPACS
  - ☑ Editors for geometry, EM physics code generation



# One area: Tracking



⌘ What a simulation code needs to do for each step of particle:

- ☑ Determine the **step length**
  - ☒ Corresponding to the applicable physics processes
  - ☒ Checking if it crosses a geometrical boundary
- ☑ Model the **final state** of the track,
  - ☒ Advancing it, potentially in an EM field,
  - ☒ Applying the actions of the physics processes,
    - which can create **secondary** particles.
- ☑ **Deposit** energy in current position ('hit').

# ***Actions during a Step***

## ⌘ During each step

- ☑ Each physics process is given the opportunity to limit the step,
  - ☒ as is the geometry module (at a boundary), and
  - ☒ leading to the decision on this step's length.
- ☑ Physics processes are allowed to apply their effect
  - ☒ If they occur along a step ('continuous')
  - ☒ If they caused the 'hard' event that limited the step ('discreet').



# ***Actions during a Step (cont)***

## ⌘ During a step (continued)

- ☑ An (optional) user-written 'action' is called,
  - ☒ Which can be used eg to create histograms or tallies.
- ☑ If the current volume contains a sensitive detector, that is addressed, allowing it eg
  - ☒ to record the energy deposited,
  - ☒ to record the exact positionin general to create a 'hit' that store all information that is relevant for that detector .

# ***Actions during a Step (cont)***

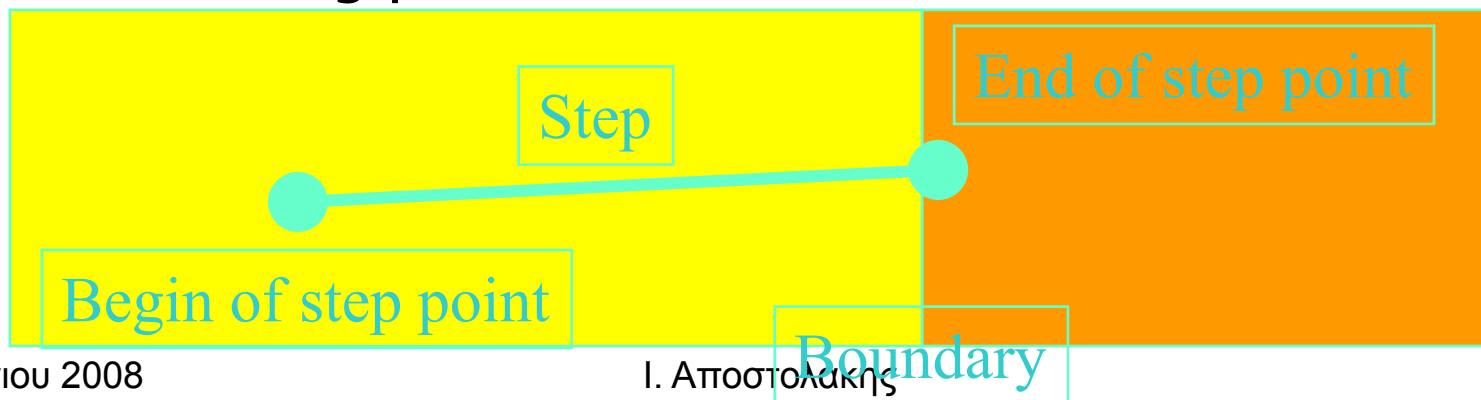
## ⌘ During a step (continued)

☑ A parametrisation can be triggered (Geant4)

☒ Taking over from 'detailed' simulation

☒ Generating directly several hits

This application-specific operates instead of 'normal' physics processes until it returns control and/or resulting particles for further 'detailed' simulation.



# GEANT 4



⌘ Detector simulation **tool-kit** for HEP

☑ offers alternatives, allows for tailoring

⌘ Software Engineering and OO technology

☑ provide the method for building, maintaining it.

⌘ **Requirements** from:

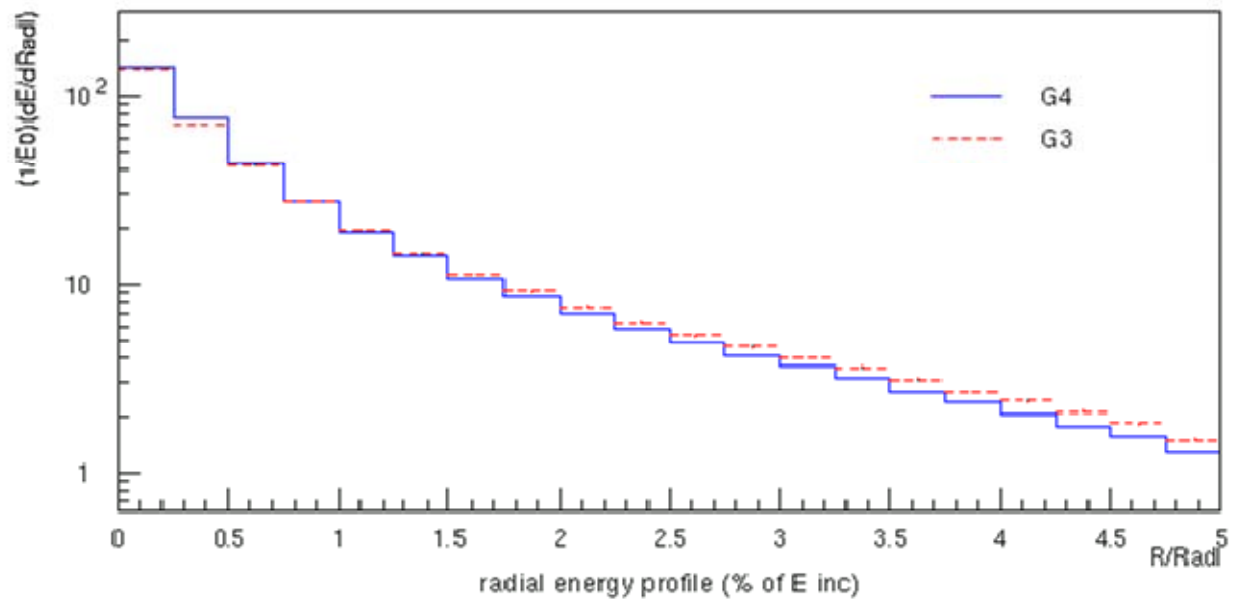
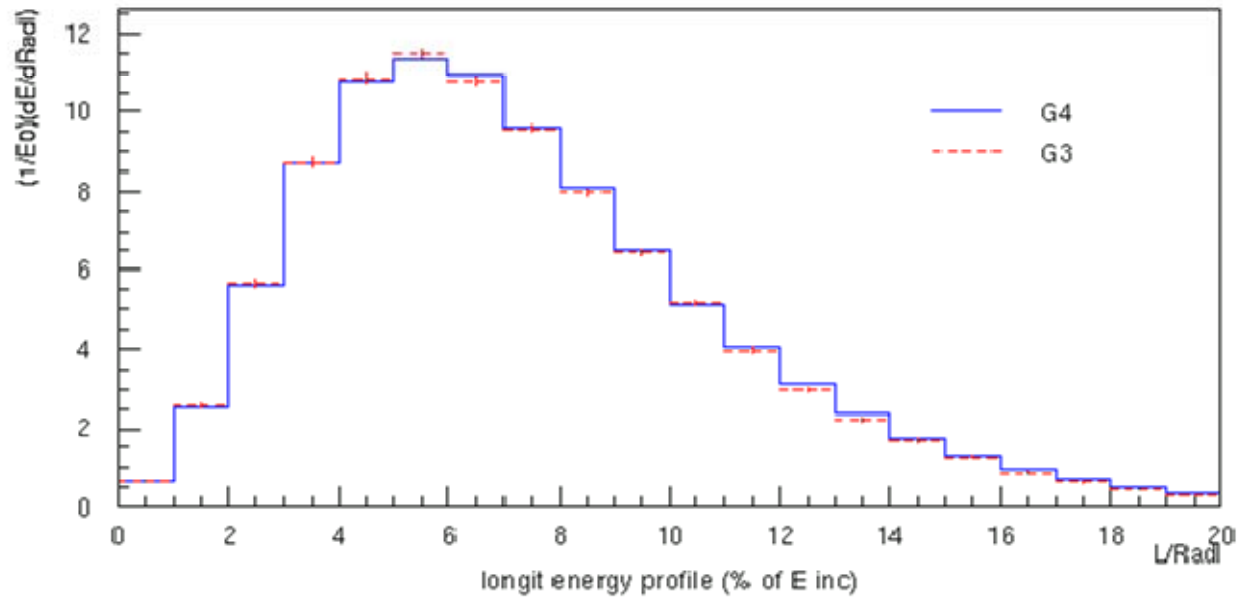
☑ LHC

☑ heavy ions, CP violation, cosmic rays

☑ medical and space science applications

⌘ **World-wide collaboration**

PbWO4 e- 5 GeV G4-G3 comparison



# Sampling calorimeter

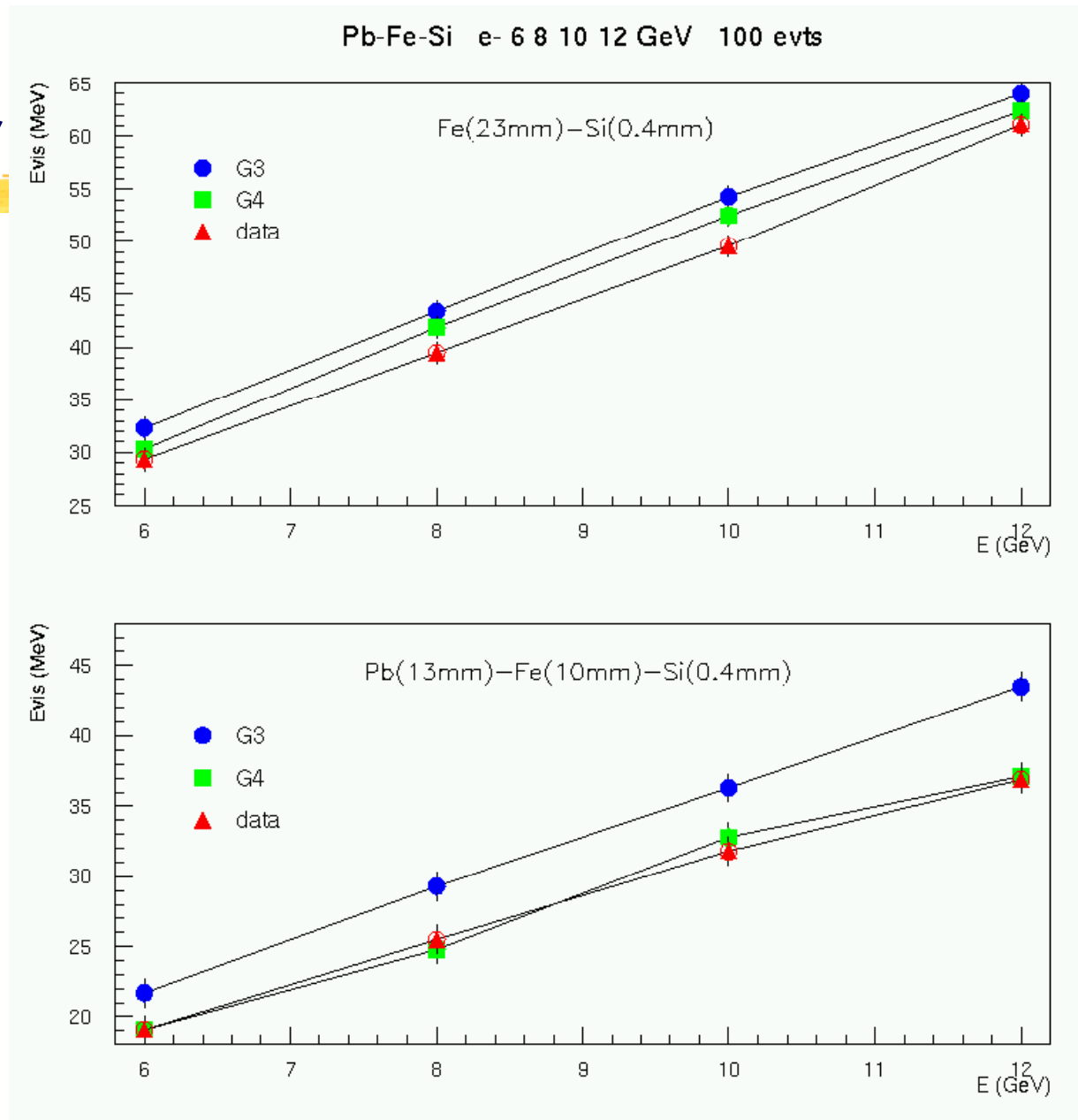
⌘ Sampling calorimeter

⌘ visible energy

⌘ tests

⌘ all EM processes for  $e^-$ ,  $e^+$  and photon

⌘ Data from Sicapo Col. NIM A332 (85-90) 1993



# ***Multiple scattering model***

- ⌘ A new model for multiple scattering based on the Lewis theory is implemented
  - ☑ since public  $\beta$  release in 1998.
- ⌘ It randomizes momentum direction and displacement of a track.
  - ☑ Step length, time of flight, and energy loss along the step are affected, and
  - ☑ It does not constrain the step length.

# Multiple scattering

Examples of comparisons:

15.7 MeV e-

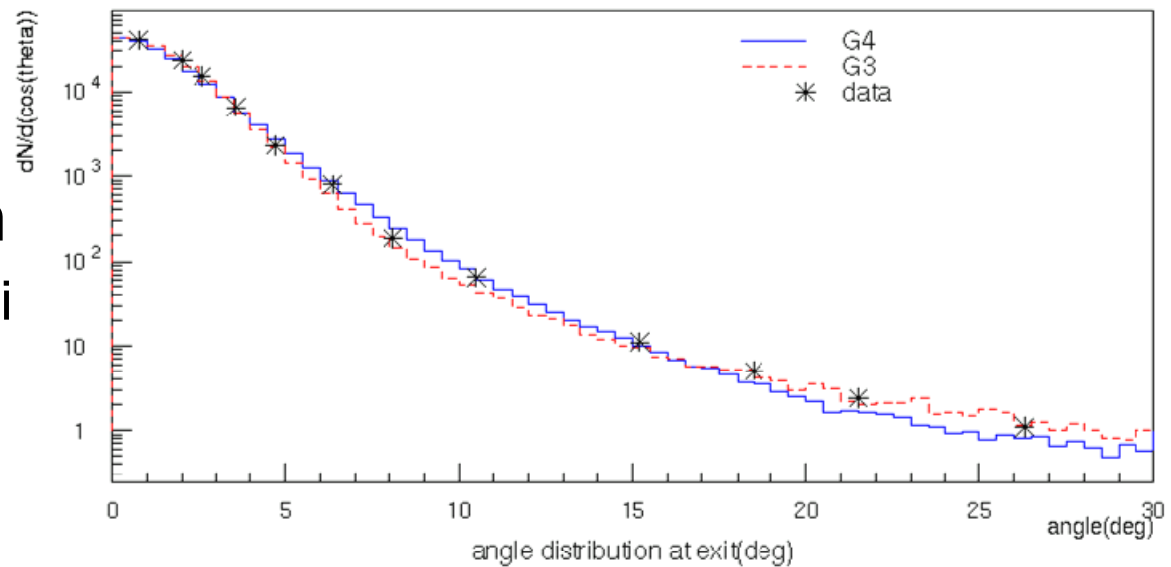
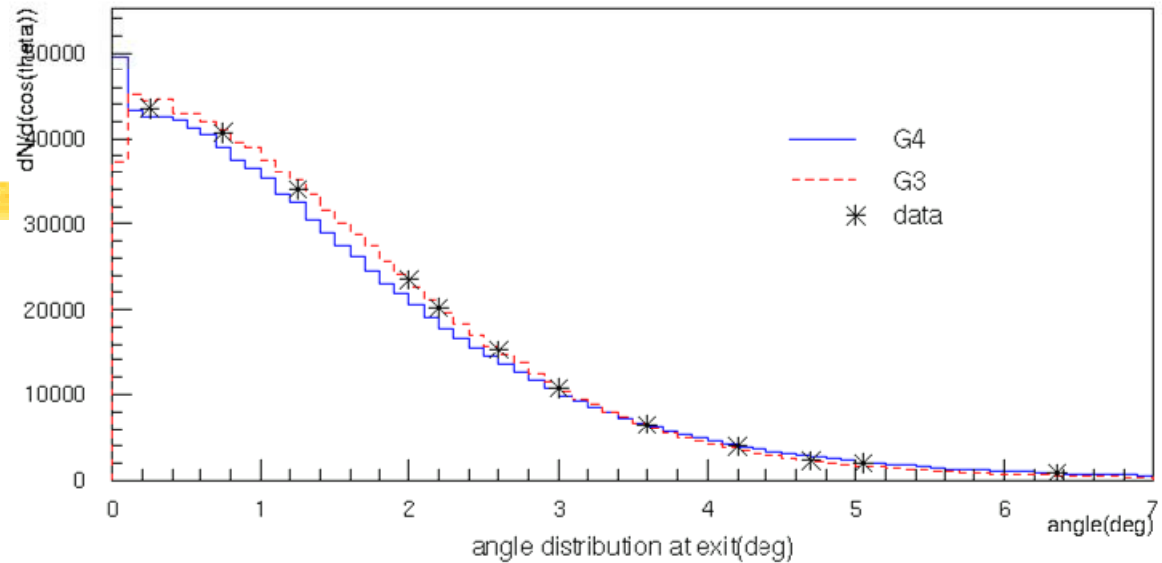
on 19 mg/cm<sup>2</sup>  
gold foil (8 μm)  
figure

6.56 MeV proton

on 93 microns Si

70 GeV/c proto

15.7 MeV e<sup>-</sup>, 18.66 mg/cm<sup>2</sup> gold



Energy deposit of 1 MeV e<sup>-</sup> in 0.530 mm Si

