

Οι Υπολογιστές στη Φυσική Υψηλών Ενέργειων

Γιαννης Αποστολακης
CERN

V0.95

John.Apostolakis@cern.ch

Πλano της ομιλιας

- Η χρηση 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"

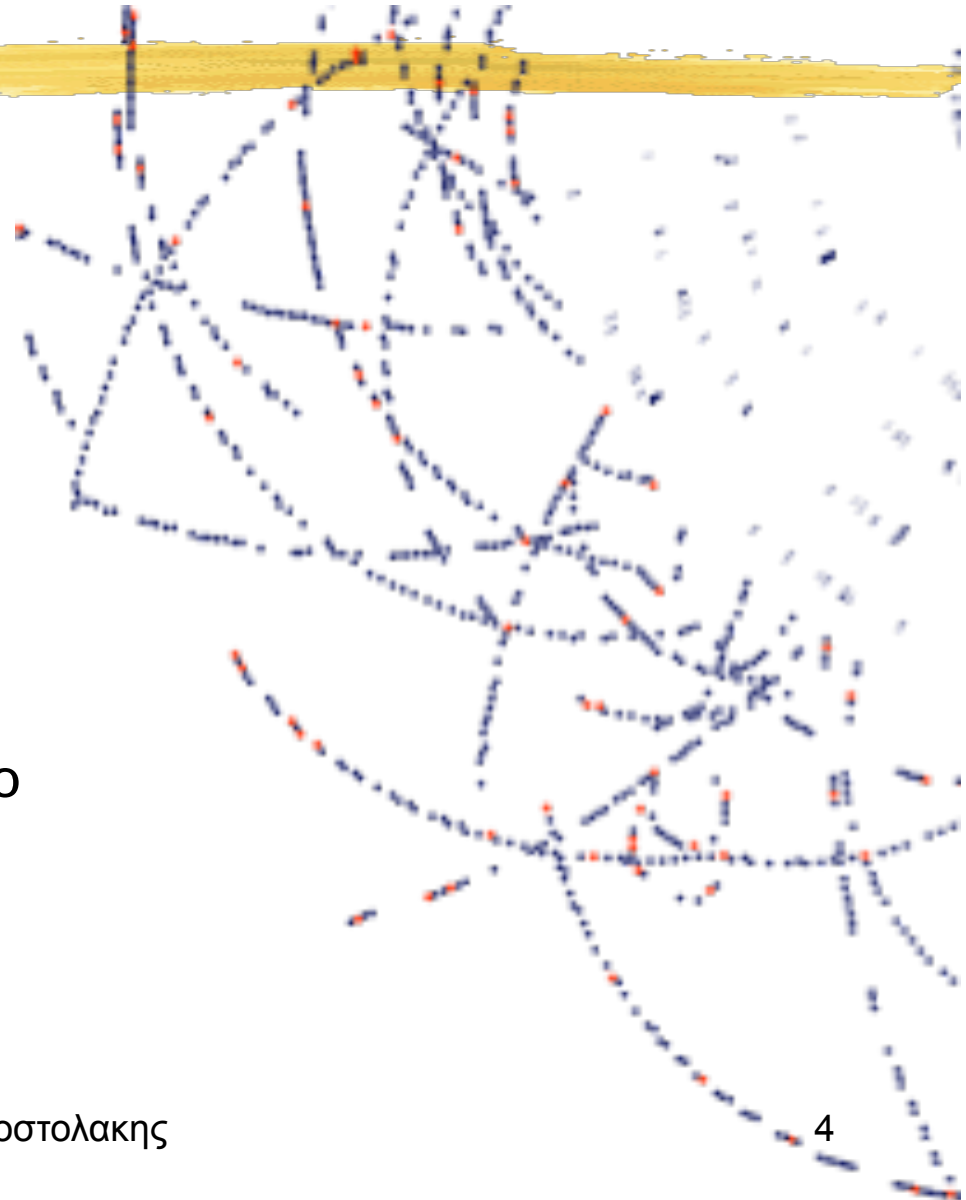
Ανακατασκευή



Μια γρήγορη εισαγωγή

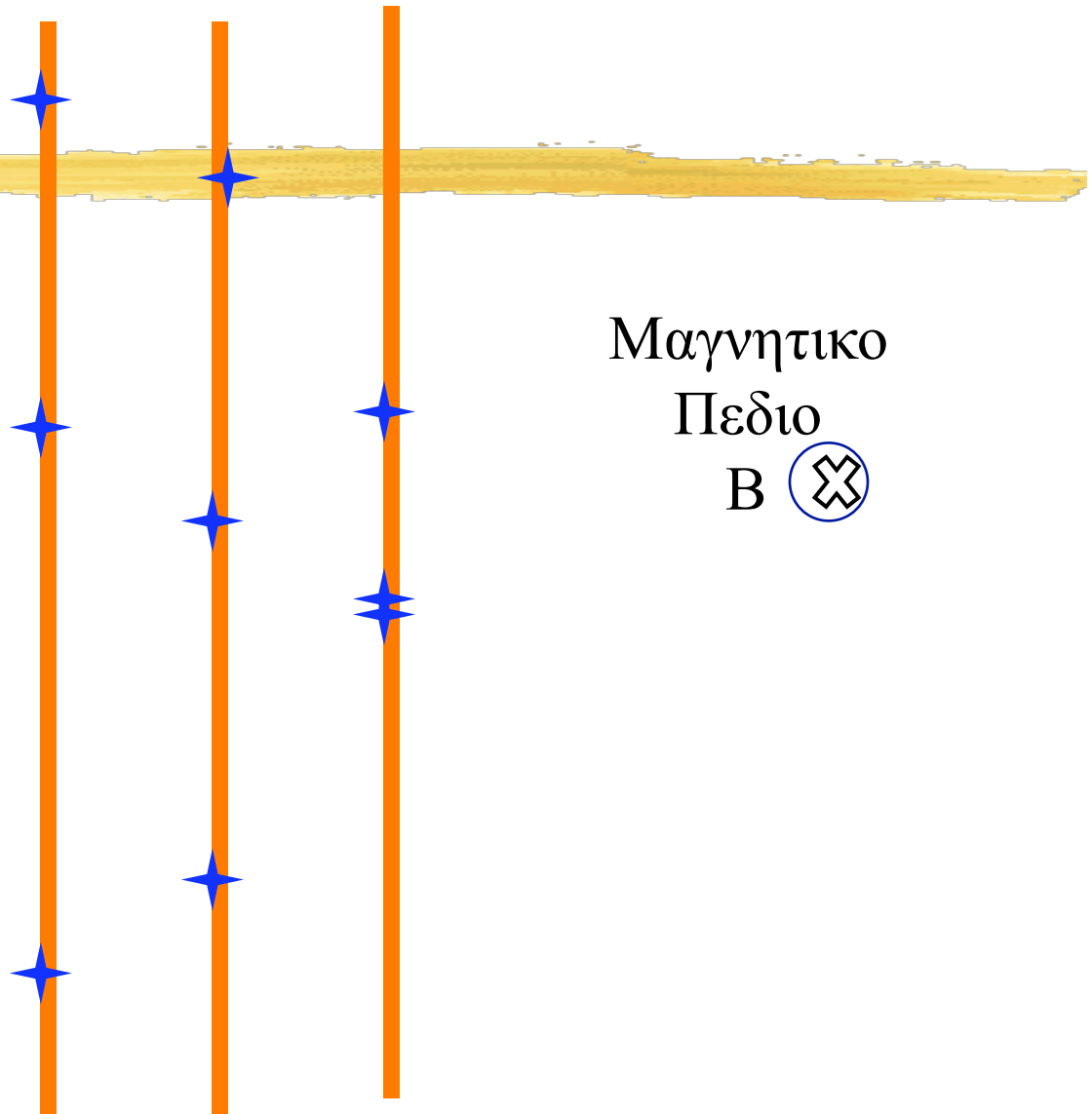
Τι είναι η ανακατασκευή?

- Οι μετρήσεις είναι σαν ένας γρίφος
 - Τι τροχιές τις προκάλεσαν?
- Κάθε μέτρηση θέσης βοηθάει
 - Υπάρχουν όμως 100-άδες ως **χιλιάδες μετρήσεις**
- Η ανακατασκευή πρέπει να βρει τη **λύση!**
 - Ξεροντας καλά το μαγνητικό πεδίο
 - Βρισκουμε ποιες μετρήσεις ανοικουν σε ποιες τροχιες



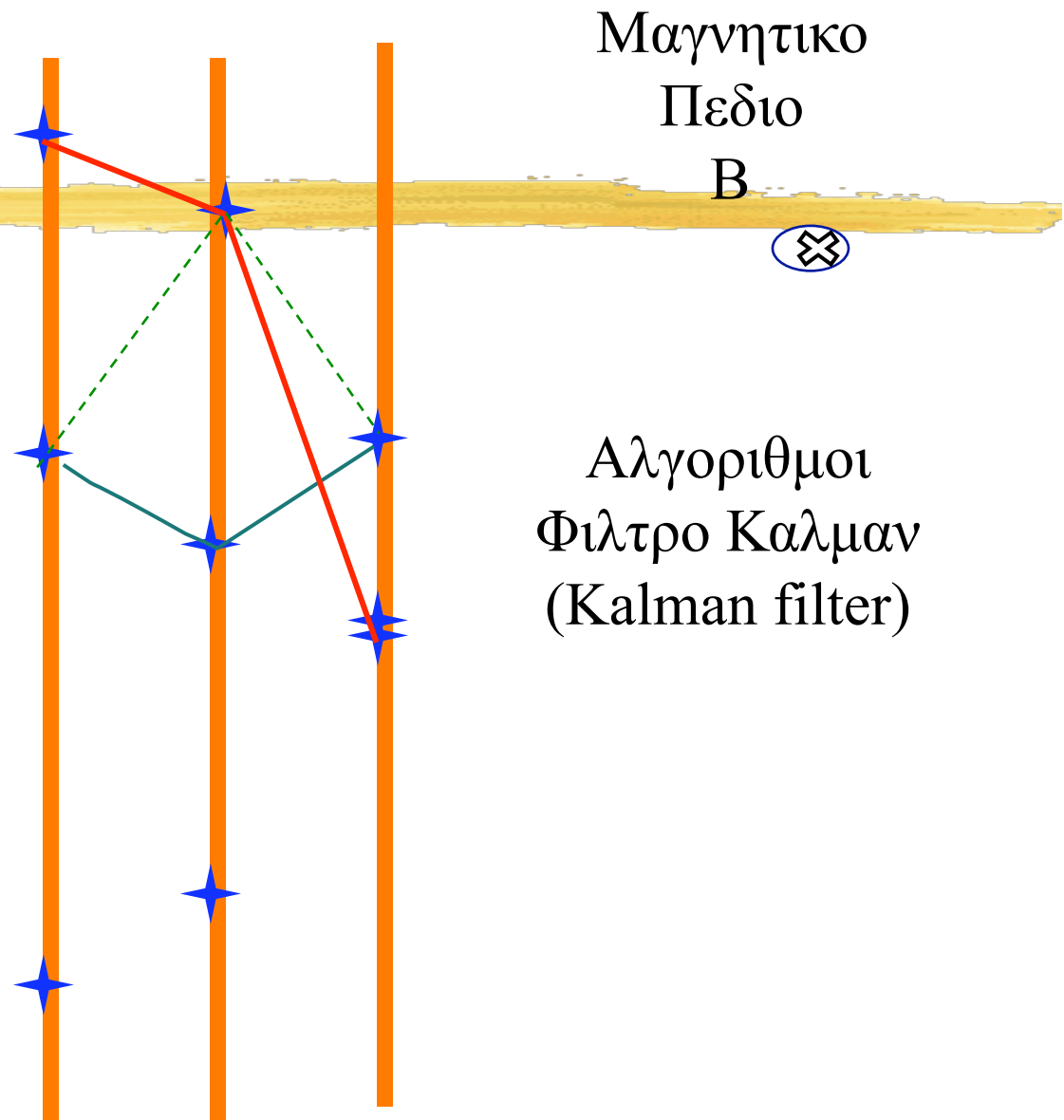
Ανακατασκευή στην πράξη

- Αρχίζει με τις θέσεις διαβάσης των σωματιδίων



Ανακατασκευή στην πράξη

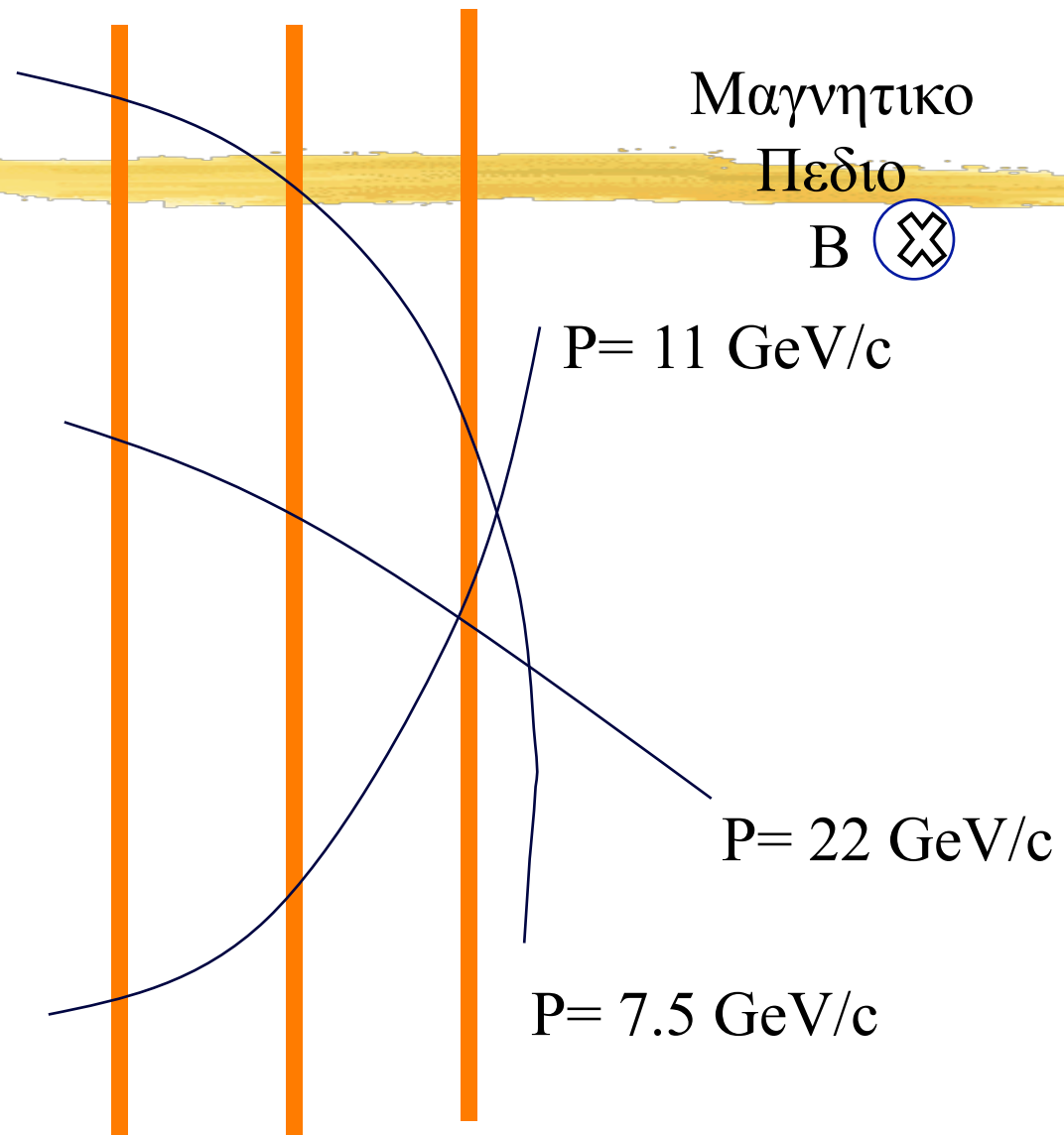
- Αρχίζει με τις θέσεις διαβάσης των σωματιδίων
- Δωκιμάζονται διαφοροί συνδιασμοί
 - και υπολογίζεται η διαφορά μετρήσης-προβλεψής
 - Και έτσι πιθανότητα του καθε συνδιασμου



Αλγοριθμοί
Φίλτρο Καλμαν
(Kalman filter)

Ανακατασκευή : αποτέλεσμα

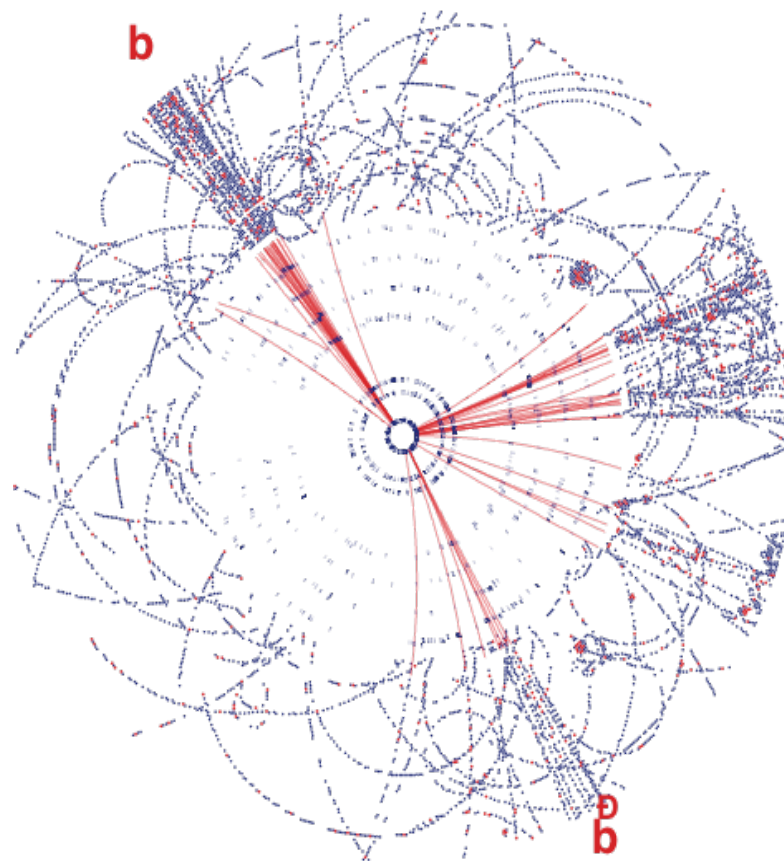
- Αρχίζει με τις θέσεις διαβάσης των σωματιδίων
- Δωκιμάζονται διαφοροί συνδιασμοί
 - και υπολογίζεται η διαφορά μετρήσης-προβλεψής
 - Και έτσι πιθανότητα του καθε συνδιασμου
- Τελικά έχουν βρεθεί όλες οι τροχιές
 - ή «στα γρηγορά» αυτές με μεγάλη ορμή-οι κυρίες τροχιές



Atlas : Physics Signatures and Event Rates

- ❑ Οι δεσμες διασταυρονονται με ρυθμο 40 MHz
- ❑ $\sigma_{inelastic} = 80 \text{ mb}$
 - Σε καθε περασμα πολλές 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 (~ 1 in $10^5 - 10^9$)

ATLAS Barrel Inner Detector
 $H \rightarrow b\bar{b}$



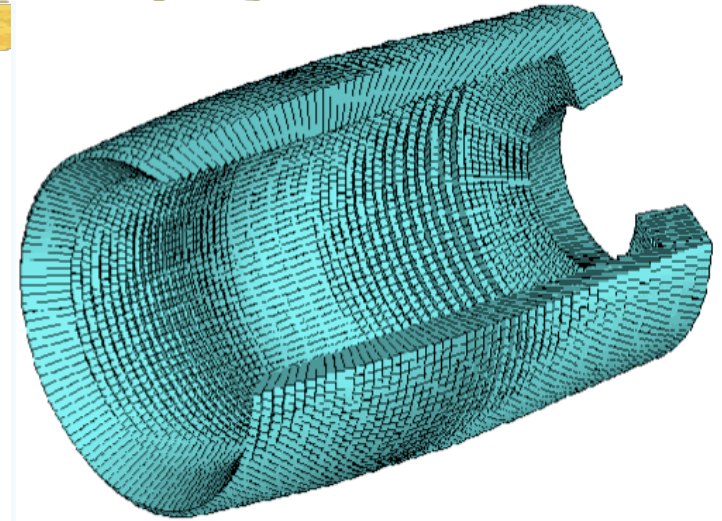
Προσομοίωση και Ανιχνευτές



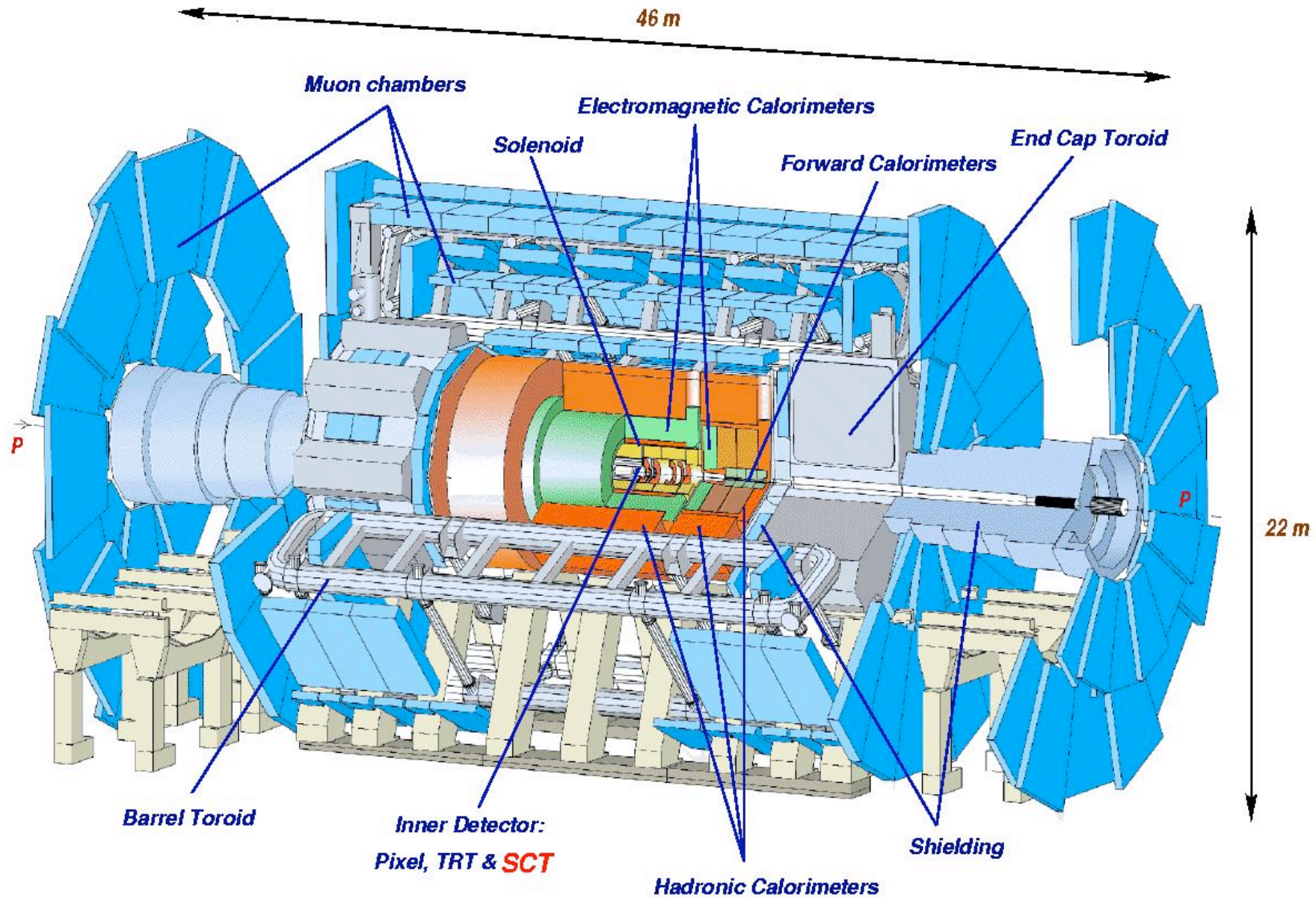
Τι είναι η προσομοίωση
Γιατί υπάρχει
Πως γίνεται

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

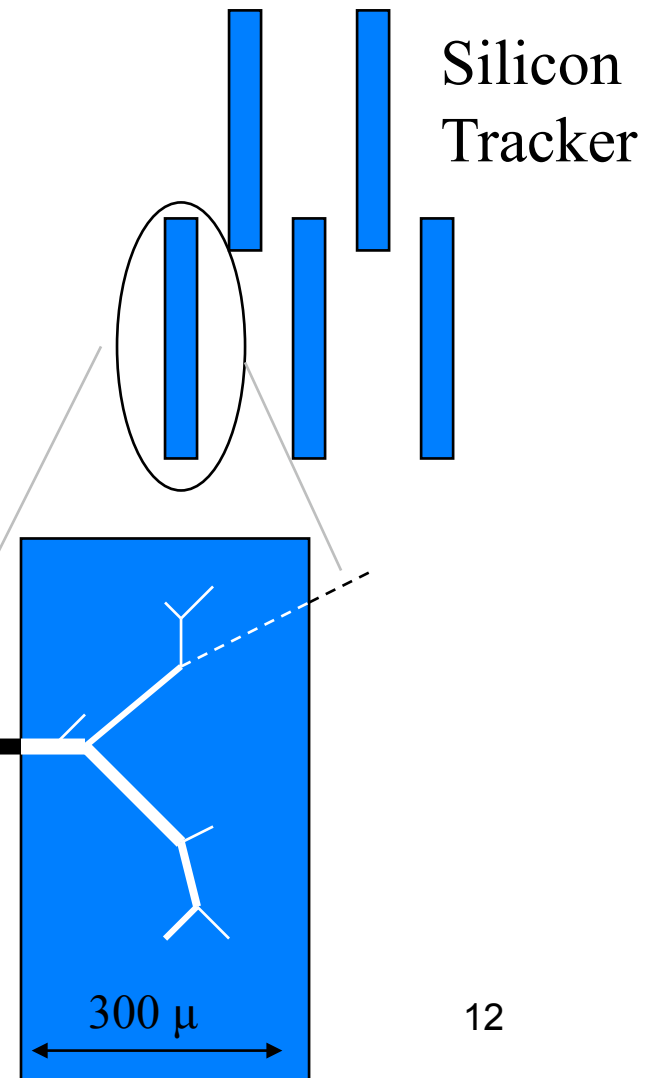


Τι είναι προσομοίωση?

- Φτιάχνουμε μοντελα
 - Του ανιχνευτη
 - Γεωμετρια
 - Υλικά
 - Των αλληλεπιδρασεων
 - Καθε γνωστου τυπου
 - Ηλεκτρομαγνητικου
 - Υσχηρου πυρηνικου

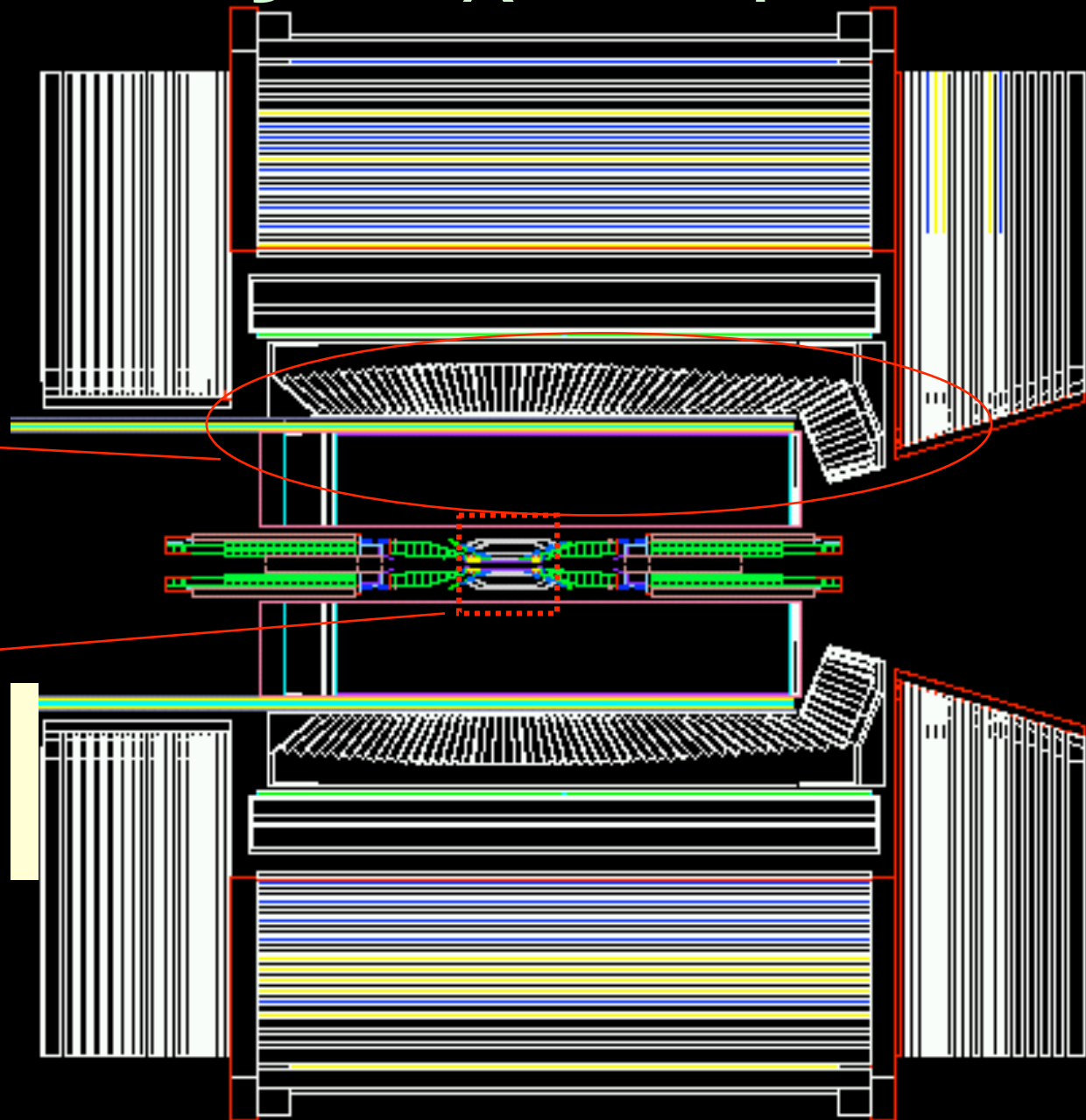
$$\sigma_{\text{συνολο}} = \sum \sigma_{\text{φαινομενου}}$$

2.5 MeV e⁻
ηλεκτρονιο



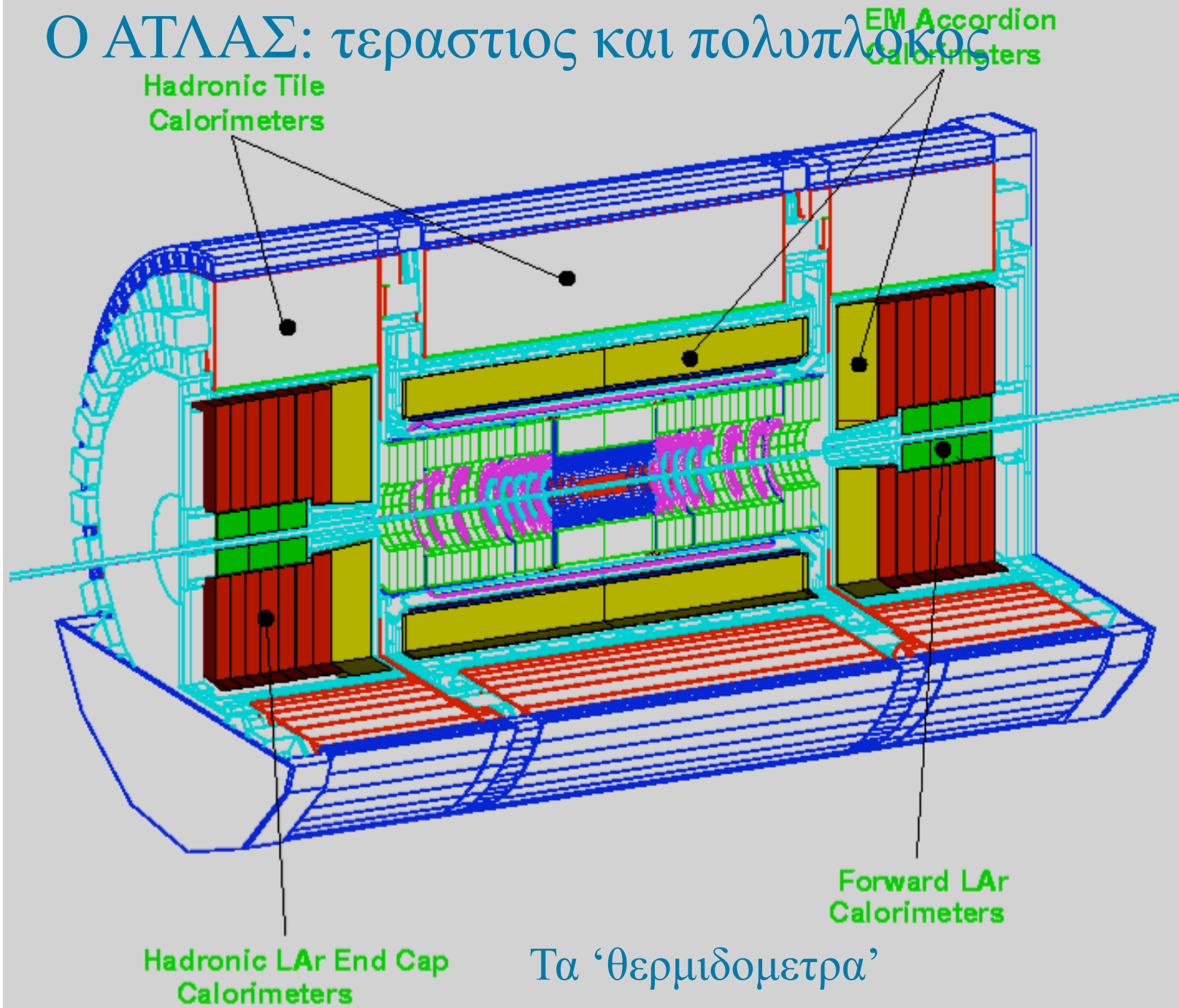
Γεωμετρία ενός ανιχνευτή

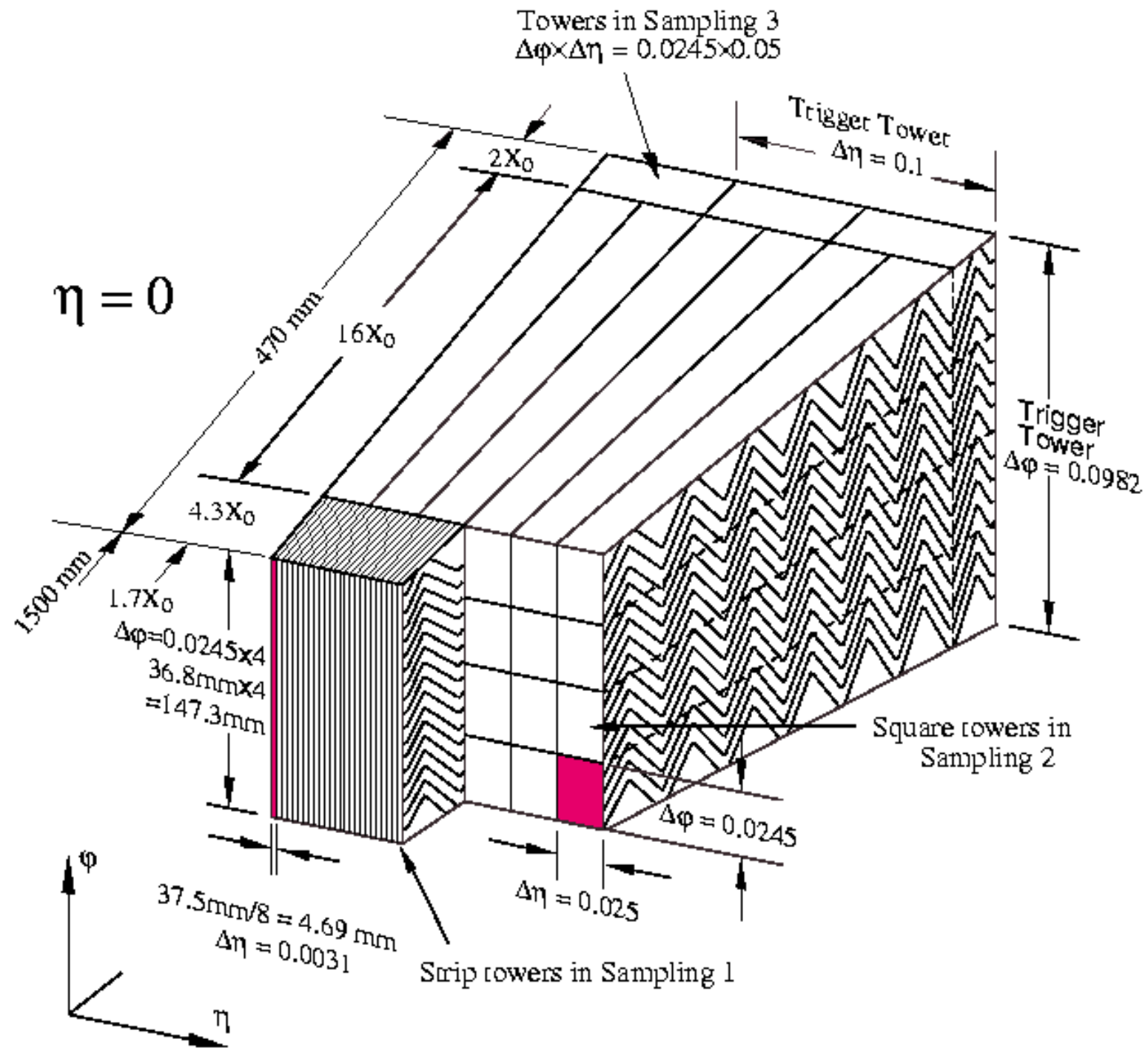
- EM Calorimeter
 - Crystal
- Tracker
 - Precision
 - Drift Chamber



BABAR
(SLAC, US)
27 Iouviou

Ο ΑΤΛΑΣ: τεραστιος και πολυπλοκος

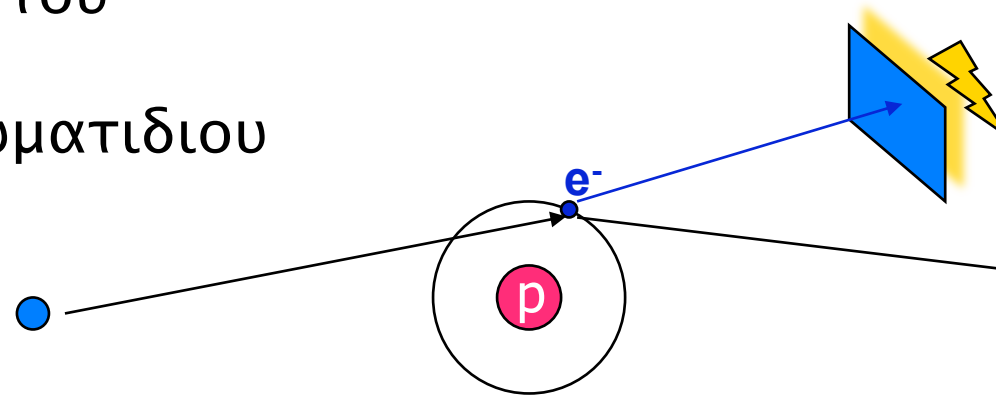




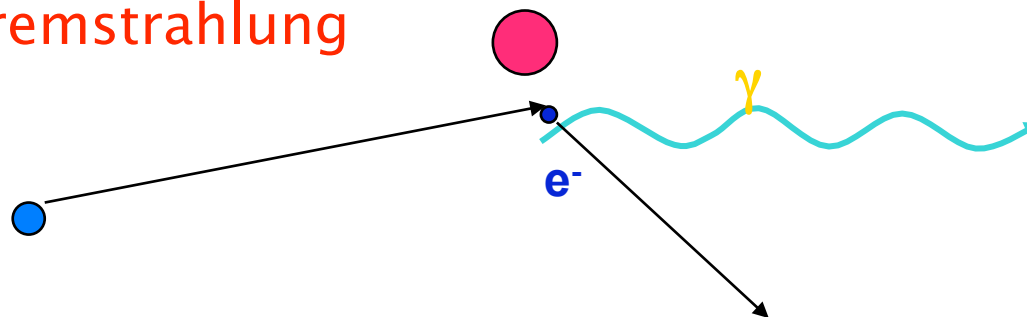
Βασικές Αλληλεπιδρασεις

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

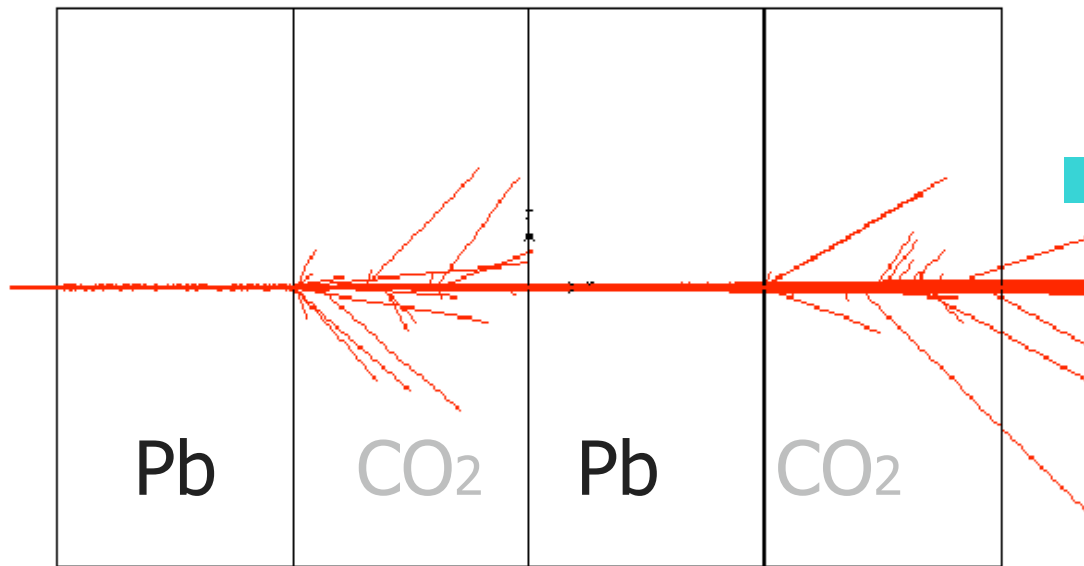
■ Ιοντισμός



■ Bremsstrahlung

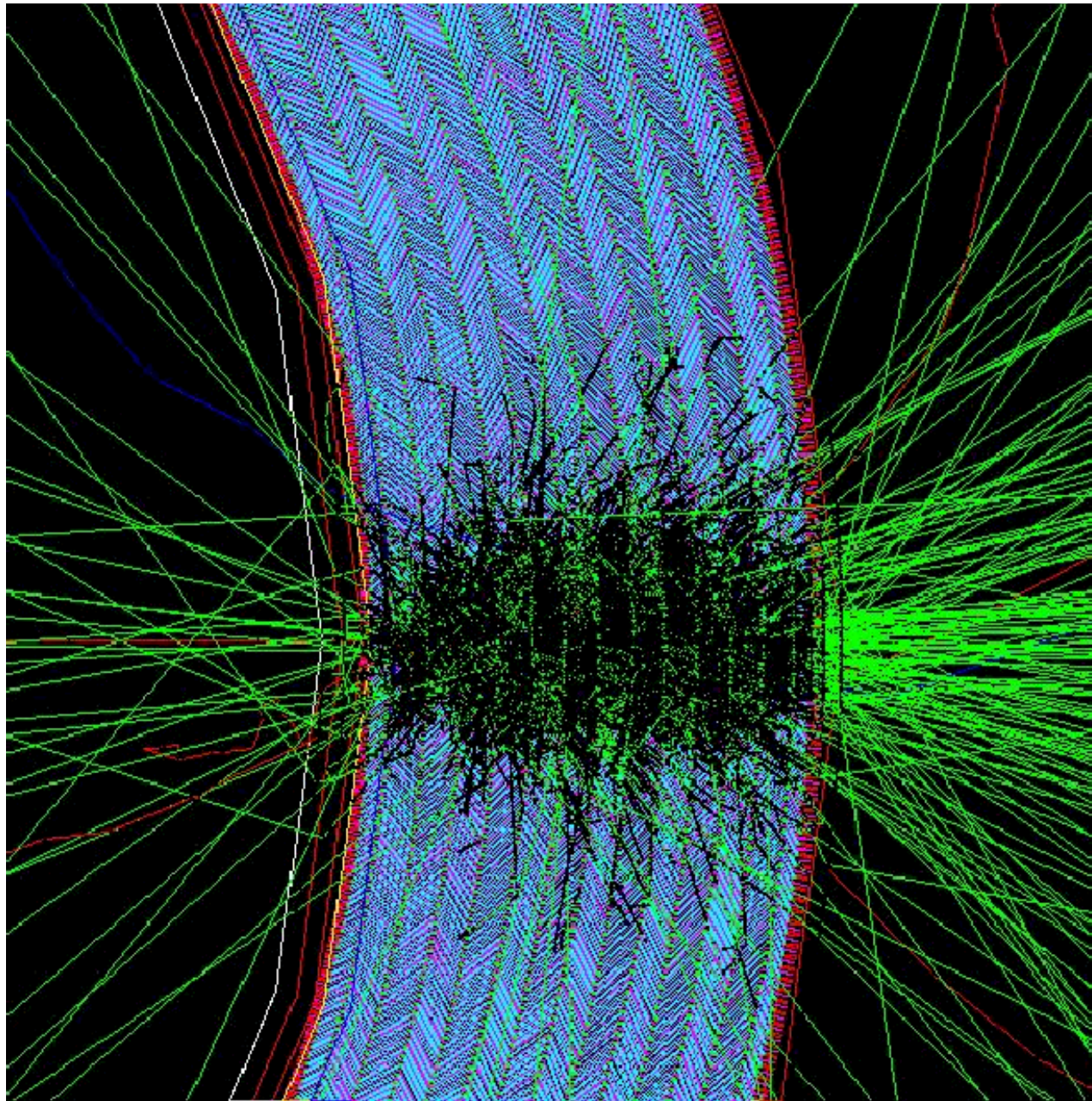


Ένα απλο παραδειγμα



GEANT 3

- 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



27 lo

. Kordas "Geant4 for the ATLAS EM calo" — CALOR2000, Annecy, 12 October 2000 (4

18

Γιατι προσομοιωση ?



- Για να σχεδιασουμε ΤΟΥΣ ΑΝΙΧΝΕΥΤΕΣ
- Για να ετοιμασουμε τις μεθόδους ανακατασκευής
- Για να καταλαβούμε τον ανιχνευτή

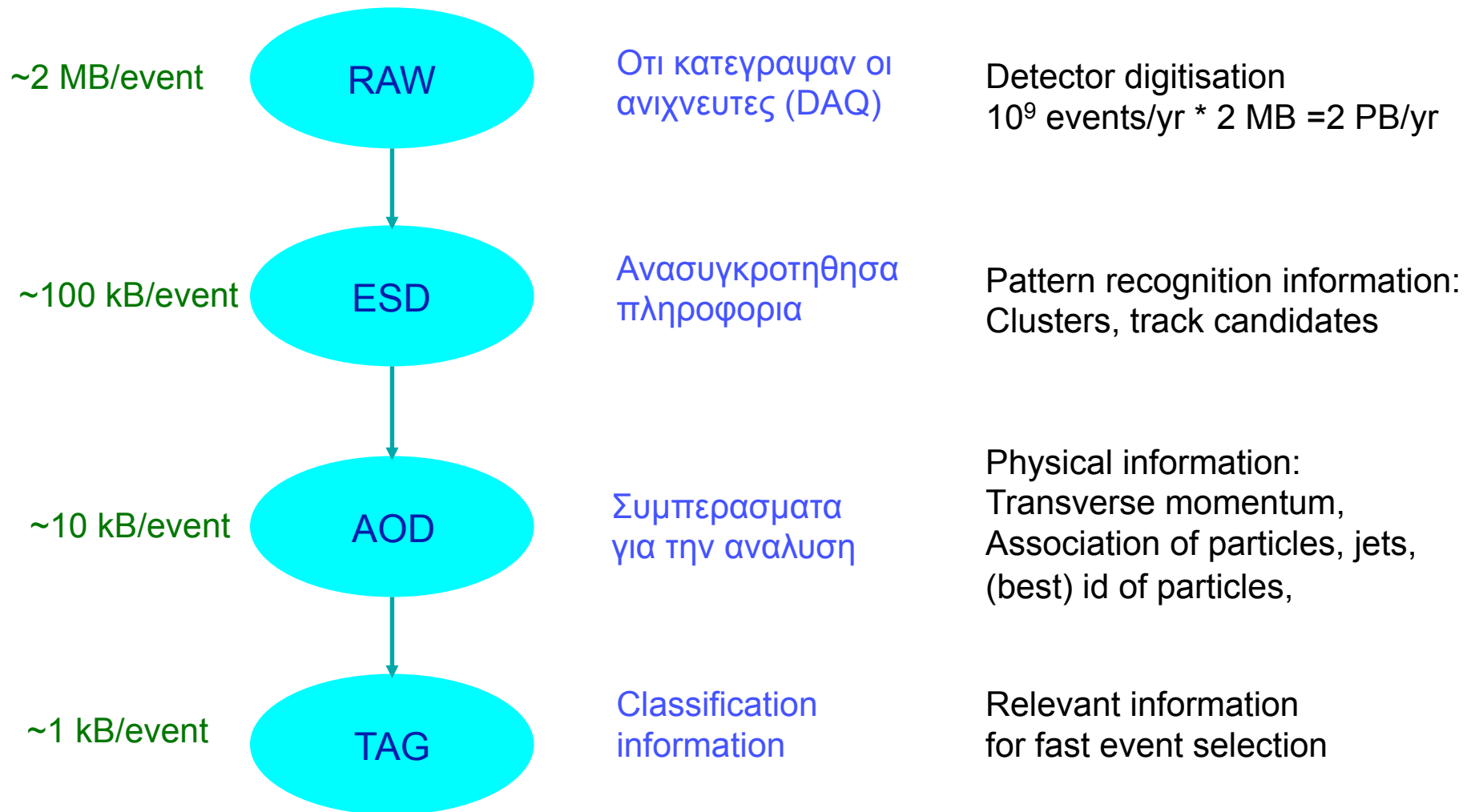
Υπάρχει απλη λυση ?

- Αρκει η μεση τιμη της αποθεσης ενεργειας (π.χ.) ?
 - Για μερικες απλες ερευνες, Ναι
 - Για πολλες (τις περισοτερες) 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)





Distribution of Computing Services

Τα νέα δεδομένα θα μεγαλώνουν 15 PetaBytes
καθε χρόνο – με δυο αντιγραφα

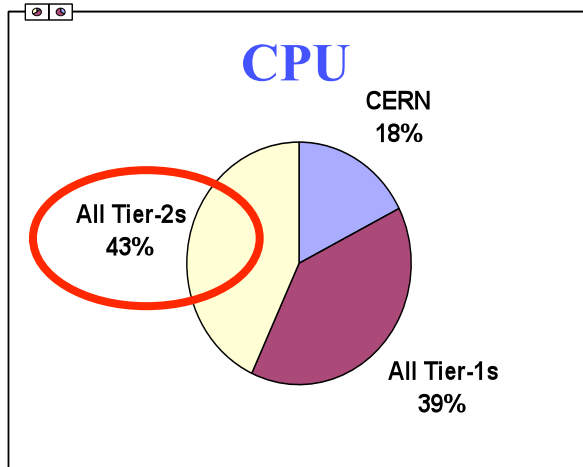
Περίπου 100,000
'πυρήνες' cores

Summary of Con

All experiments - 2008

From LCG TDR - June 2008

	CERN	All Tier-1s	All Tier-2s	Total
CPU (MSPECint2000s)	25	56	61	142
Disk (PetaBytes)	7	31	19	57
Tape (PetaBytes)	18	35		53



Disk

Tape

$$\text{MSPECint} = 10^6$$

Μεγάλο ποσοστό των υπολογιστών και δίσκων είναι διαδεδωμένα σε 120 υπολογιστικά κέντρα



Λυση : το Πλεγμα (Grid)

- Χρήσιμοποιουμε το Πλεγμα να ενώσουμε τους υπολογιστικούς πόρους των ινστιτουτων ανά τον κόσμο

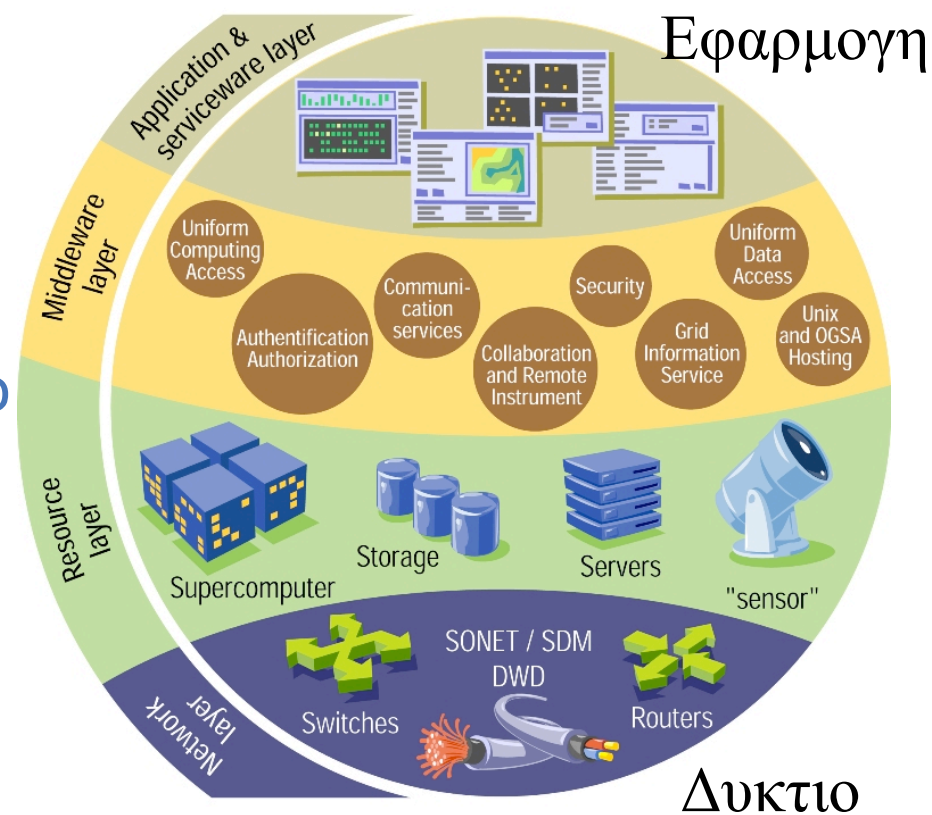
Το World Wide Web παρέχει απλη πρόσβαση σε πληροφορίες που είναι αποθηκευμένες σε πολλά εκατομμύρια διαφορετικές γεωγραφικές τοποθεσίες

The **Grid** είναι μια υποδομή που παρέχει αδιάλειπτη πρόσβαση σε υπολογιστική ισχύ και χωρητικότητα αποθήκευσης δεδομένων καταναεμημένη σε όλη την υφήλιο

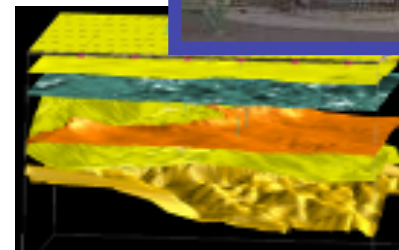
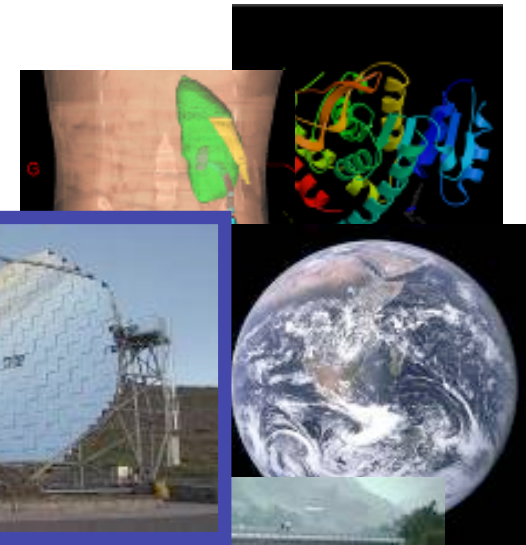
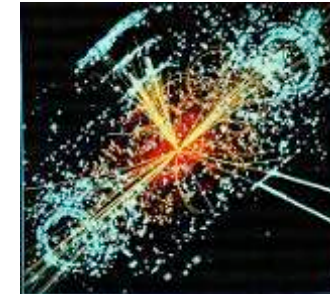


Πως δουλεει το Πλεγμα (Grid)

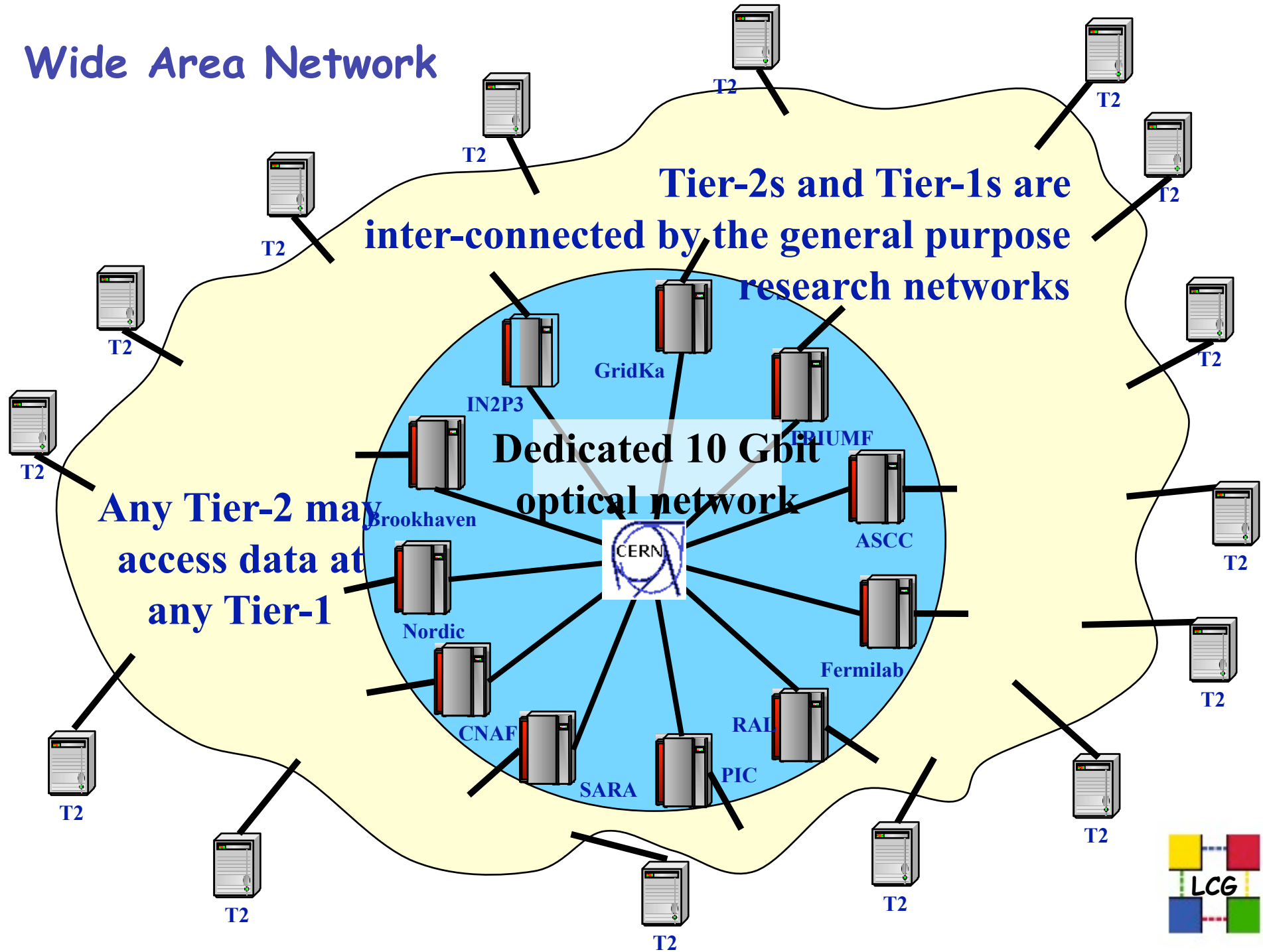
- Βασίζεται σε ειδικευμενο λογισμικο, το λεγομενο **μεσο-λογισμικο** (middleware).
- Η ιδεα του Middleware ειναι να βρει αυτοματα τα **στοιχεια** που χρειαζεται ο/η ερευνητης, και το **υπολογιστικο δυναμικο** να τα επεξεργαστει.
- Middleware εξισοροπει το φορτιο σε διαφορους πορους και εγκαταστασεις. Χειριζεται επισης **ασφαλεια**, λογιστικη, επιτηρηση και πολλα αλλα.

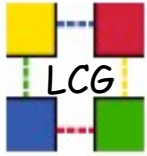


- Περισσότερες από 20 εφαρμογες σε 7 τομείς
 - Φυσική Υψηλών Ενέργειων (Pilot domain)
 - 4 LHC experiments, DESY, Fermilab
 - Βιοϊατρική (Pilot domain)
 - Βιοπληροφορική (Bioinformatics)
 - Ιατρική απεικόνιση (Medical imaging)
 - Γεωεπιστήμες
 - Γεω-επισκόπηση
 - Φυσική Στερεάς Γης? (Solid Earth Physics)
 - Υδρολογία
 - Κλίμα
 - Υπολογιστική Χημεία
 - Fusion
 - Αστρονομία
 - Cosmic microwave background
 - Gamma ray astronomy
 - Γεωφυσική
 - Industrial applications



Wide Area Network

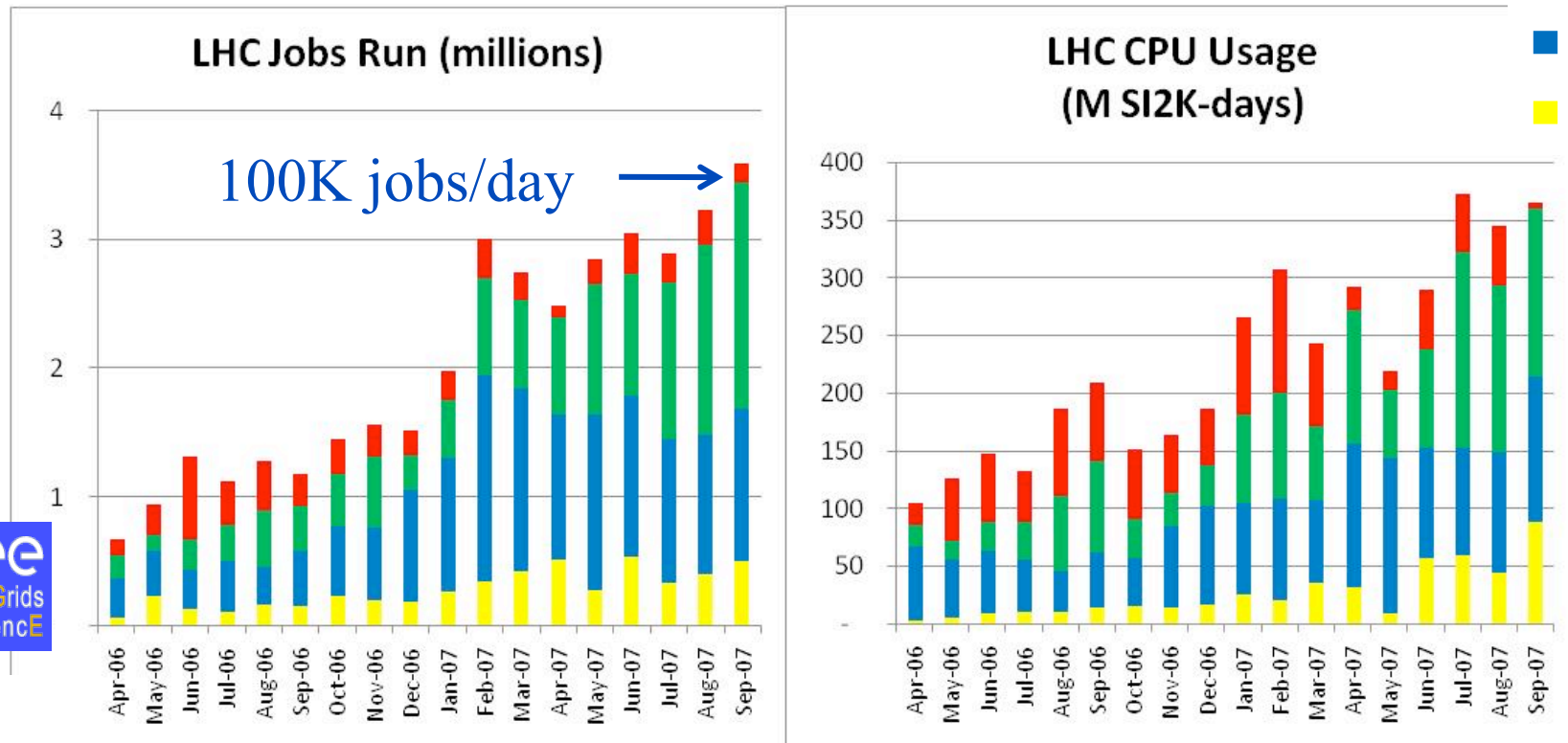




Grid Activity

- Continuing increase in usage of the EGEE and OSG grids
- All sites reporting accounting data (CERN, Tier-1, -2, -3)
- Increase in past 17 months - 5 X number of jobs
- 3.5 X cpu usage

- lhcb
- cms
- atlas
- alice

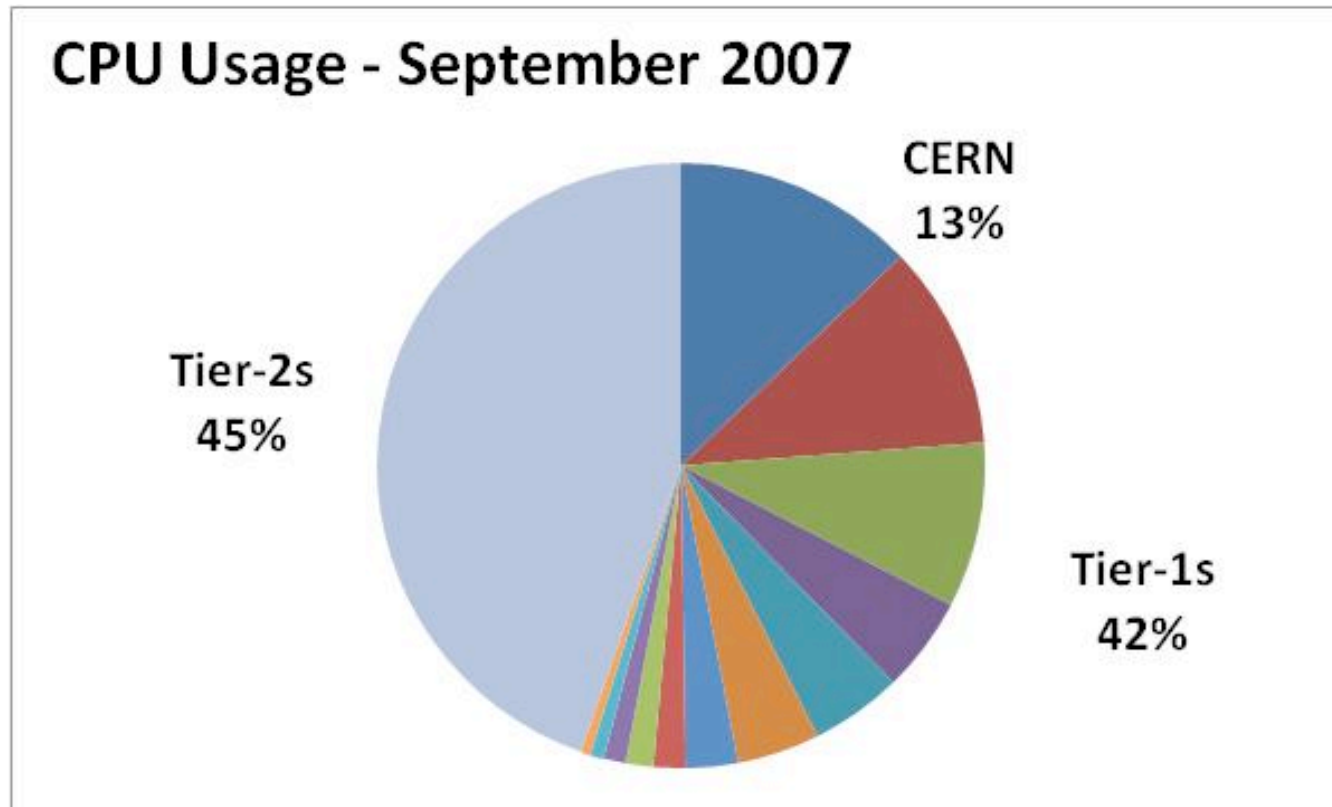


Οι φυσικοί προγραμματίζουν

- Ολη 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 χρησιμοποιώ τα ‘εργαλεία’ να δουν τις περιλήψεις των μετρήσεων

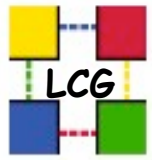


September 2007 - CPU Usage CERN, Tier-1s, Tier-2s



- > 80% of CPU Usage is external to CERN





Backup

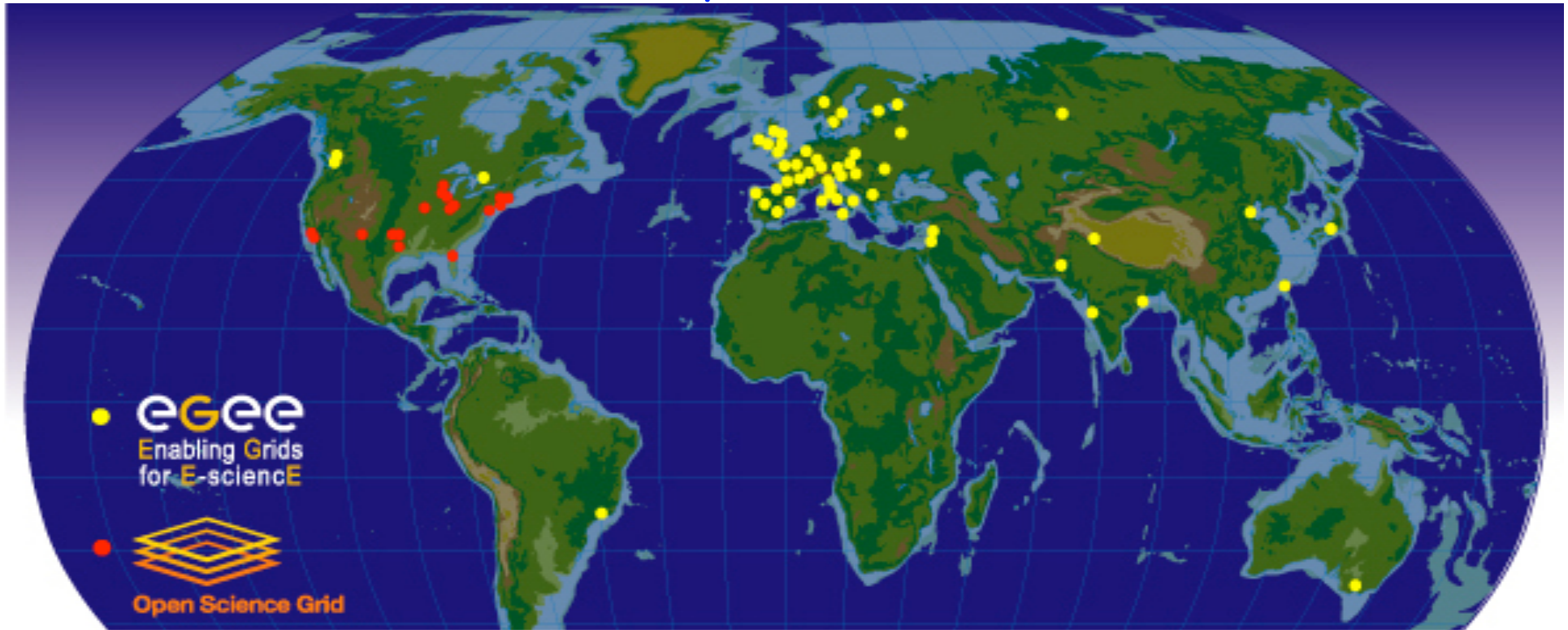




LCG depends on two major science grid infrastructures

EGEE - Enabling Grids for E-Science

OSG - US Open Science Grid



A map of the worldwide LCG infrastructure operated by EGEE and OSG.



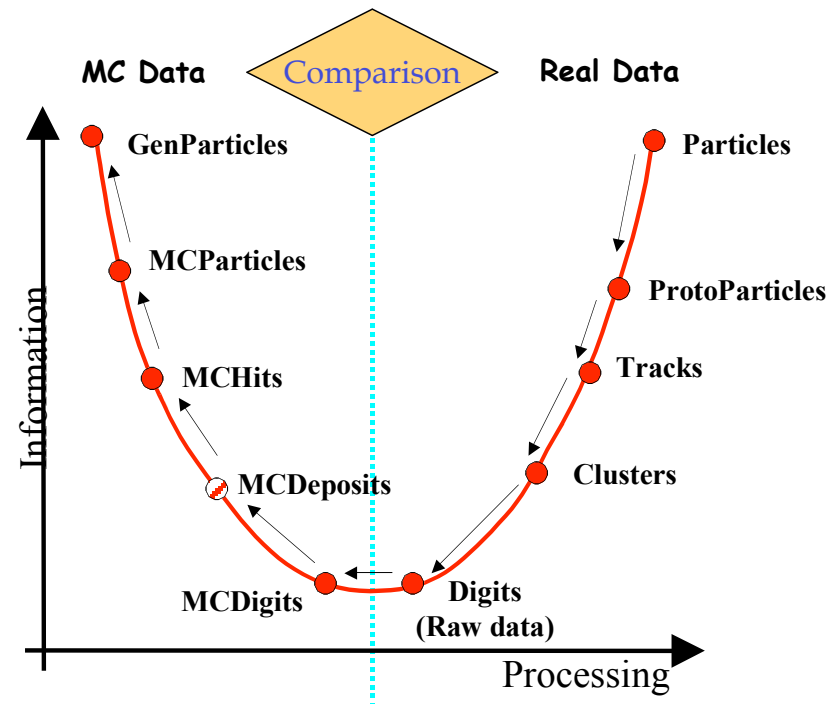
The Worldwide LHC Computing Grid

- The LHC physics data analysis service distributed across the world
 - CERN, 11 large Tier-1 centres,
~ 140 active Tier-2 centres
- Status in May 2007
 - Established the 10 Gigabit/sec optical network that interlinks CERN and the Tier-1 centres
 - Demonstrated data distribution from CERN to the Tier-1 centres at 1.3 GByte/sec - the rate that will be needed in 2008
 - ATLAS and CMS can each transfer 1 PetaByte of data per month between their computing centres
 - Running ~2 million jobs each month across the grid
 - The distributed grid operation, set up during 2005, has reached maturity, with responsibility shared across 7 sites in Europe, the US and Asia
 - End-user analysis tools enabling "real physicists" to profit from this worldwide data-intensive computing environment





Data Organisation



Backup



More on simulation

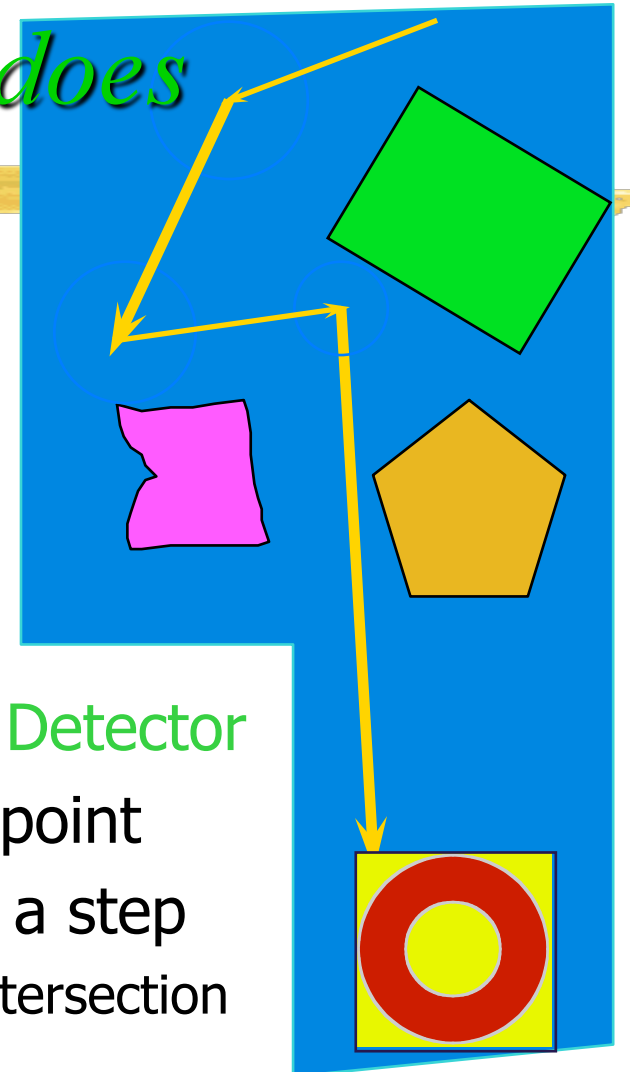
Geant4 geometry: what it does

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

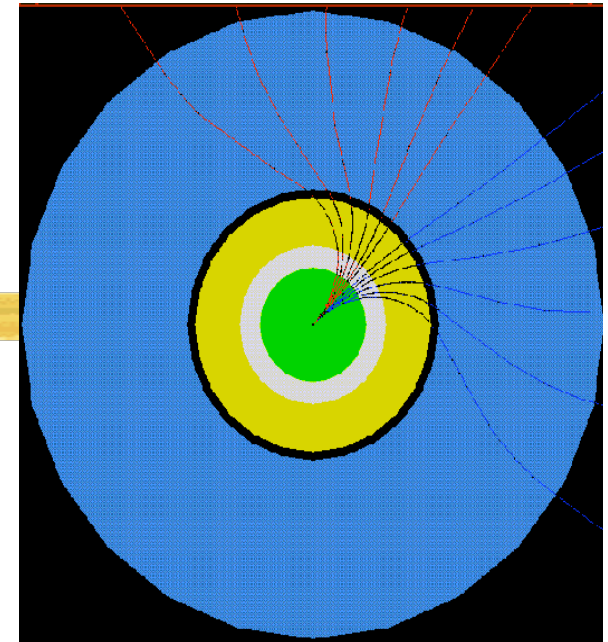
Navigates in Detector

- 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



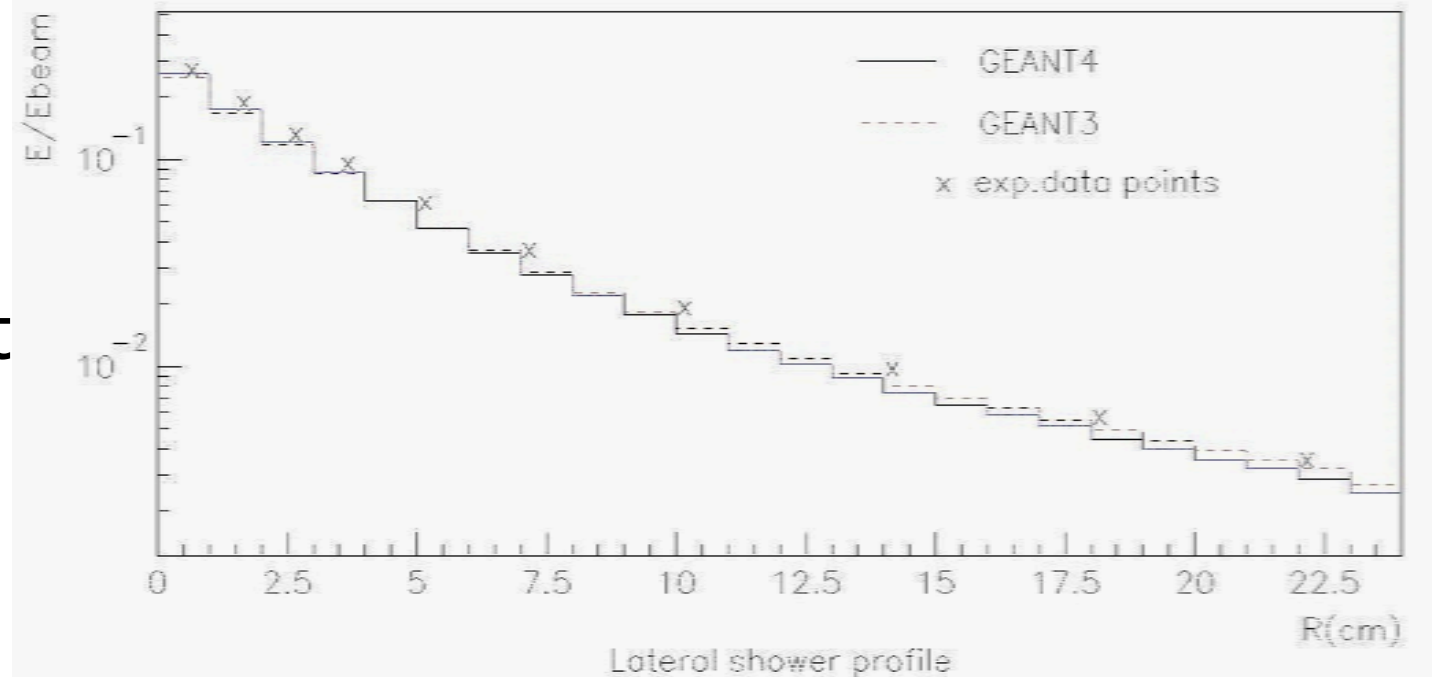
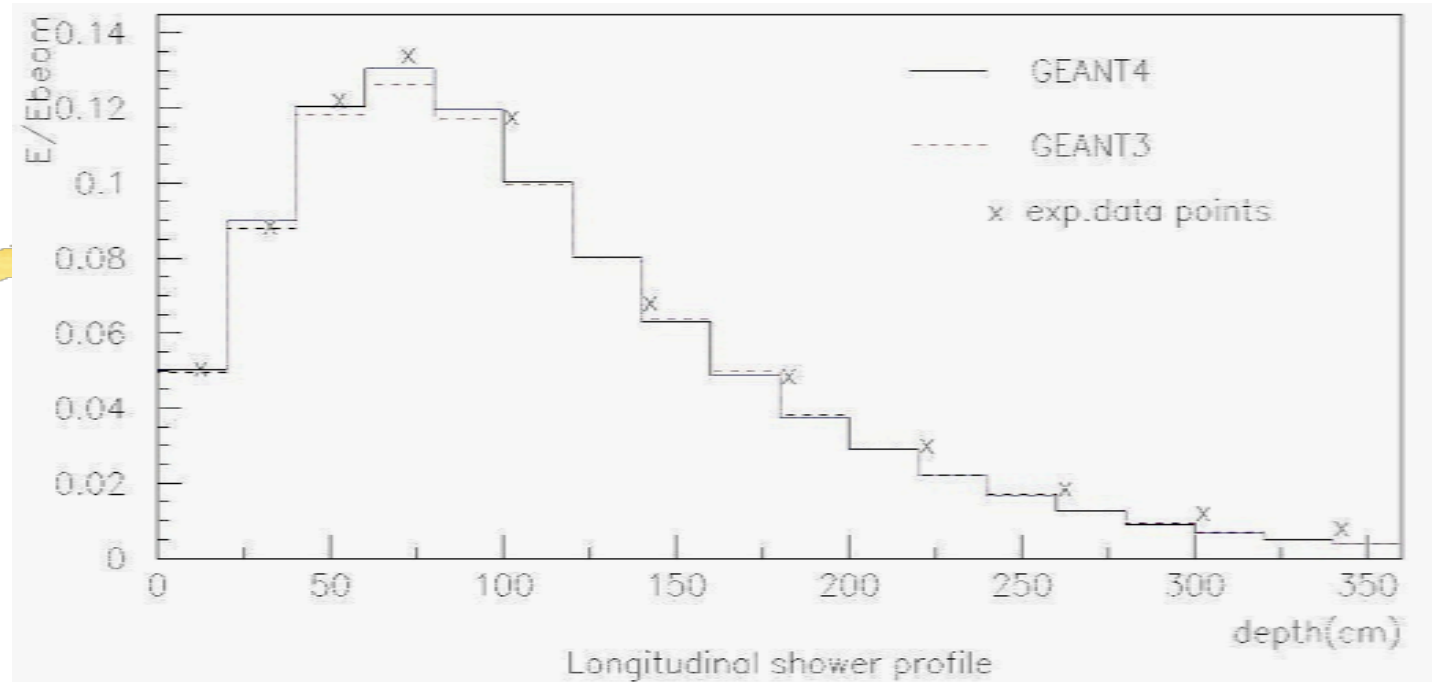
- **Gammas:**
 - Gamma-conversion, Compton scattering, Photo-electric effect
- **Leptons(e , μ), charged hadrons, ions**
 - Energy loss (Ionisation, Bremstrahlung) or PAI model energy loss, Multiple scattering, Transition radiation, Synchrotron radiation,
- **Photons:**
 - Cerenkov, Rayleigh, Reflection, 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 H₂O

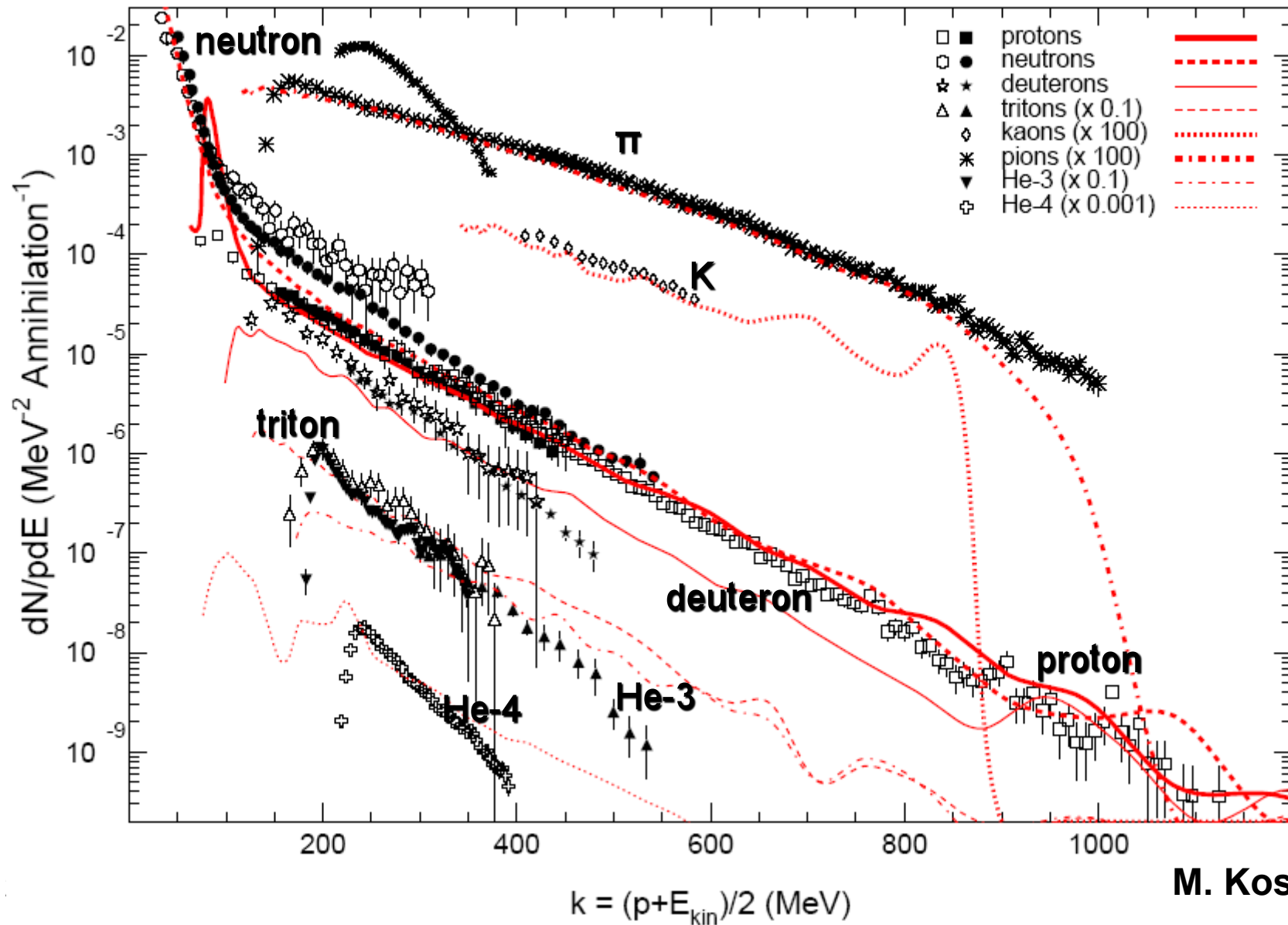
G4,
Data
G3

■ Good agreement seen with the data



Antiproton annihilation - CHIPS Model

Antiproton annihilation on ^{238}U nucleus



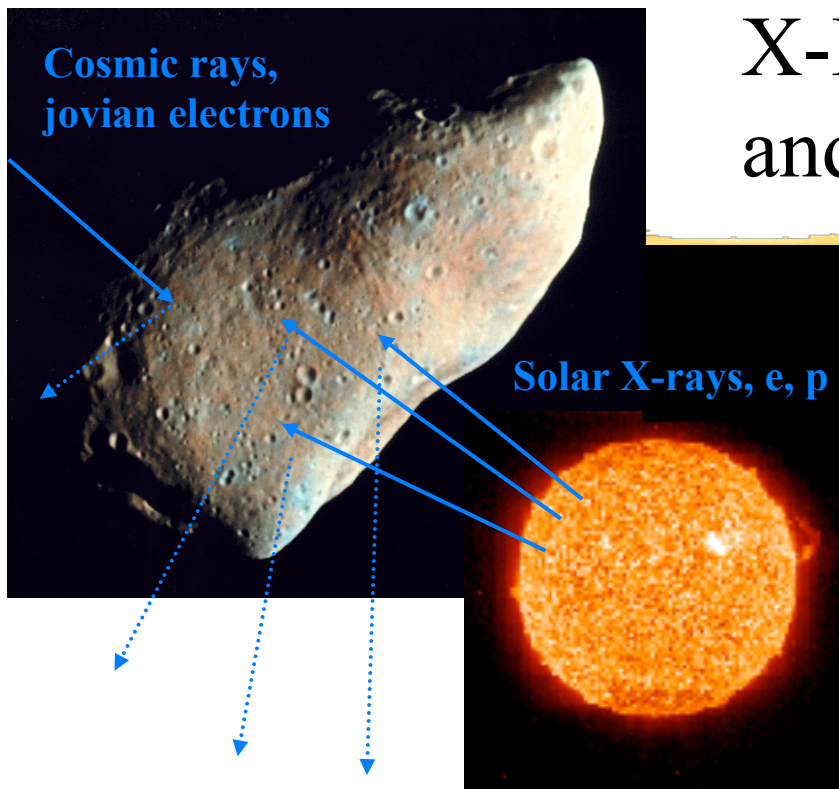
M. Kossov

Simulation 'packages'



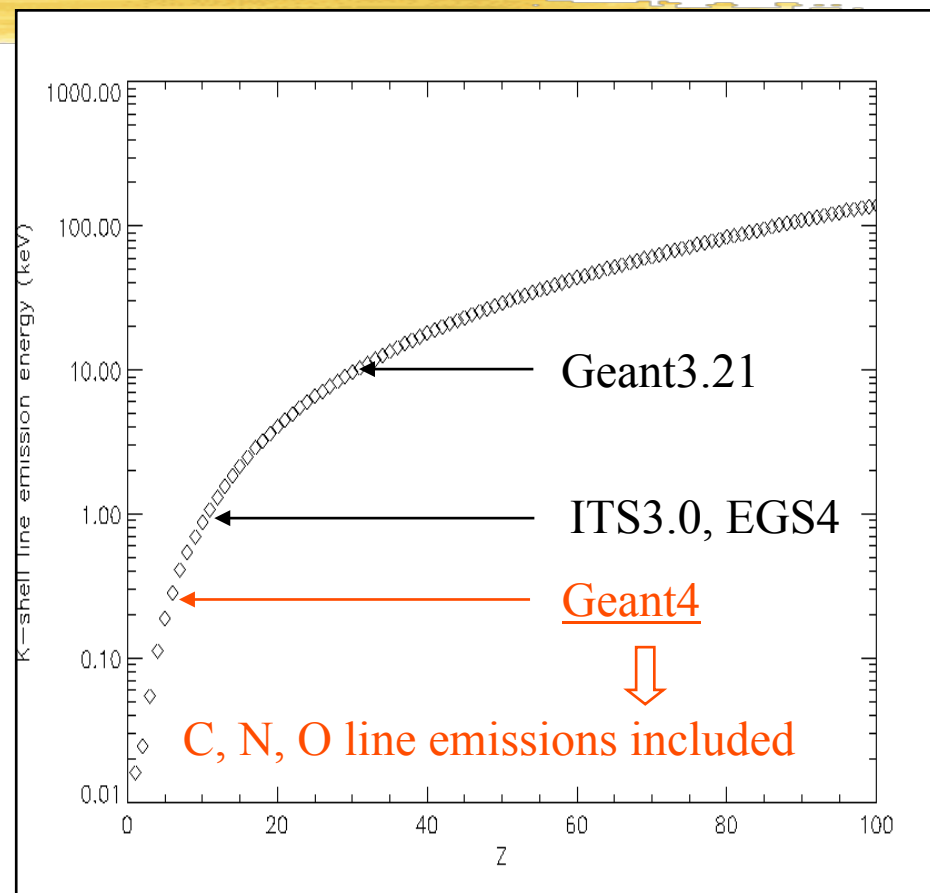
- Provides the means to simulate
 - the **physical processes** and
 - **detector response** of an experiment.
- As was realised by many in the past,
 - **most of the parts** needed can be **common** between experiments (eg physics, geometry blocks) .
- So it makes eminent sense to create and use a **general 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

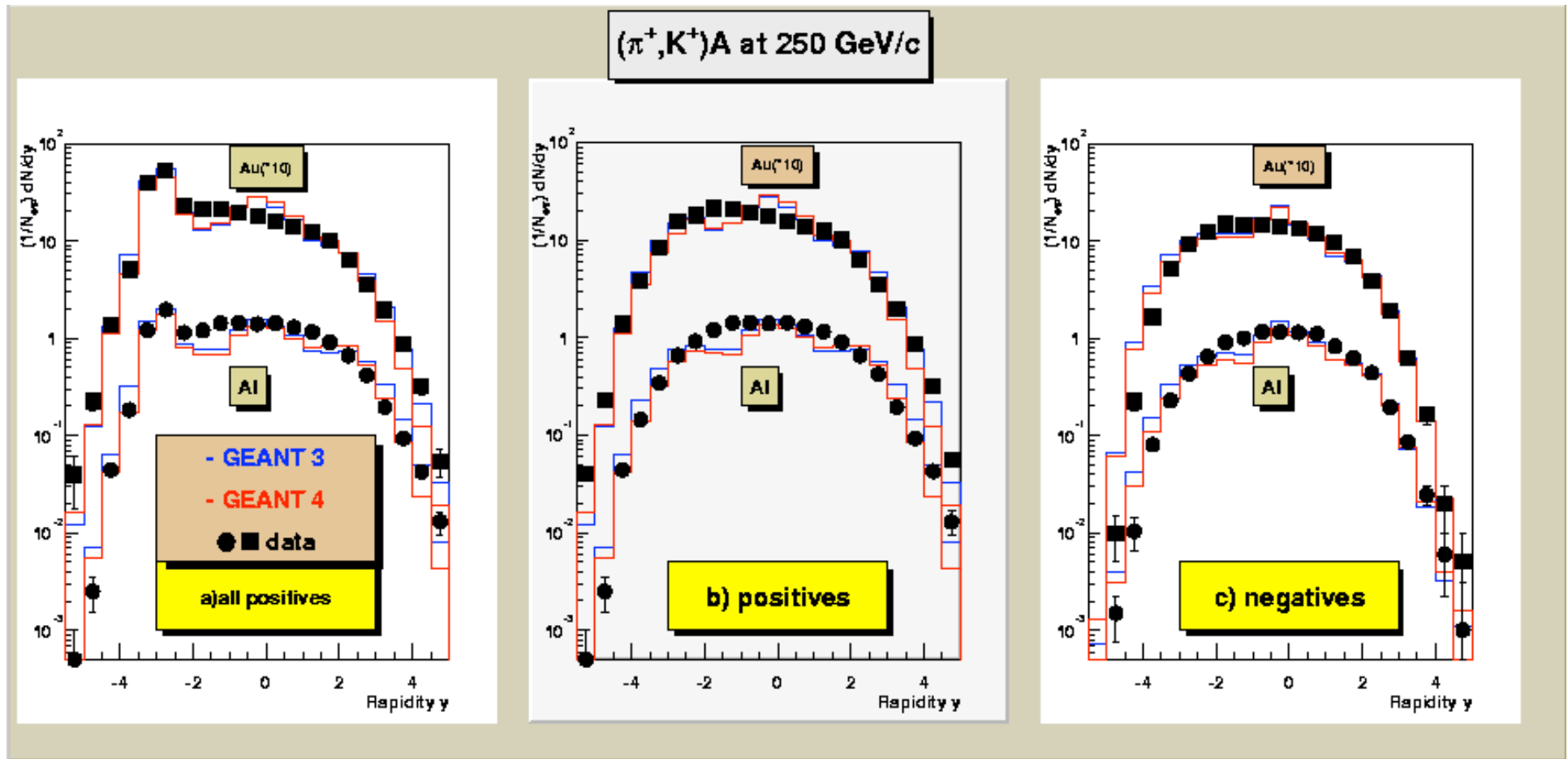


Courtesy SOHO EIT

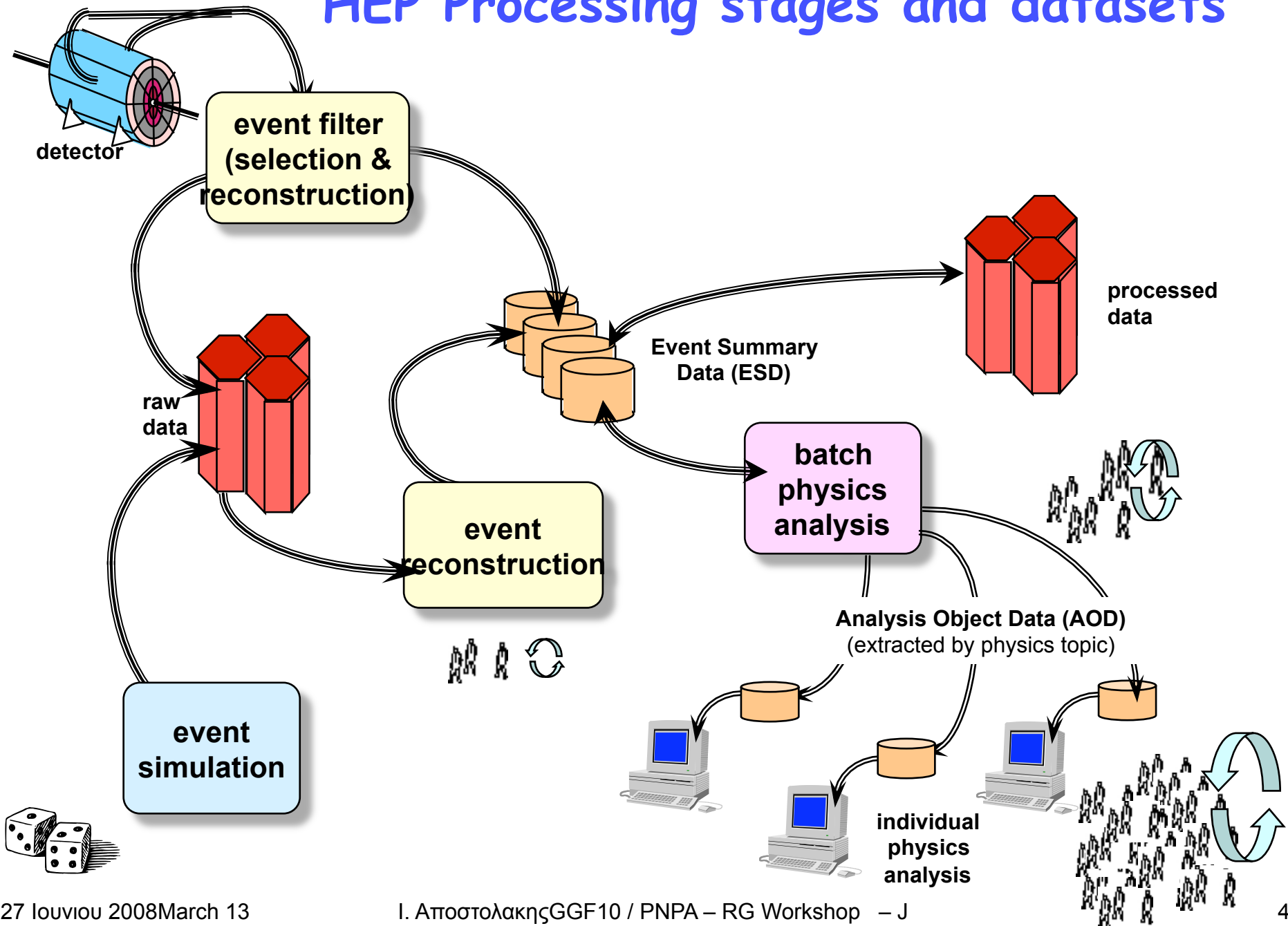
Induced X-ray line emission:
indicator of target composition
(~100 μm surface layer)



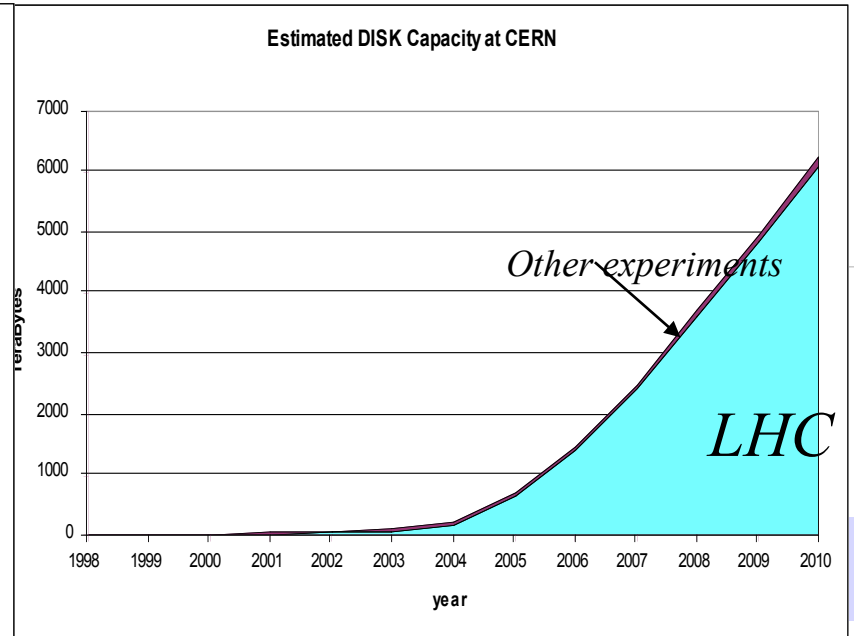
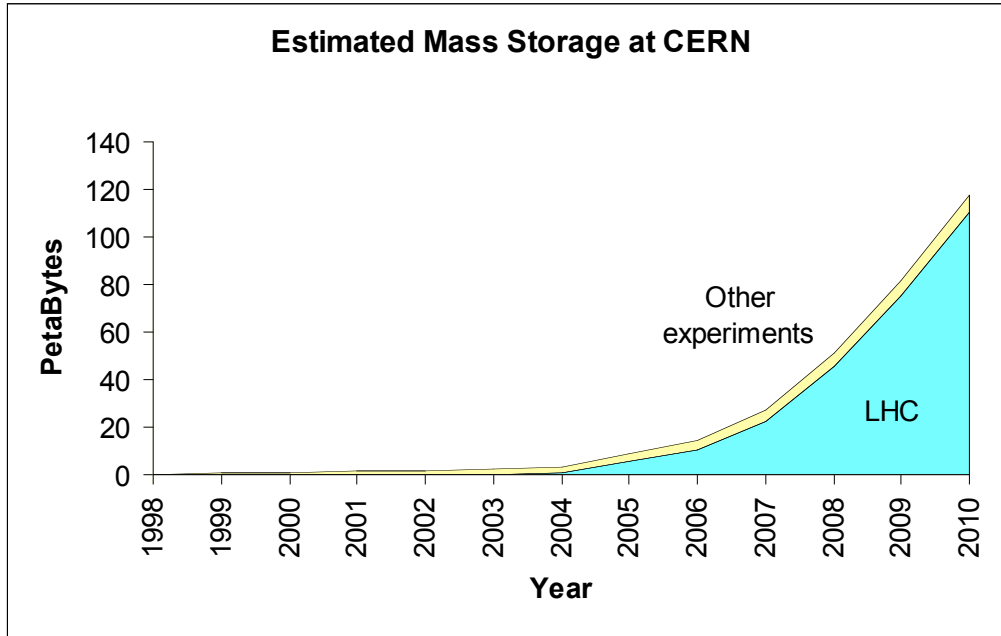
High Energy K,pi on Al, Au



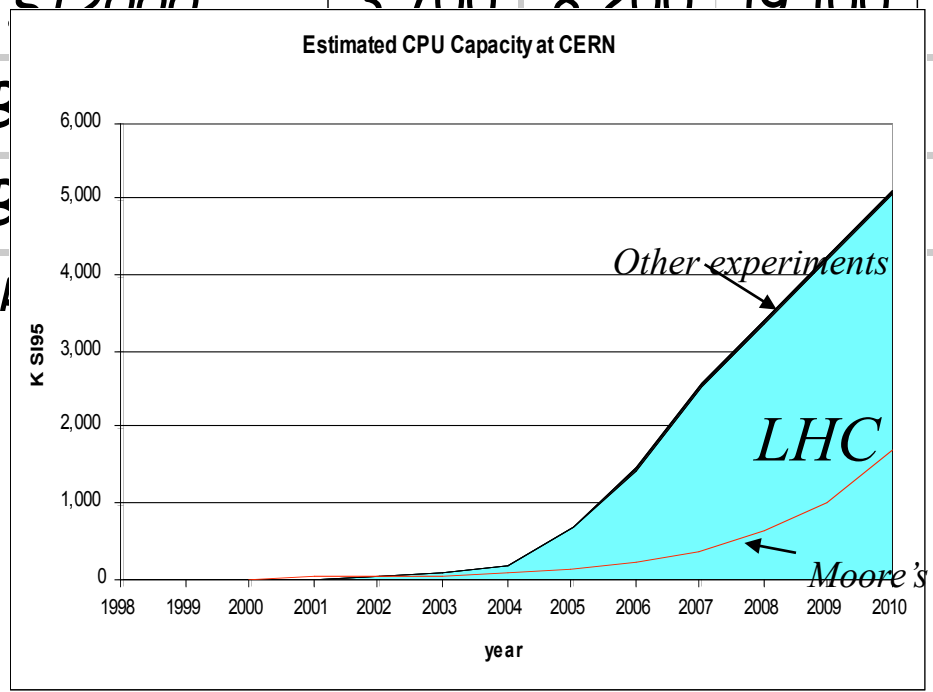
HEP Processing stages and datasets



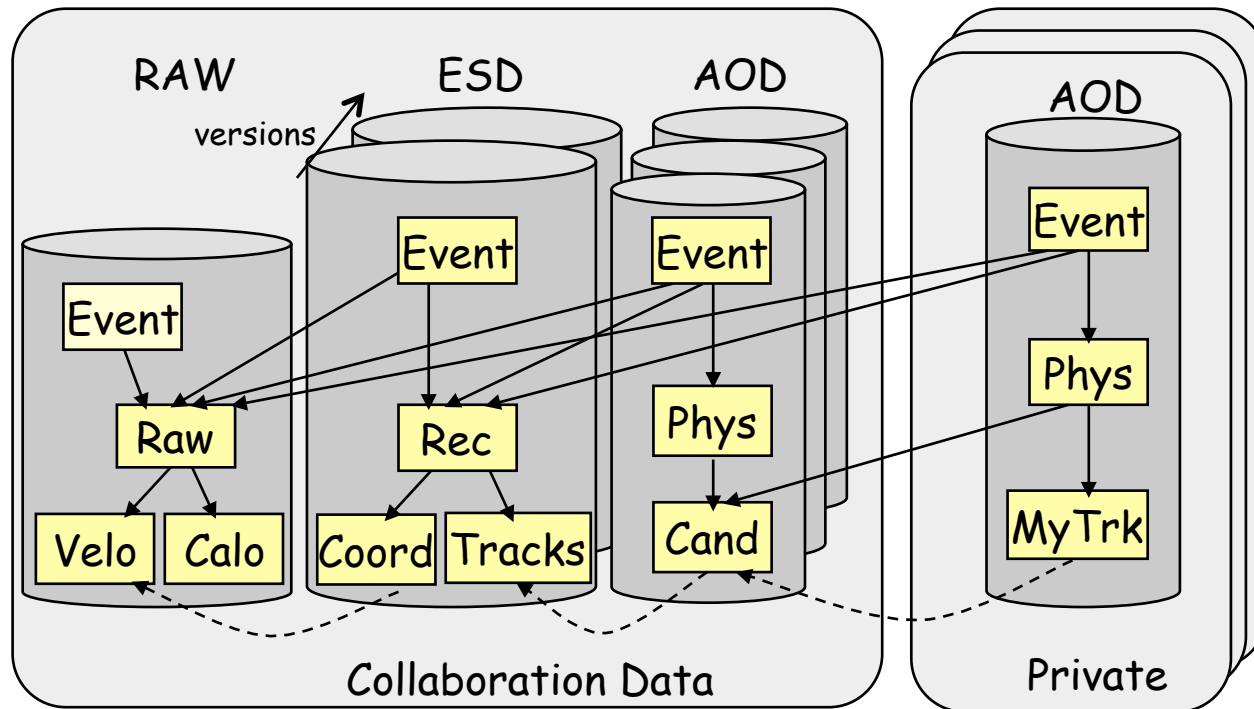
CERN Centre Capacity Requirements for all



processing	K ST2000	3 700	8 200	19 100	25,000	34,000
disk	PE				5.0	6.7
tape media	PE				36	48
tape T/O	GA				39	39

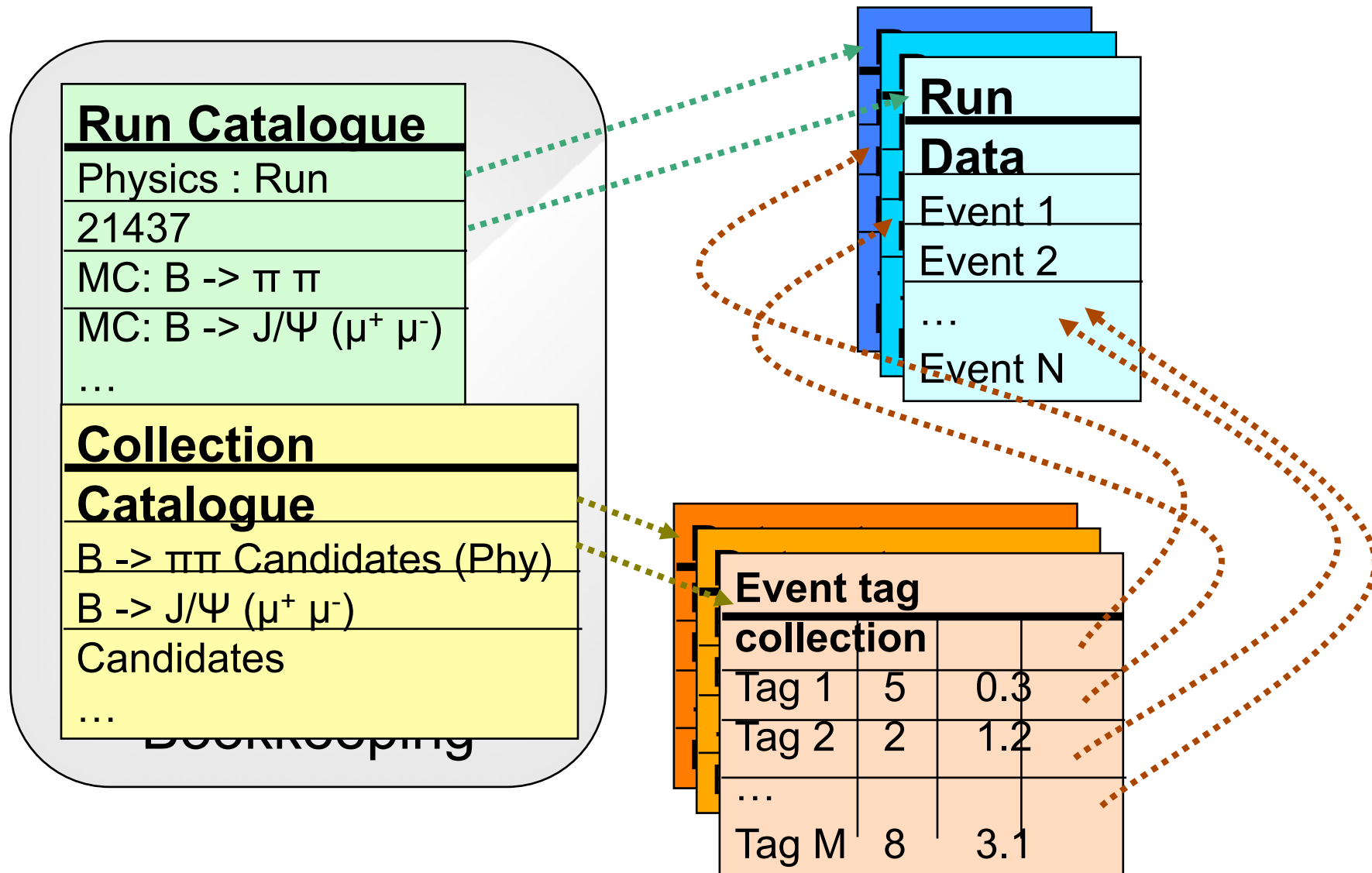


Event Data



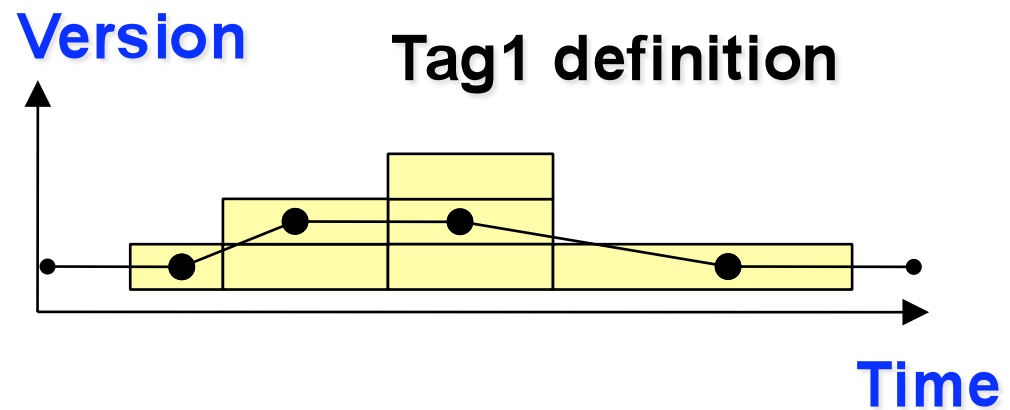
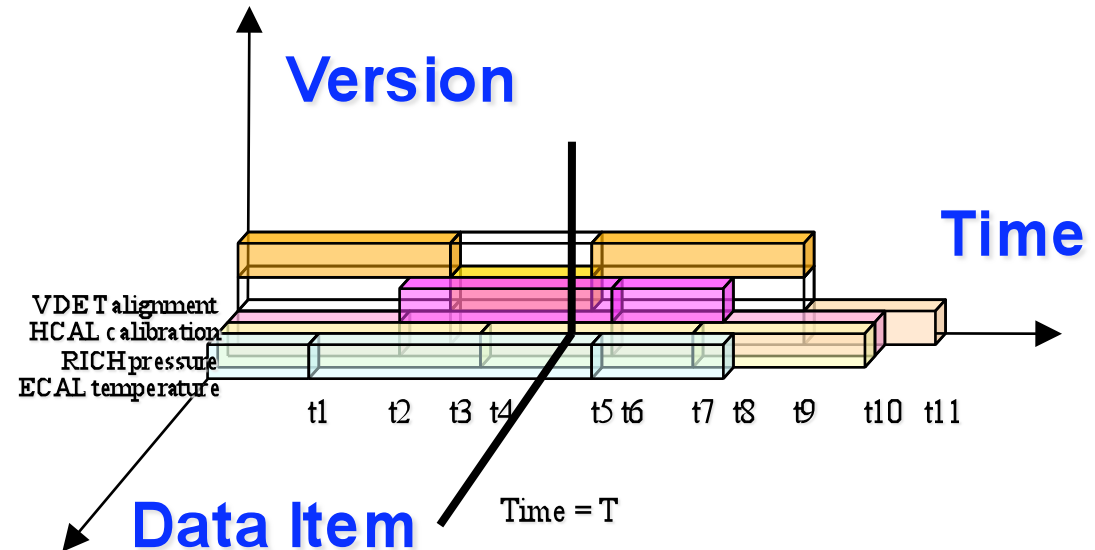
- ❑ Complex data models
 - ~500 structure types
- ❑ References to describe relationships between event objects
 - unidirectional
- ❑ Need to support transparent navigation
- ❑ Need ultimate resolution on selected events
 - need to run specialised algorithms
 - work interactively
- ❑ Not affordable if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

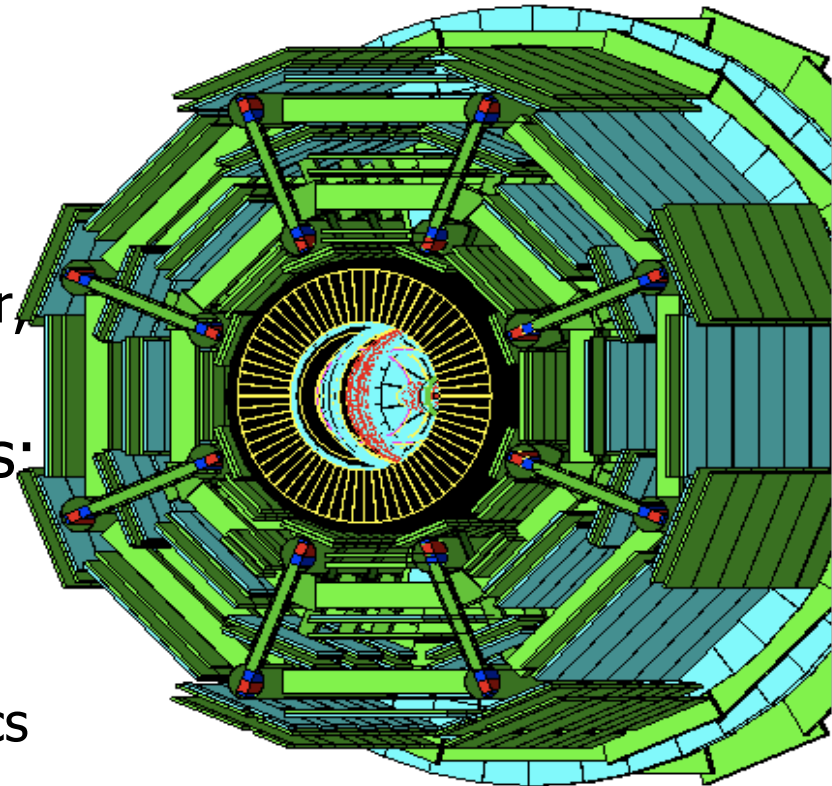


Distributed Analysis - the real challenge

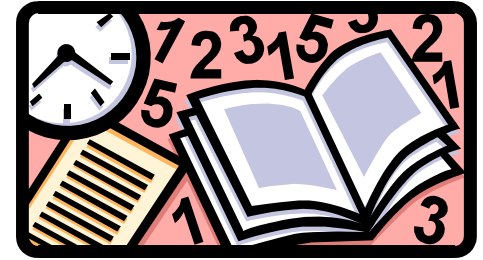
- ❑ Analysis will be performed with a mix of "official" experiment software and private user code
 - How can we make sure that the user code can execute and provide a correct result wherever it "lands"?
- ❑ Input datasets not necessarily known a-priori
- ❑ Possibly very sparse data access pattern when 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').

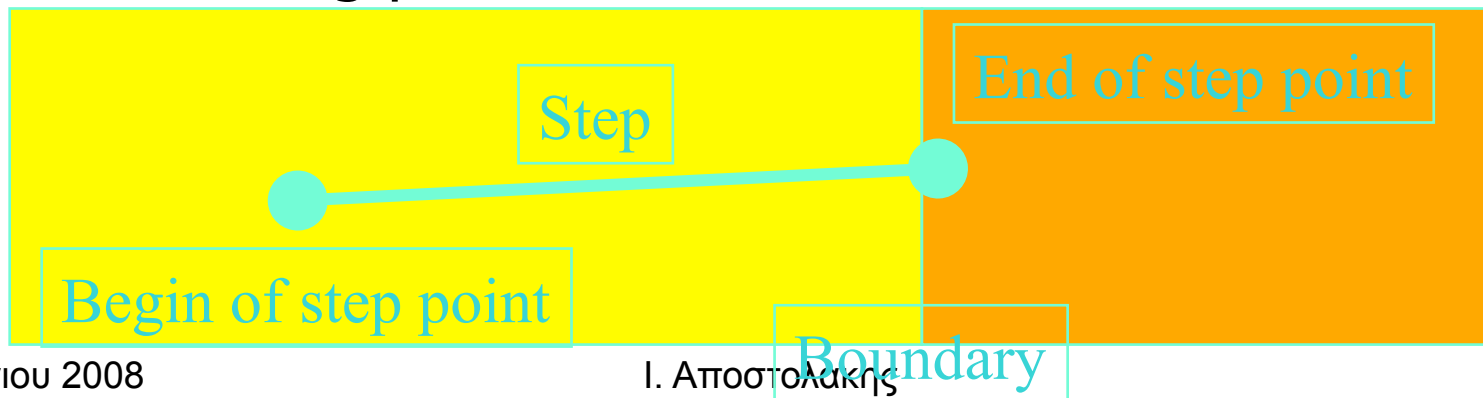
Actions during a Step (cont)

- 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.

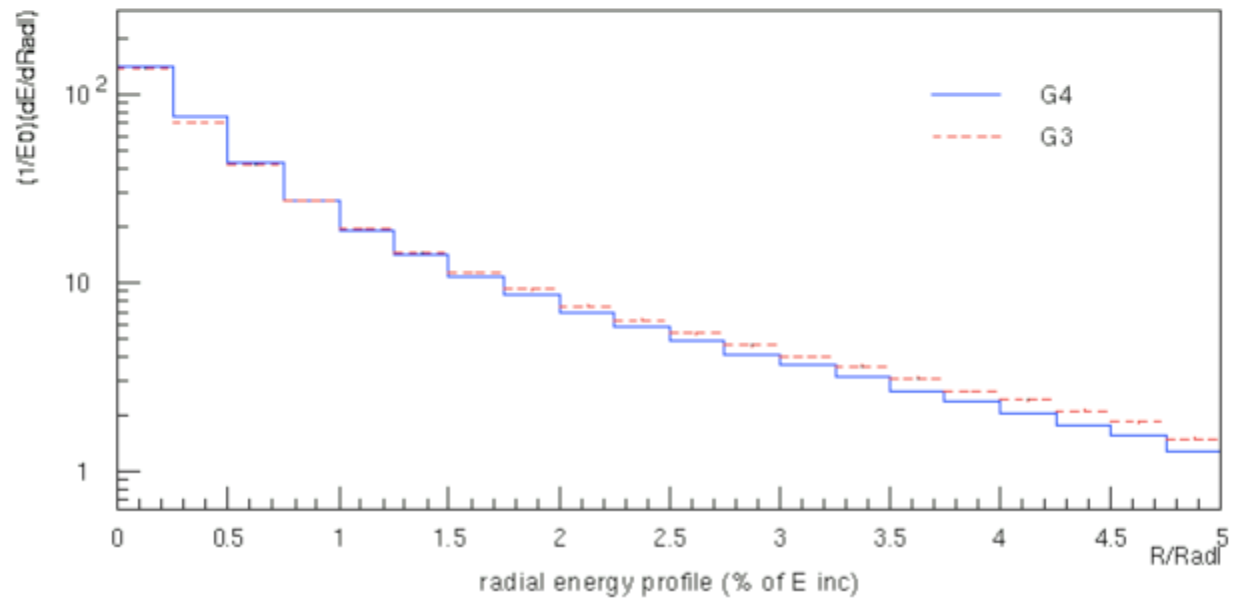
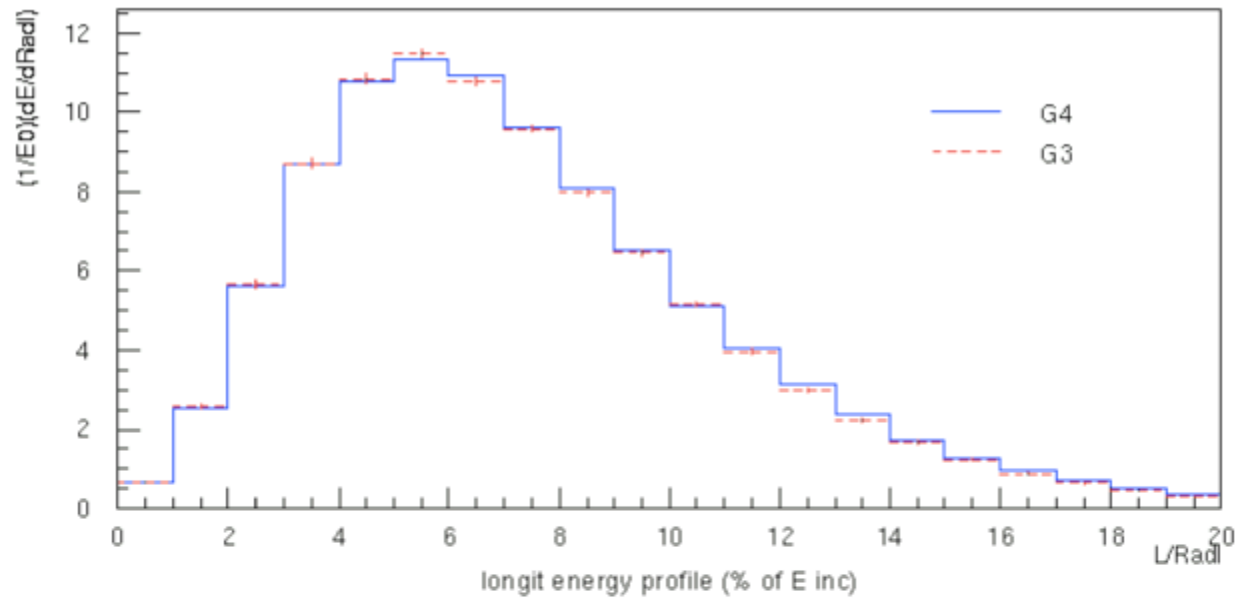


GEANT 4



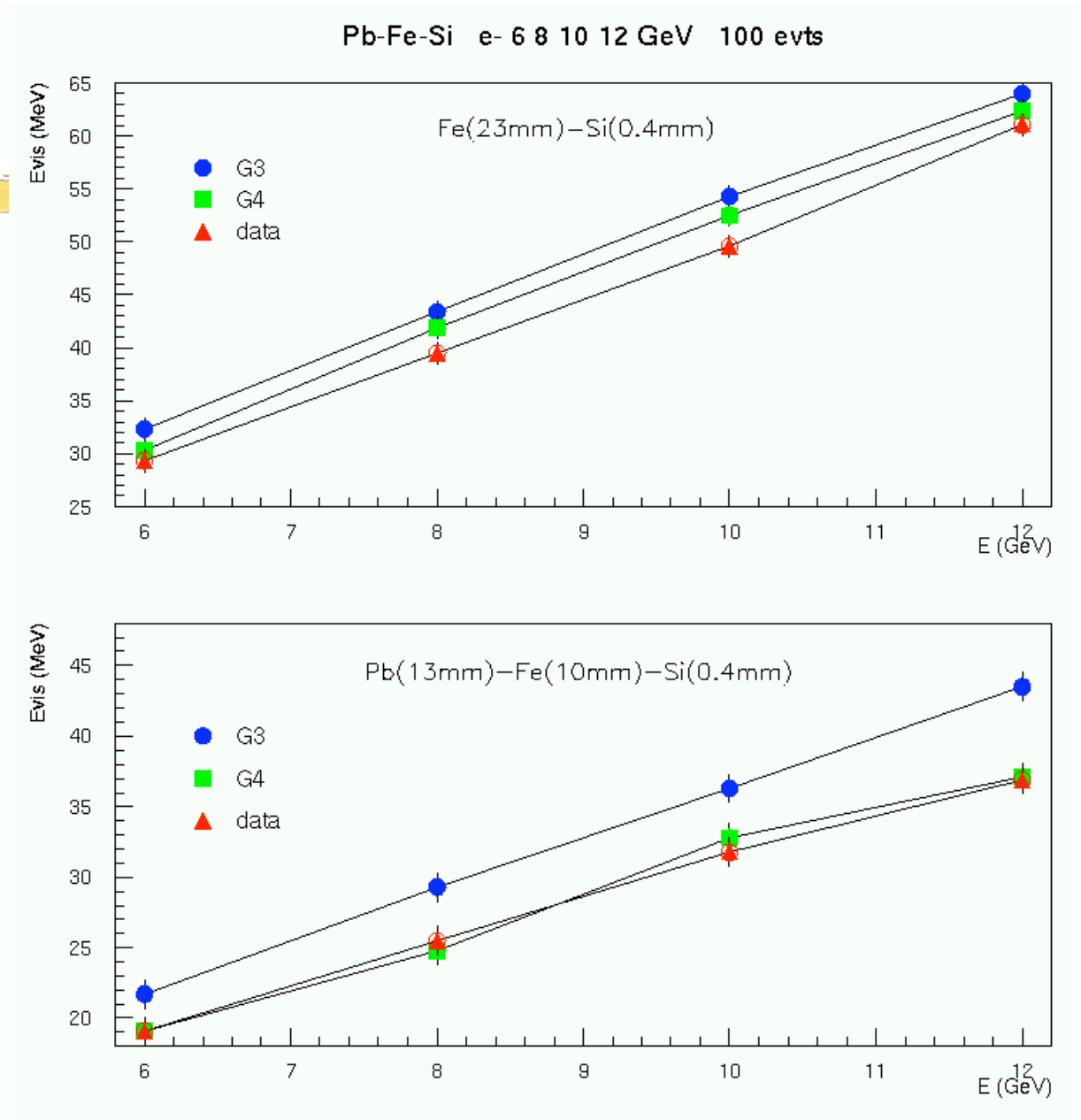
- Detector simulation **tool-kit** for HEP
 - offers alternatives, allows for tailoring
- Software Engineering and OO technology
 - provide the method for building, maintaining it.
- **Requirements** from:
 - LHC
 - heavy ions, CP violation, cosmic rays
 - medical and space science applications
- **World-wide 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



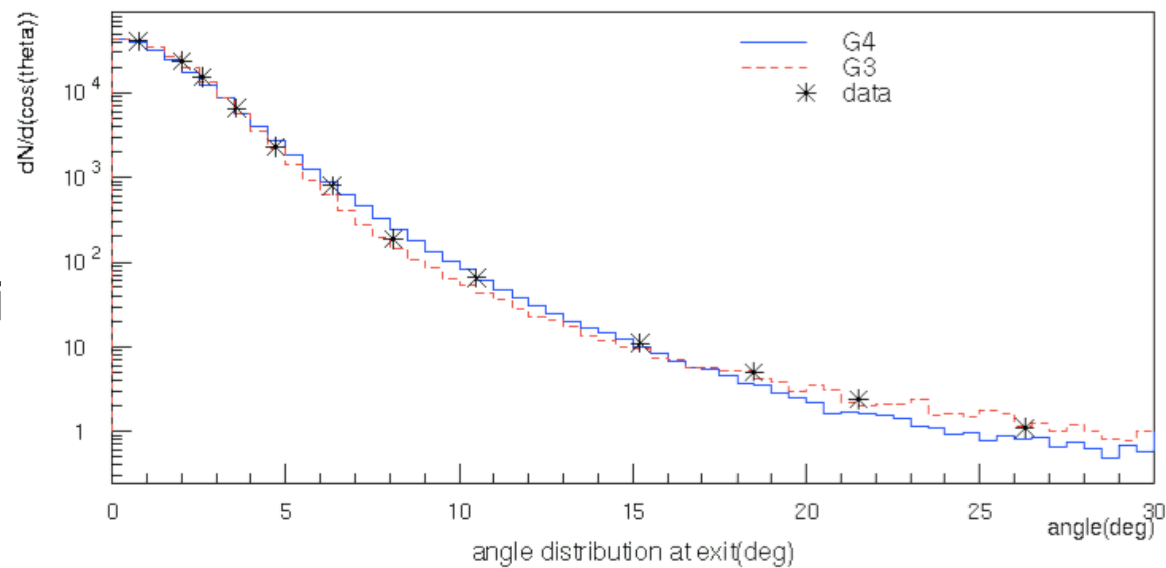
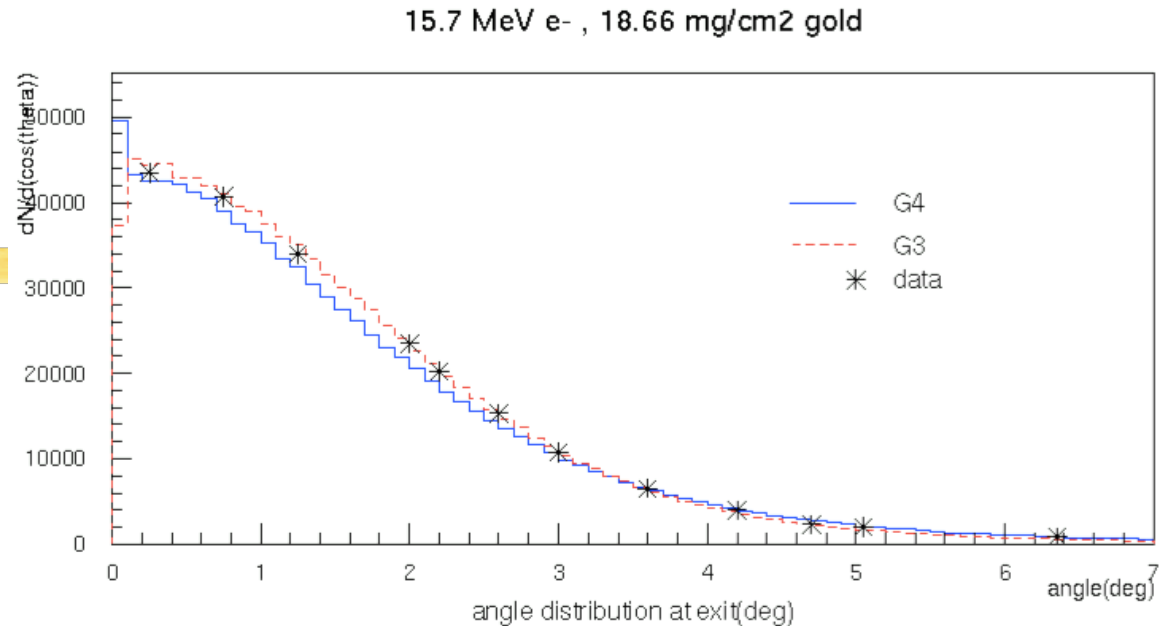
Multiple scattering model

- A new model for multiple scattering based on the Lewis theory is implemented
 - since public β 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/cm² gold foil (8 μm) figure
- 6.56 MeV proton on 93 microns Si
- 70 GeV/c proto



Energy deposit of 1 MeV e⁻ in 0.530 mm Si

