



Οι Υπολογιστες στη Φυσικη Υψηλων Ενεργειων

Ιωαννης Αποστολακηs

CERN

V0.98

2012.08.22

John.Apostolakis@cern.ch

Πλανο της ομιλιας



⌘Η χρηση των Υπολογιστων

▲Ανακατασκευη (reconstruction)

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

▲Προσομοιωση (simulation)

▲Αναλυση δεδομενων (data analysis)

⌘Μεγεθη και το πλεγμα (GRID)

▲Υπολογιστικες αναγκες και ... GRID

▲Άλλες εφαρμογες του πλεγματος

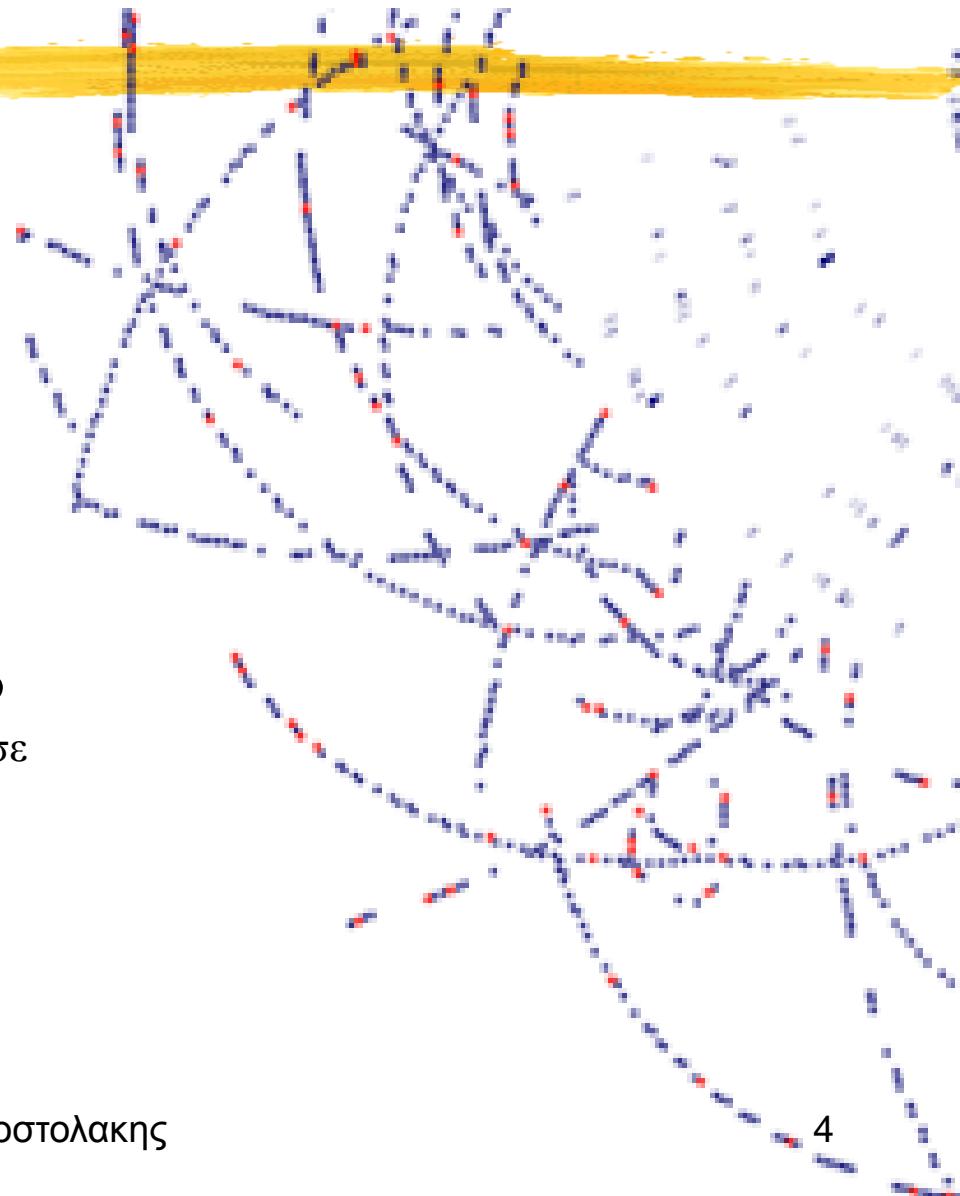
Ανακατασκευη



Μια γρηγορη εισαγωγη

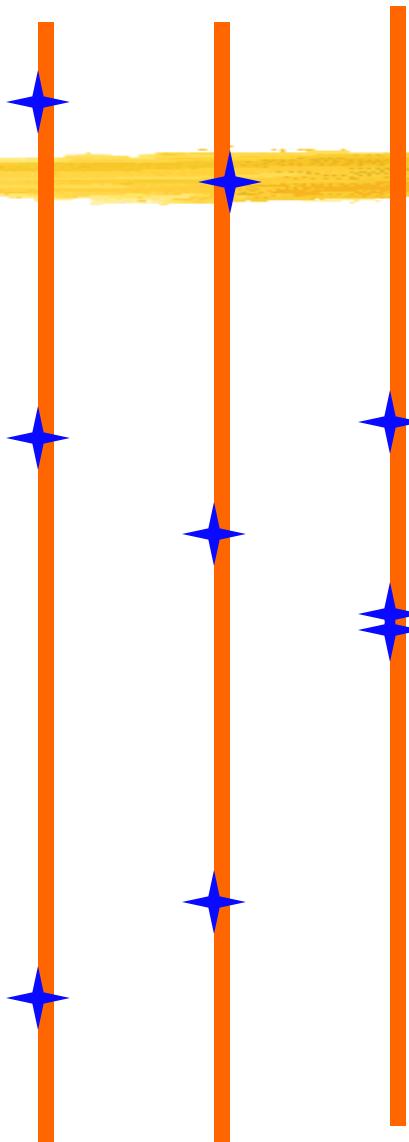
Τι είναι η ανακατασκευή?

- ⌘ Οι μετρησεις είναι σαν ενας γριφος
 - ☒ Τι τροχιες τις προκαλεσαν;
- ⌘ Καθε μετρηση θεσης βοηθαει
 - ☒ 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



Ανακατασκευη στην πραξη

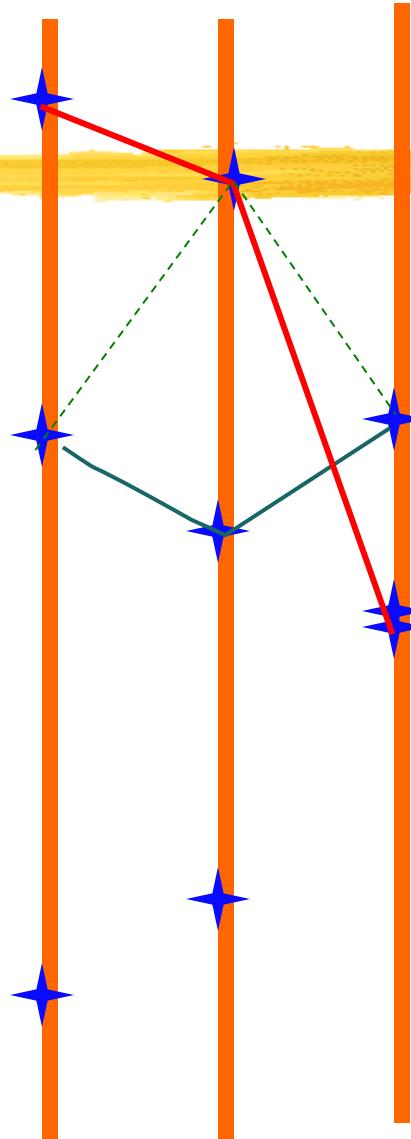
- Αρχιζει με τις θεσεις διαβασης των σωματιδιων



Μαγνητικο
Πεδιο
B

Ανακατασκευη στην πραξη

- Αρχιζει με τις θεσεις διαβασης των σωματιδιων
- Δωκιμαζονται διαφοροι συνδιασμοι
 - και υπολογιζεται η διαφορα μετρησης προβλεψης
 - Και ετσι πιθανοτητα του καθε συνδιασμου
-

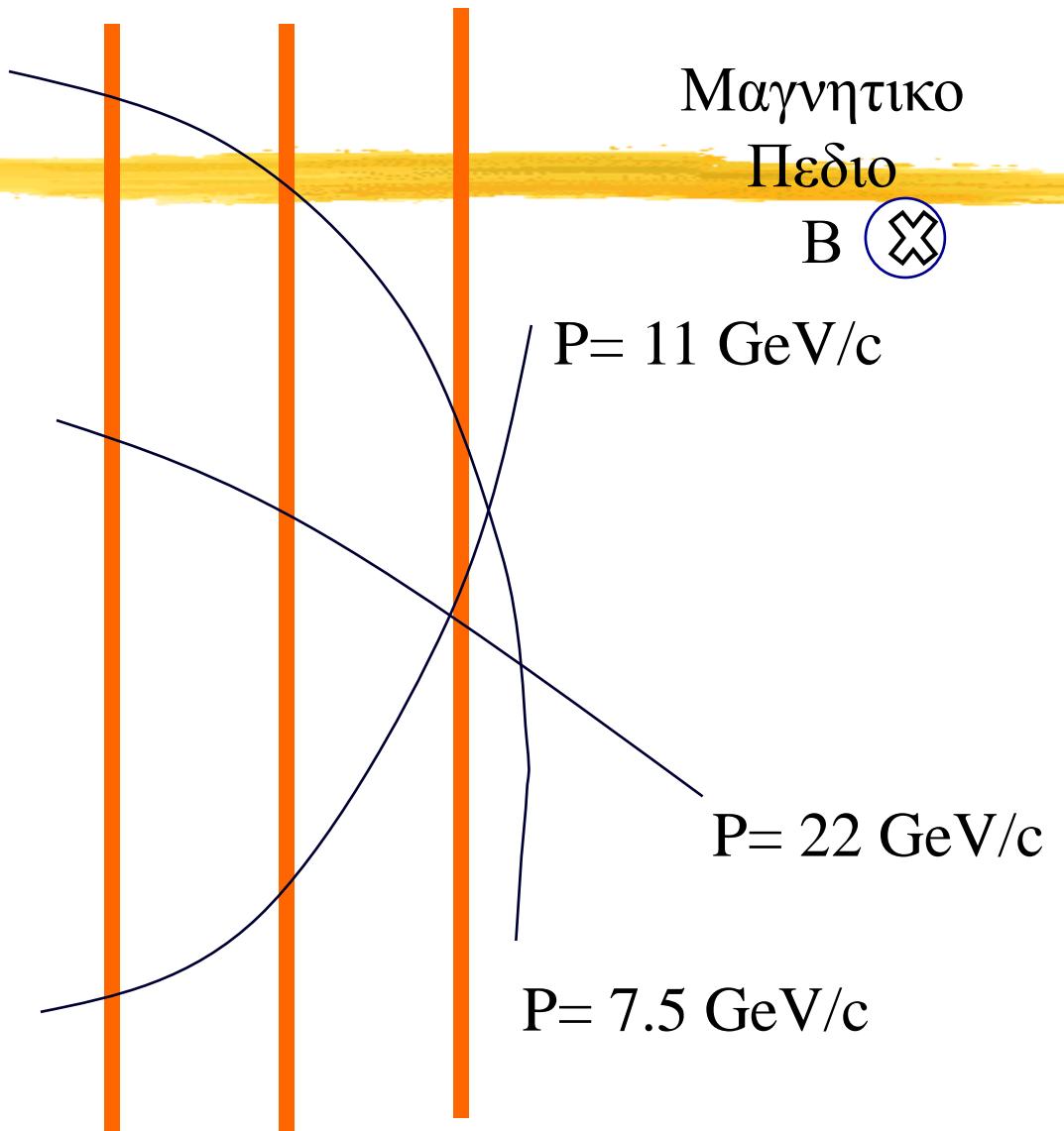


Μαγνητικό Πεδιο
B

Αλγοριθμοι
Φιλτρο Καλμαν
(Kalman filter)

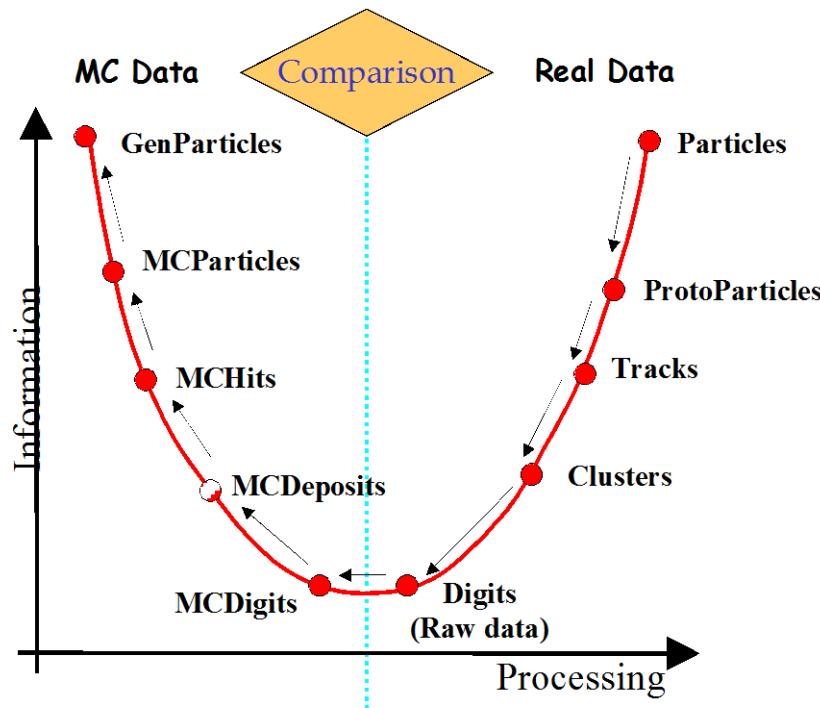
Ανακατασκευη: αποτελεσμα

- Αρχίζει με τις θεσις διαβασης των σωματιδιων
- 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





Data Organisation



Προσομοιωση και Ανιχνευτες



Τι ειναι η προσομοιωση
Γιατι υπαρχει
Πως γινεται

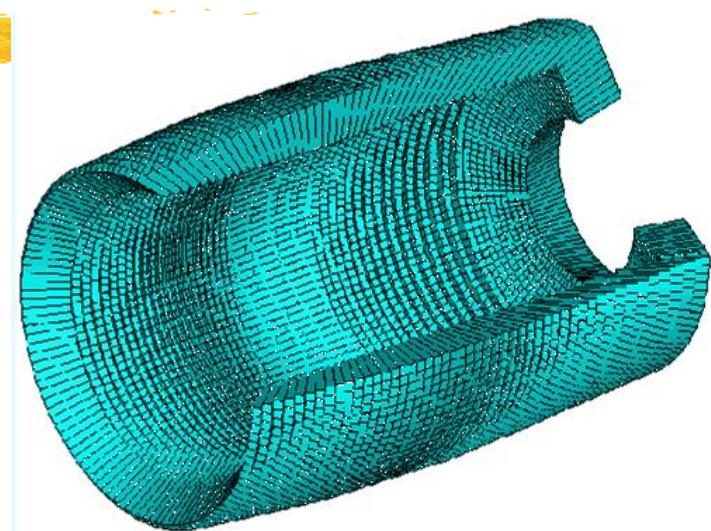
Οι σημερινοί ανιχνευτές

⌘ Πολλα τμηματα

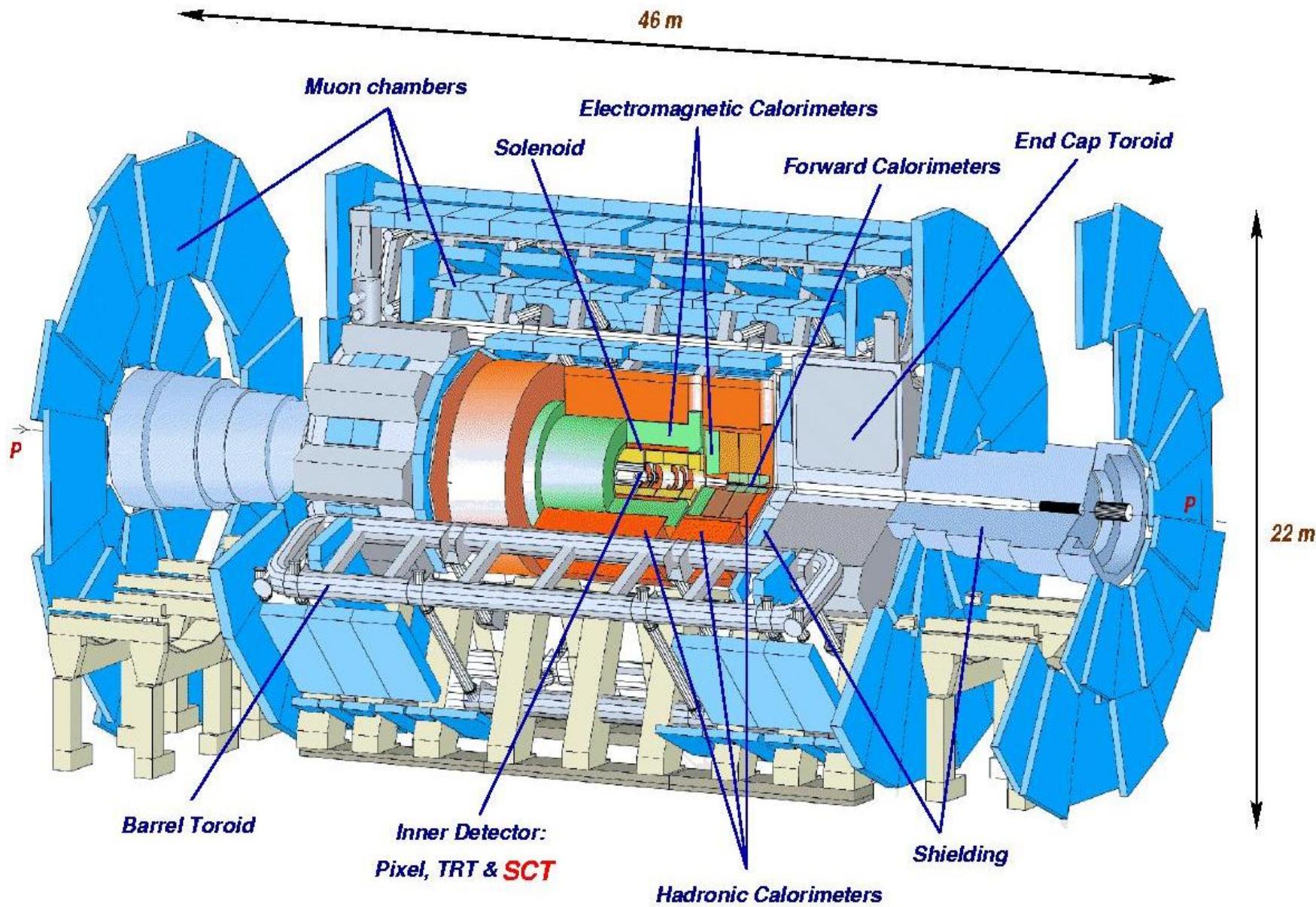
- └ Διαφορετικες αναγκες
 - └ Μετρηση θεσης (τρακερ - trackers)
 - └ Μετρηση ενεργειας (θερμιδομετρα)

⌘ Λογω της πολυ-πλοκοτητας

- └ οι πιο πολλες μελετες χρειαζονται πολλα υπολογιστικα εργαλεια



ΟΙ ΑΝΙΧΝΕΥΤΕΣ ΤΟΥ ΑΤΛΑΣ



Τι είναι προσομοιση?

⌘ Φτιαχνουμε μοντελα

↗ Του ανιχνευτη

☒ Γεωμετρια

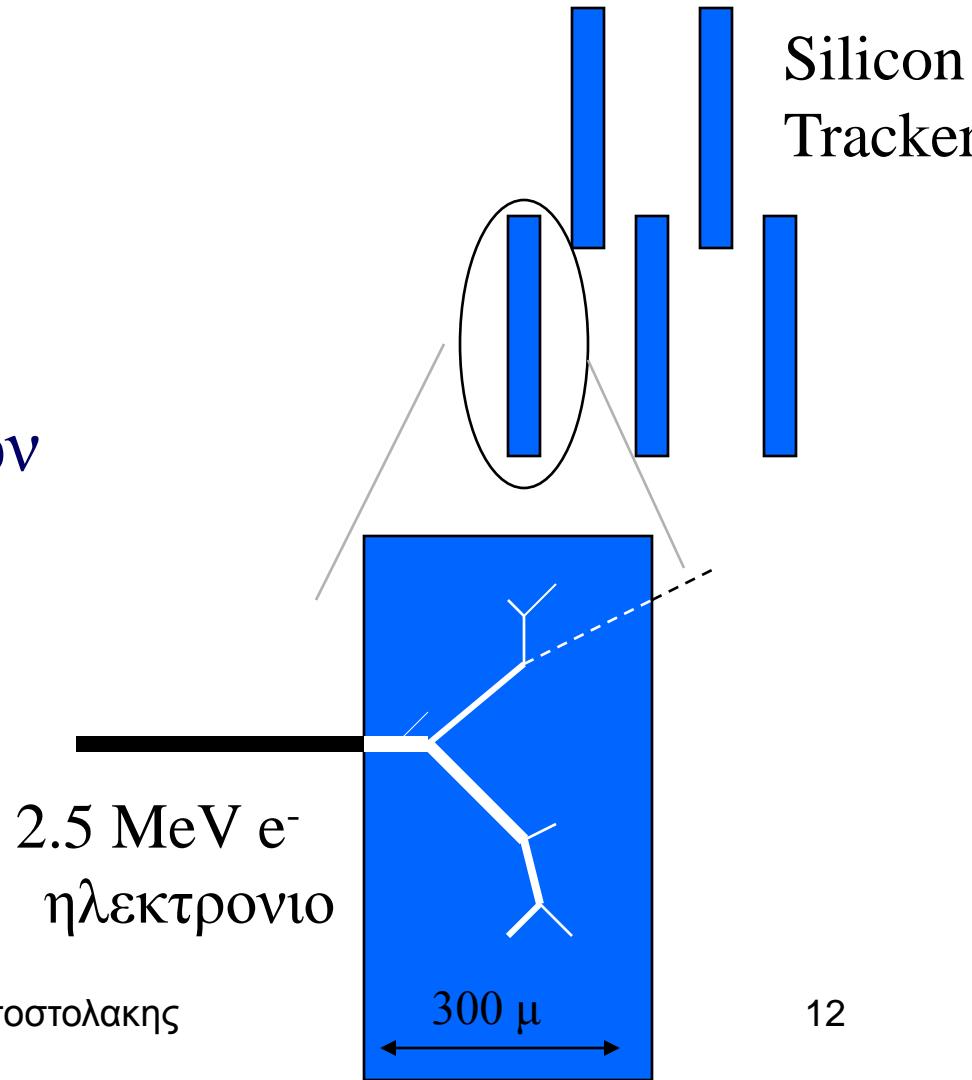
☒ Υλικα

↗ Των αλληλεπιδρασεων

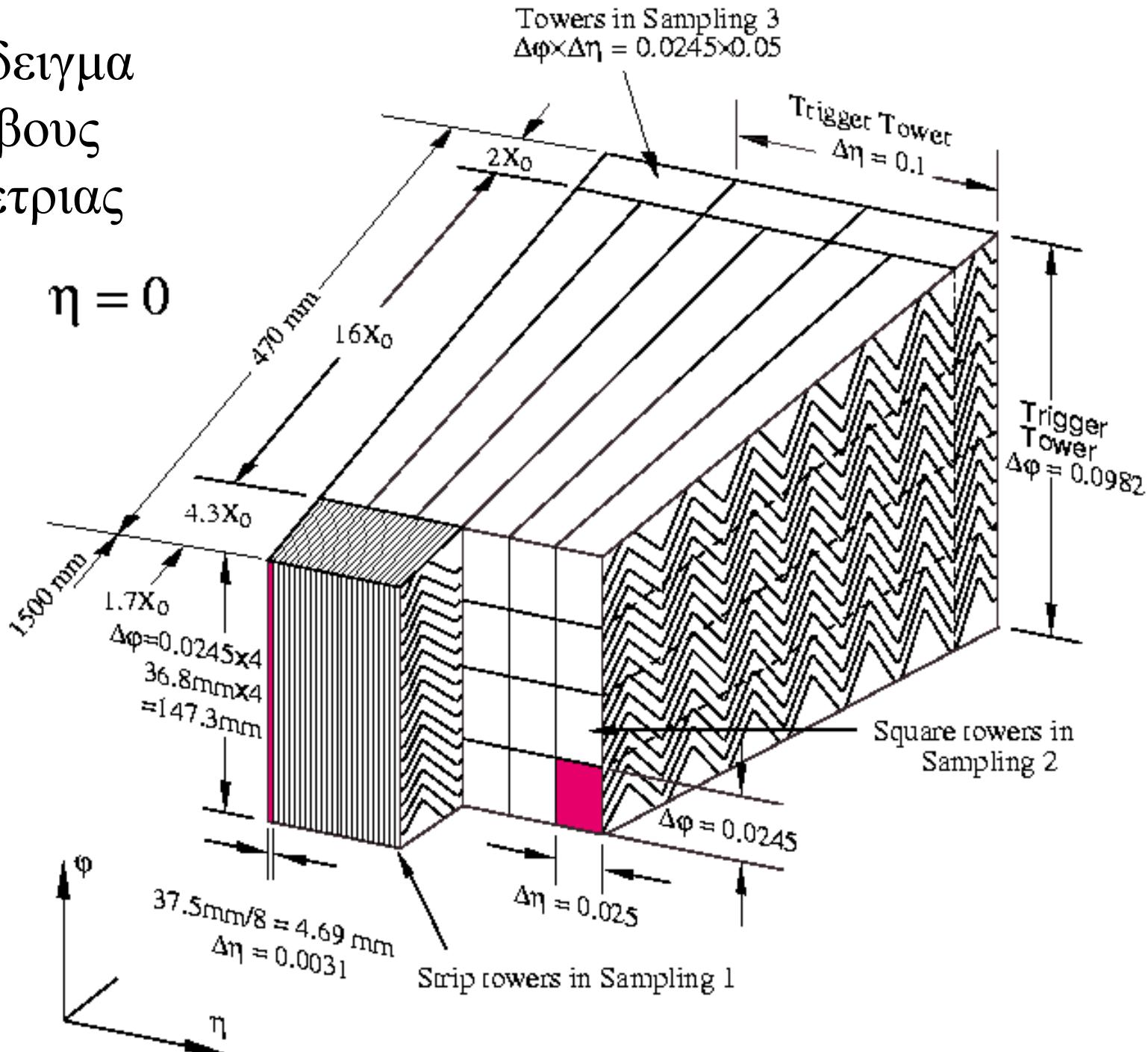
☒ Καθε γνωστου τυπου

- Ηλεκτρομαγνητικου
- Υσχηρου πυρηνικου

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



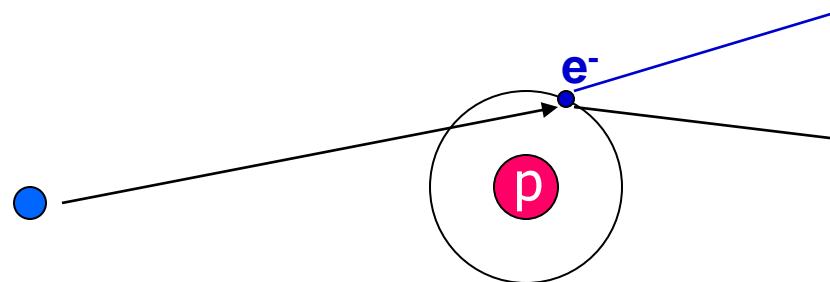
Παραδειγμα ακριβους γεωμετριας



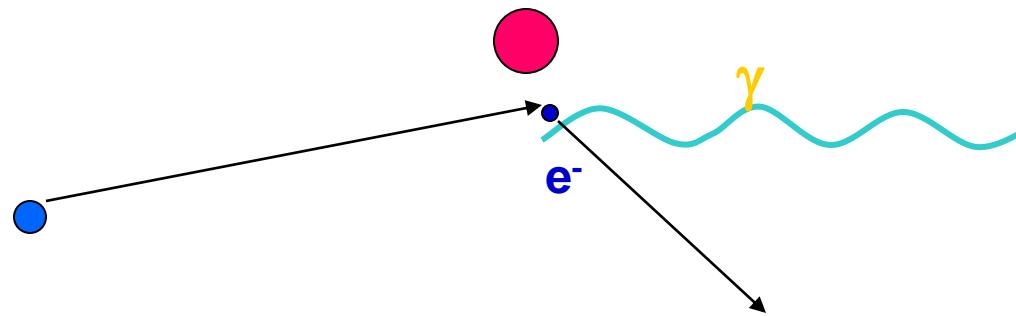
Βασικές Αλληλεπιδρασεις

- ⌘ Οι διαφορες αλληλεπίδρασης του σωματιδίου με το υλικό (τμημα του ανιχνευτή ή αλλο)
 - ↗ Παραγωγή δευτερευοντος σωματιδιου

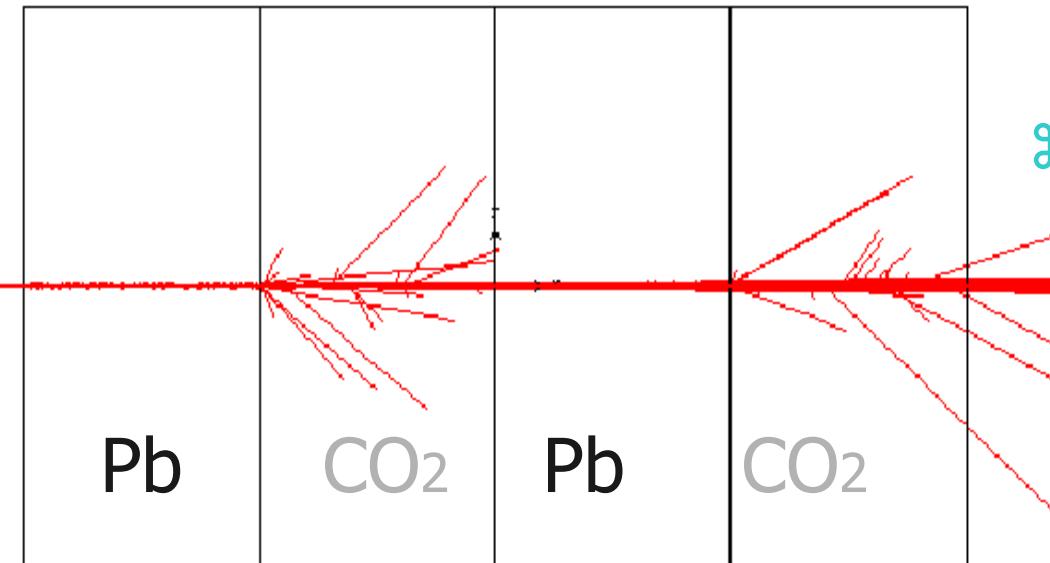
☒ Ιοντισμός



☒ Bremstrahlung

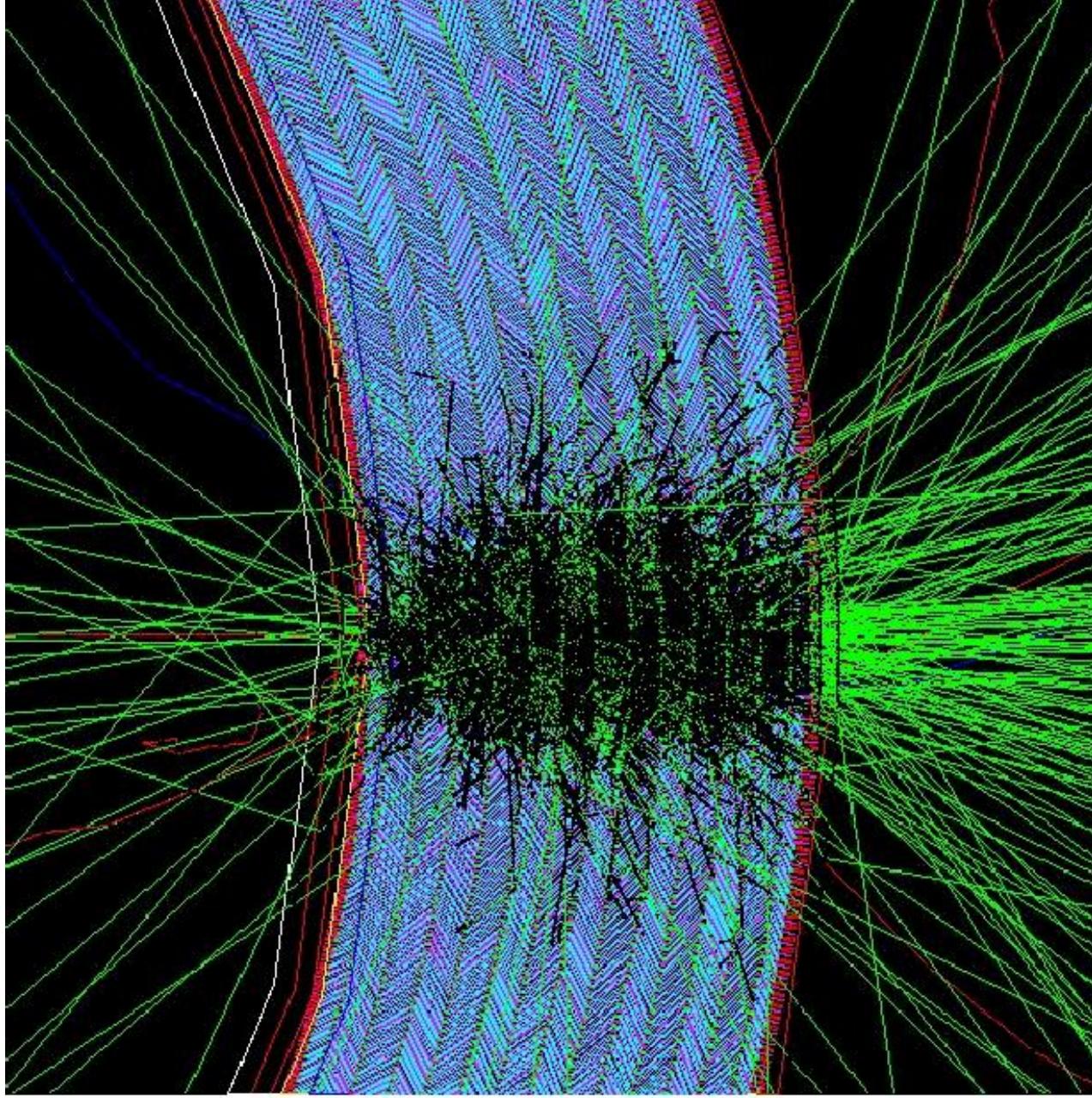


Ενα απλό παραδειγμα



- ⌘ Στο μολυβδο παραγονται πολλα δευτερευοντα σωματιδια
 - ↗ Τα περιστερα μενουν κοντα,
 - ↗ μερικα ξεφευγουν.
- ⌘ Το διοξειδιο του ανθρακα, σαν αεριο, εχει μικρη πυκνοτητα
 - ↗ Οσα σωματιδια φτανουν η παραγωνται, πανε μακρια
 - ↗ Παραγονται λιγοτερα

GEANT 3

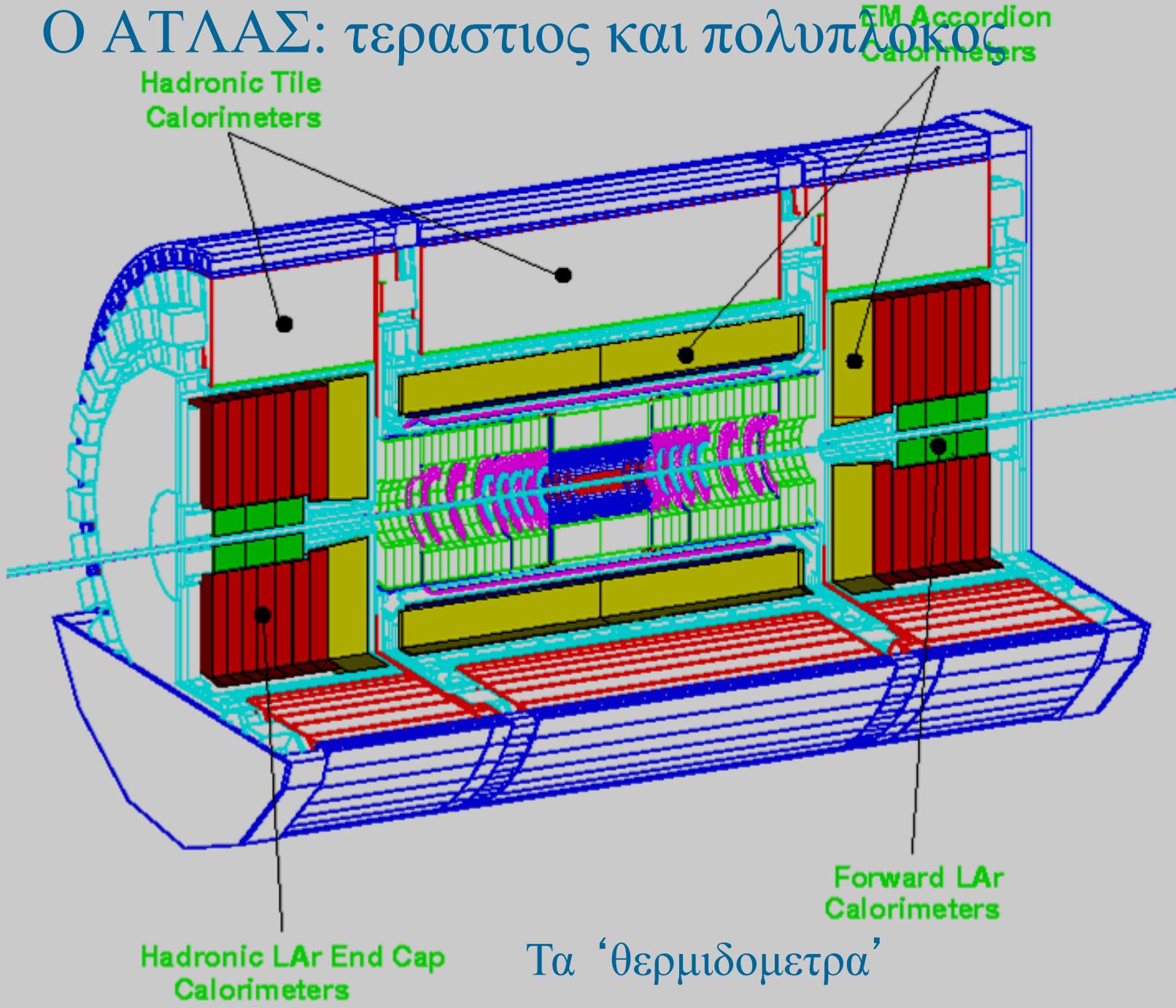


25 Au

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

16

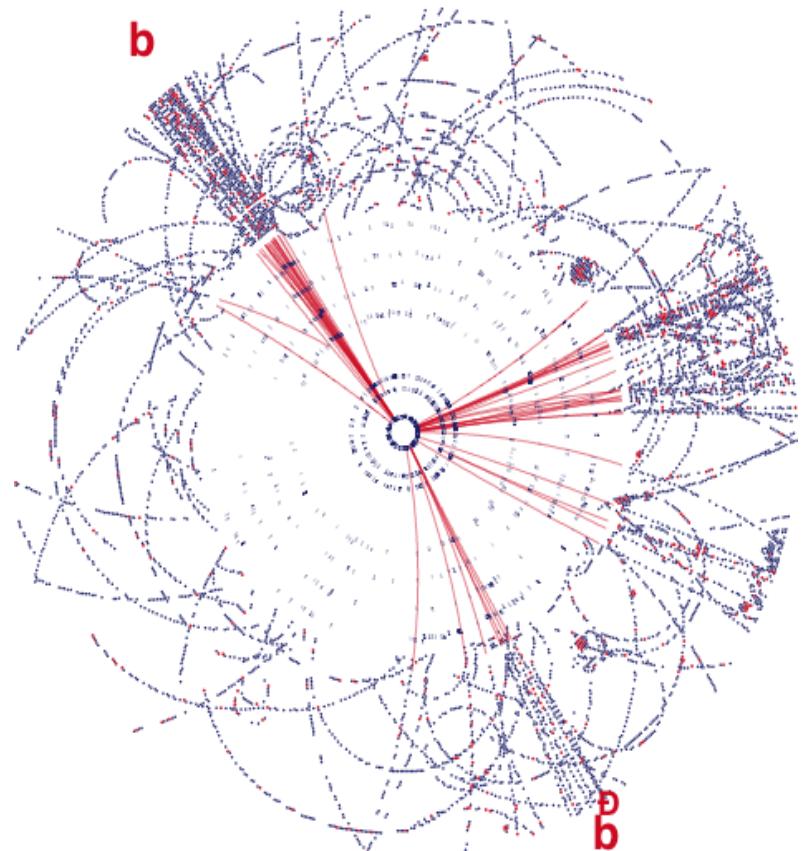
Ο ΑΤΛΑΣ: τεραστιος και πολυπλοκος



Atlas : Physics Signatures and Event Rates

- ❑ Οι δεσμες διασταυρονονται 40 MHz
- ❑ $\sigma_{inelastic} = 80 \text{ mb}$
 - Σε καθε περασμα πολλες συγκρουσεις (αυξουσα μεση τιμη. φετος ~ 25)
- ❑ Διαφορετικοι στοχοι, ο καθενας με τη δικια του «υπογραφη»
 - Το Χιγκς (Higgs) μεσονιο
 - Υπερσυμμετρια (Supersymmetry)
 - Το αγνωστο
 - Οι συμμετριες στα B μεσονια
- ❑ Το καθε καναλι χρειαζεται την προσομοιωση του
- ❑ Τα ενδιαφεροντα συμβαντα ειναι καρφιτσες στα αχυρα σε ενα χωριο γιοματο σταβλους (~ 1 in $10^5 - 10^9$)

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



Γιατί προσομοιωση ?



- ⌘ Για να σχεδιασουμε τους ανιχνευτες
- ⌘ Για να ετοιμασουμε τις μεθοδους ανακατασκευης
- ⌘ Για να καταλαβου-με τον ανιχνευτη

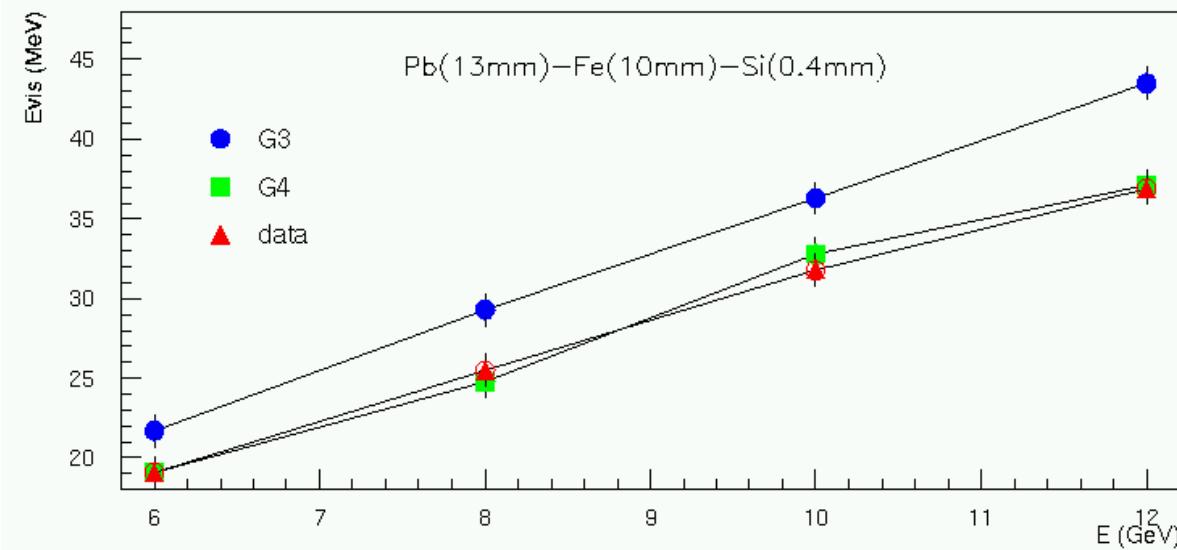
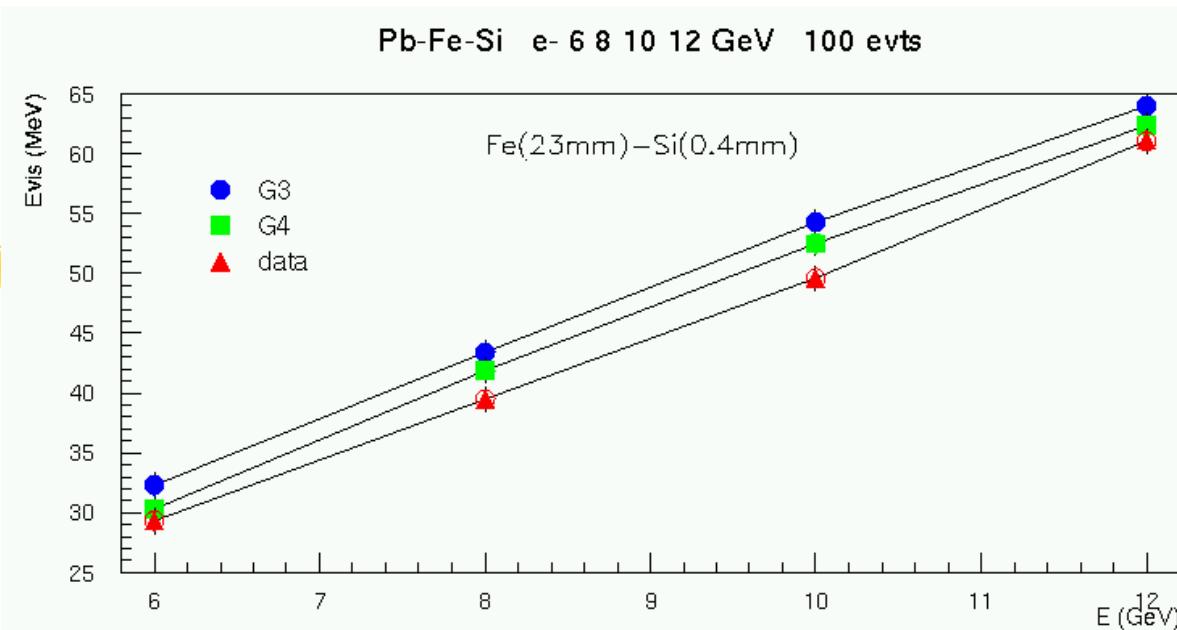
Παραδειγμα σύγκρισης

- ⌘ Δειγματοληπτικό θερμιδομετρό
 - visible energy

⌘ tests

- all EM processes for e-, e+ and photon

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



Υπαρχει απλη λυση ?

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

☒ Για μερικες απλες ερευνης, Ναι

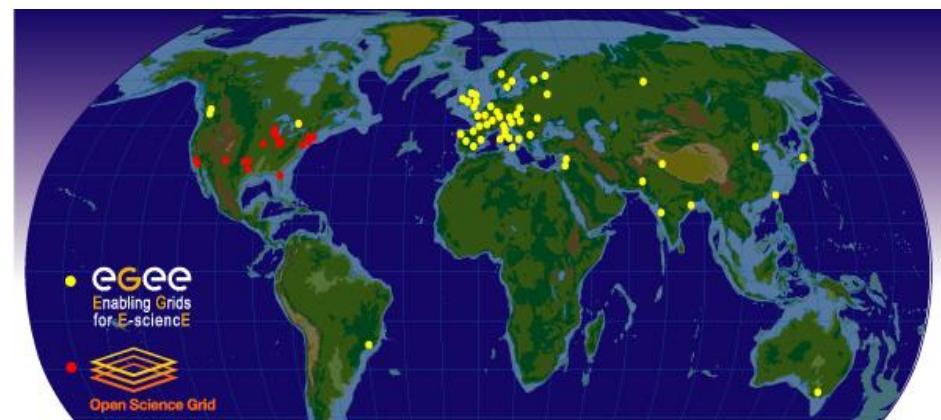
☒ Για πολλες (τις περιστερες) χρειαζεται ολη η εικονα

☒ Υπαρχει κατανομή τιμων

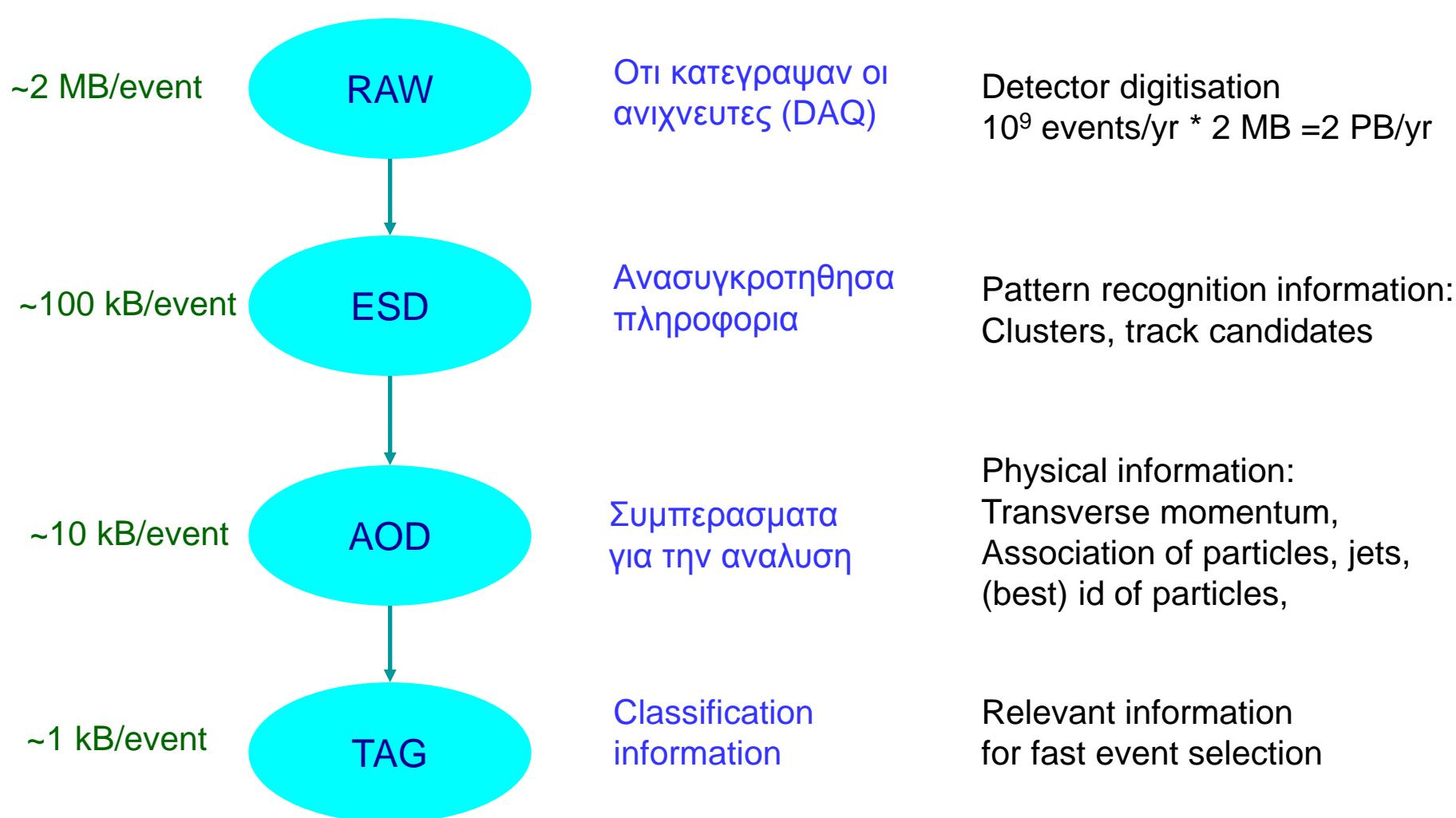
- Που δεν ειναι παντα συμμετρικη η απλη
- Οι ουρες των κατανομων μπορουν να παιξουν μεγαλο ρολο

⌘ Με ποιο σφαλμα ξερετε την ταδε ενεργεια?

Αναλυση δεδομενων



Ιεραρχεία Δεδομένων (Data)





Distribution of Computing Services

Summary of Con

All experiments - 2008

From LCG TDR - June 2009

CPU (MSPECint2000s)

Disk (PetaBytes)

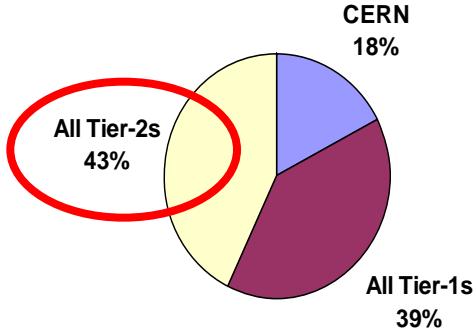
Tape (PetaBytes)

Τα νεα δεδομένα θα
μεγαλωνουν 15 PetaBytes
καθε χρονο – με δυο αντιγραφα

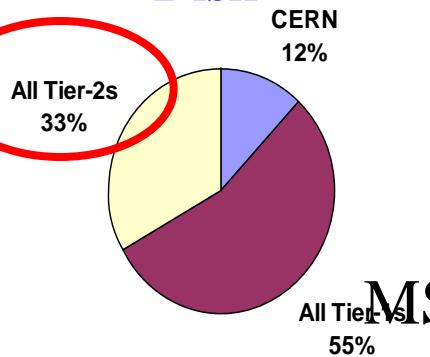
Περιπου 100,000
'πυρήνες'cores

	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

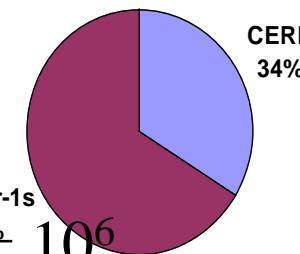
CPU



Disk



Tape



MSPECint = 10^6

Μεγαλο ποσωστο των υπολογιστων και
δισκων ειναι διαδεδομενα

σε 120 υπολογιστικα κέντρα



Λυση : το Πλεγμα (Grid)

- Χρήσιμοποιούμε το Πλεγμα να ενώσουμε τους υπολογιστικούς πόρους των ινστιτούτων ανά τον κόσμο

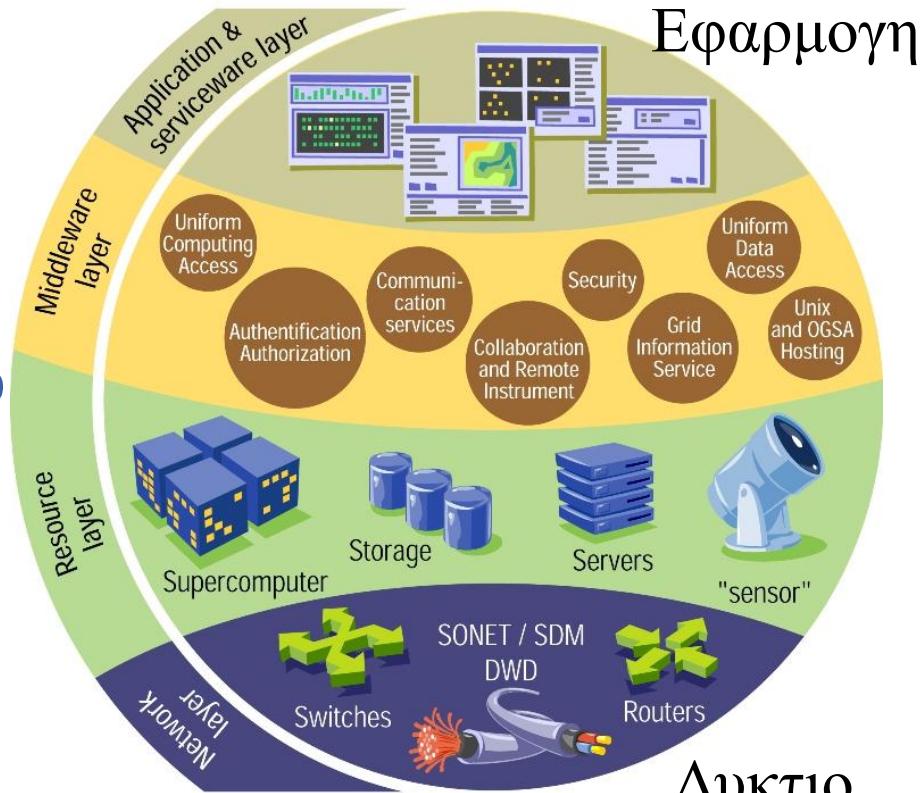
To World Wide Web παρέχει απλη πρόσβαση σε πληροφορίες που είναι αποθηκευμένες σε πολλά εκατομμύρια διαφορετικές γεωγραφικές τοποθεσίες

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



Πως δουλευει το Πλεγμα (Grid)

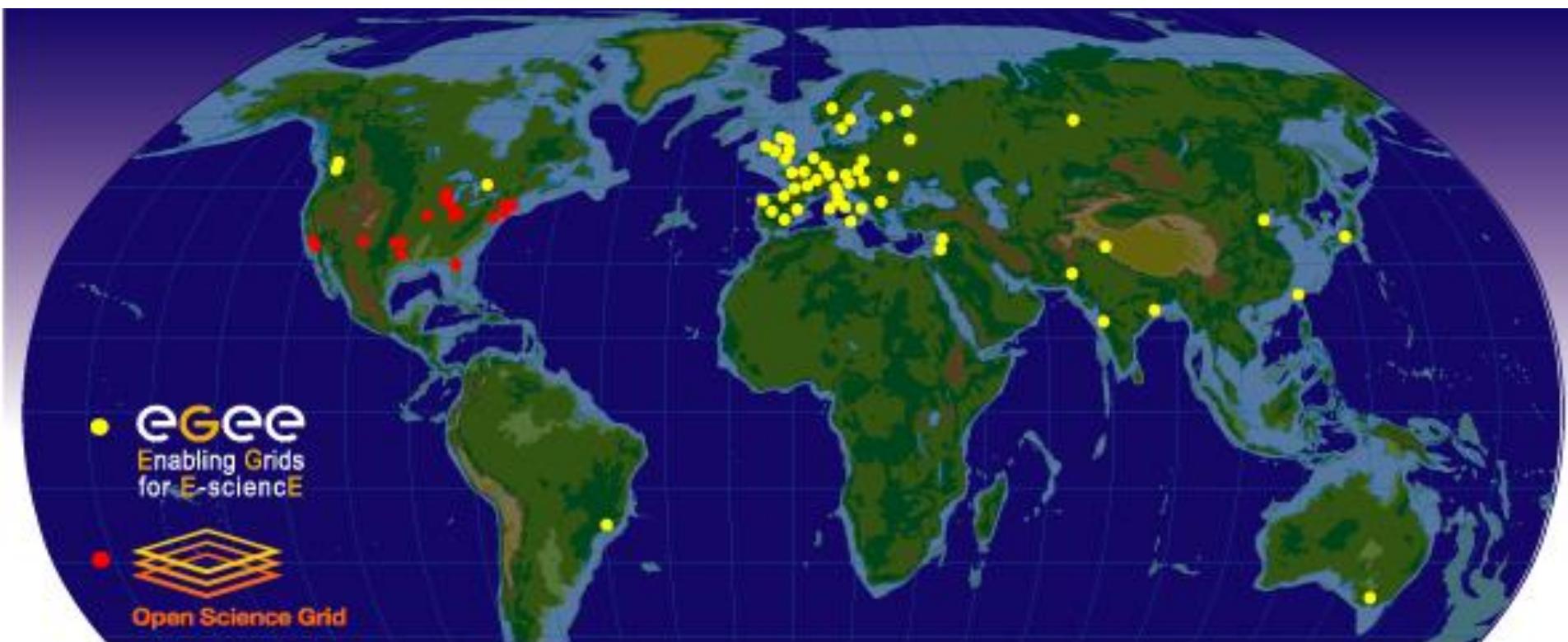
- Βασίζεται σε ειδικευμένο λογισμικό, το λεγομένο **μεσολογισμικό** (middleware).
- Η ιδέα του Middleware είναι να βρει αυτοματά τα **στοιχεία** που χρειαζεται ο/η ερευνητης, και το **υπολογιστικό δυναμικό** να τα επεξεργαστει.
- Middleware εξισοροπει το φορτίο σε διαφορούς πόρους και εγκαταστασεις. Χειρίζεται επισης **ασφαλεια**, λογιστική, επιτηρηση και πολλα αλλα.



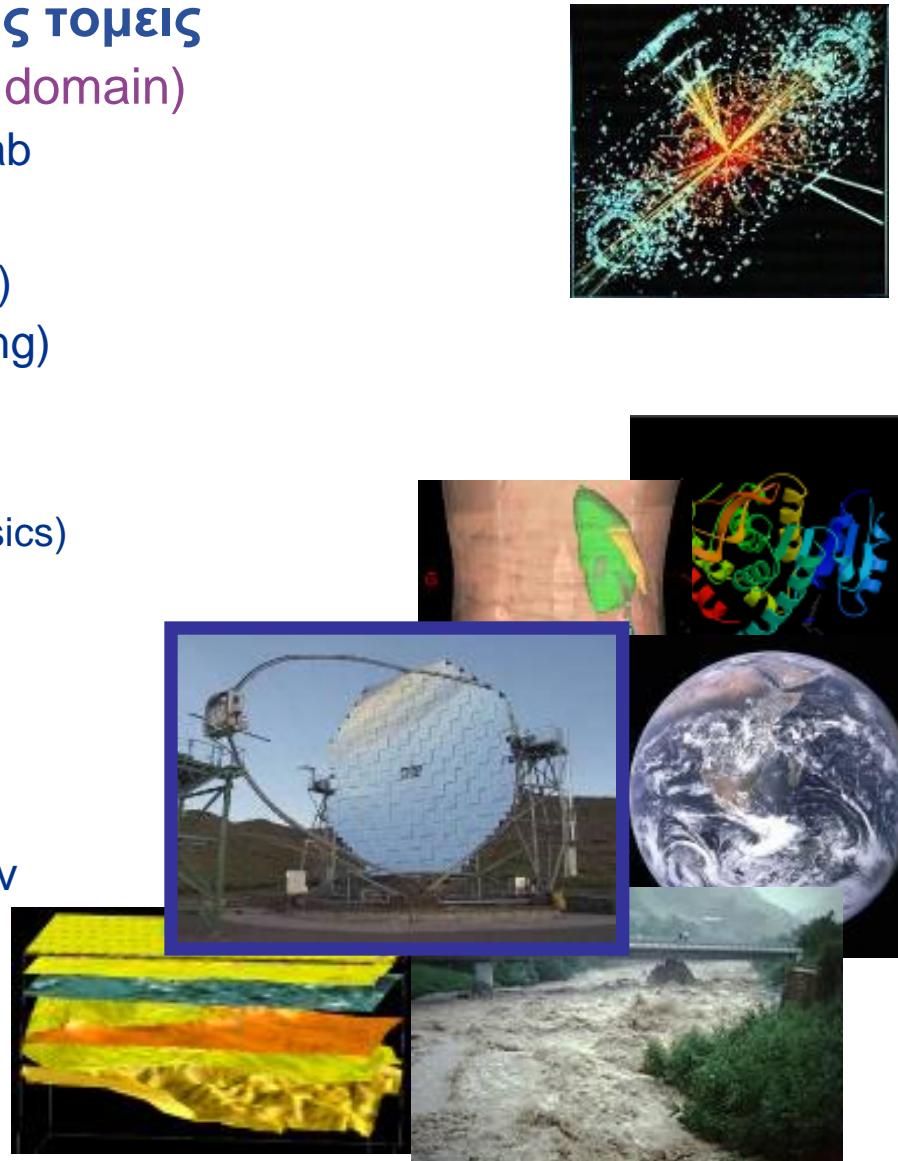


LCG depends on two major science grid infrastructures

- EGEE** - Enabling Grids for E-Science
- OSG** - US Open Science Grid



- Δεκαδες εφαρμογες σε διαφορους τομεις
 - Φυσικη Υψηλων Ενεργειων (Pilot domain)
 - 4 πειραματα LHC, DESY, Fermilab
 - Βιοϊατρική (Pilot domain)
 - Βιοπληροφορική (Bioinformatics)
 - Ιατρική απεικόνιση (Medical imaging)
 - Γεωεπιστημες
 - Γεω-επισκόπηση
 - Φυσικη Στερεας Γης (Solid Earth Physics)
 - Υδρολογία, Κλίμα
 - Υπολογιστική Χημεία
 - Τηξη (Fusion)
 - Αστρονομία
 - Κοσμικό υπόβαθρο μικροκυμάτων
 - ακτίνων-γ
 - Γεωφυσικη
 - Βιομηχανικές εφαρμογές

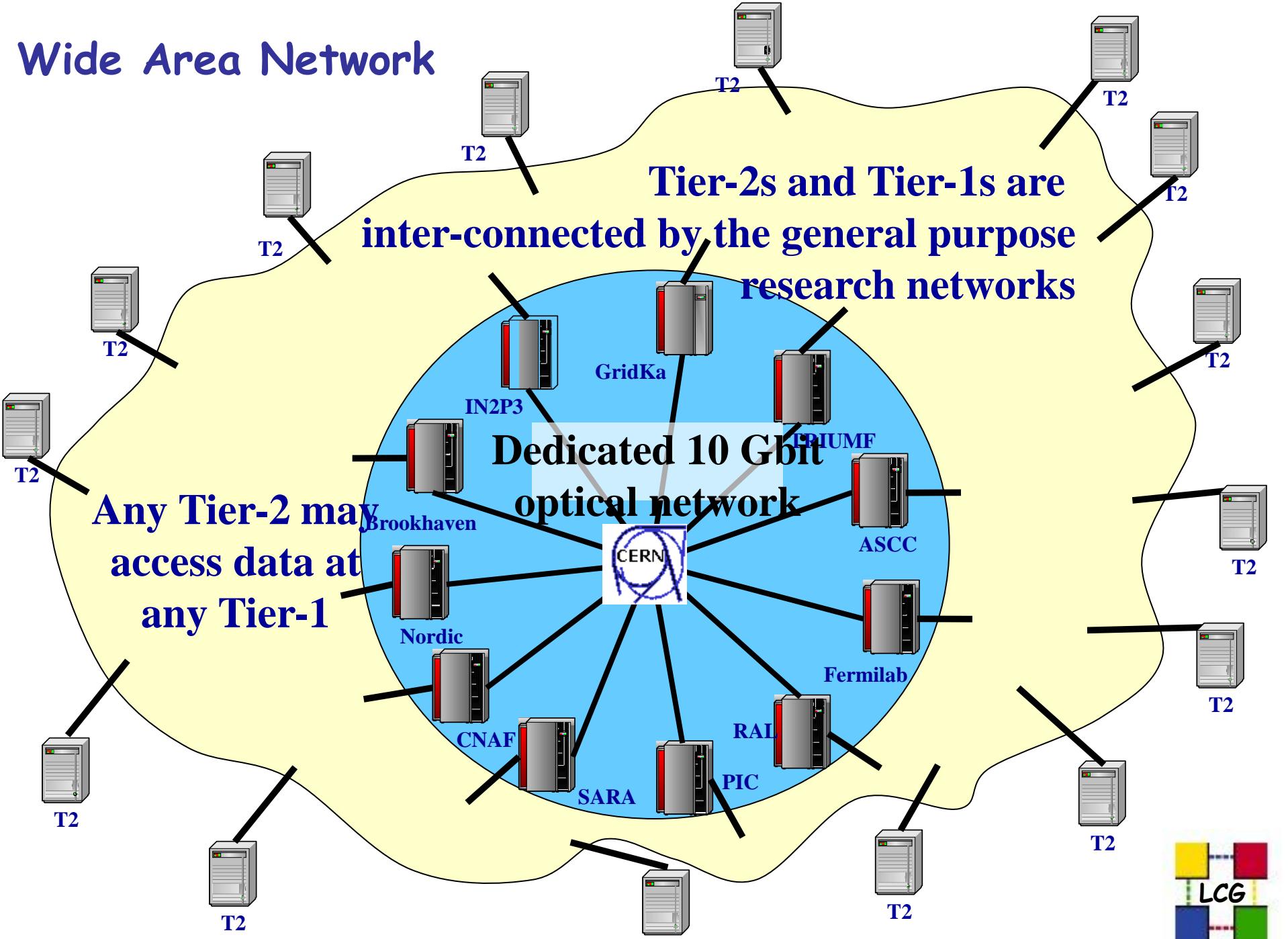


Wide Area Network

Any Tier-2 may
access data at
any Tier-1

Tier-2s and Tier-1s are inter-connected by the general purpose research networks

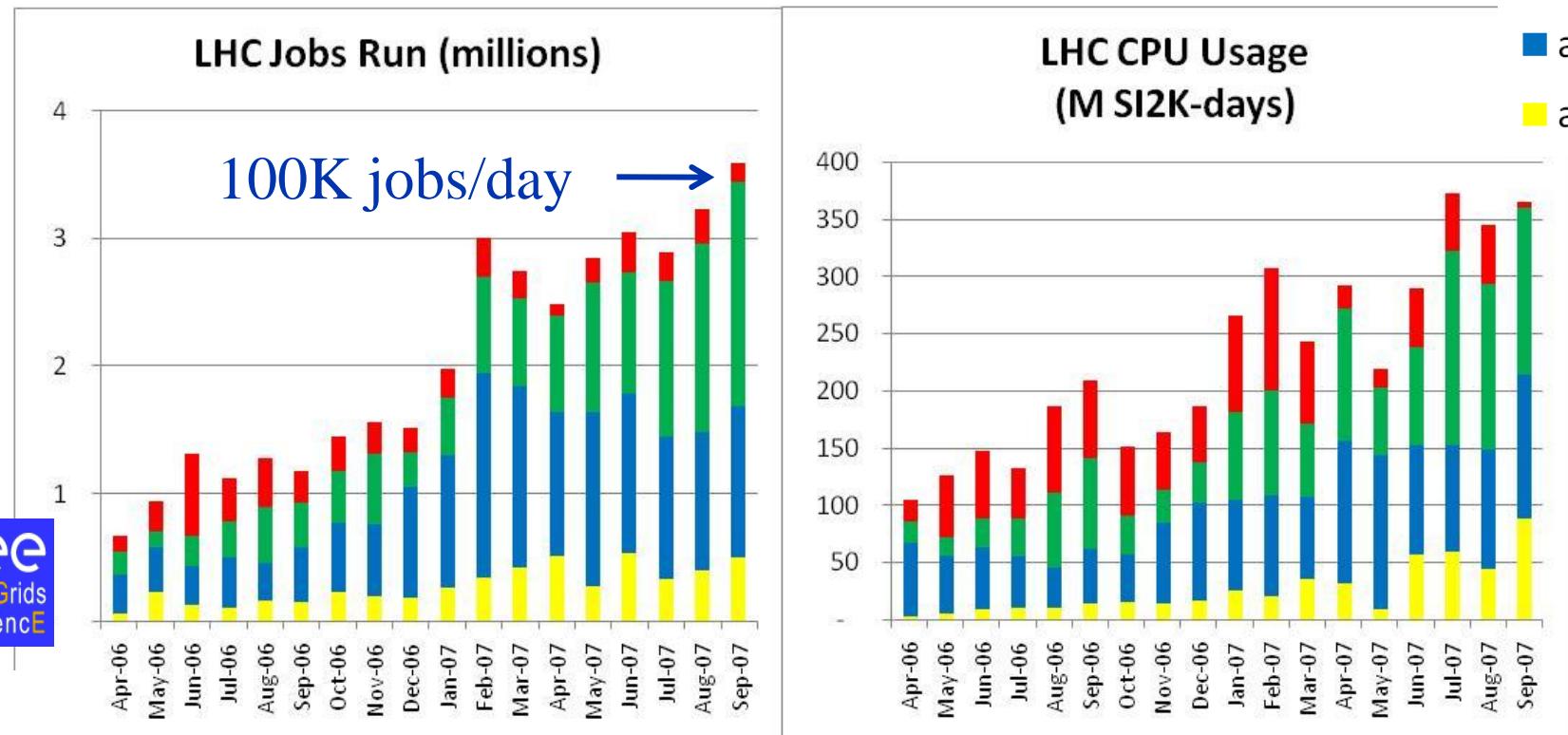
Dedicated 10 Gbit optical network





Grid Activity

- Συνεχης αύξηση στη χρήση του EGEE και OSG πλεγματων(grids)
- All sites reporting accounting data (CERN, Tier-1, -2, -3)
- Συνεχης αυξηση χρησης π.χ. 18 μηνες (ως Σεπ 2007) - επι 50 αριθμος των jobs και επι 3.5 ο χρονος χρησης



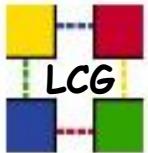
Οι φυσικοί προγραμματίζουν

- ⌘ Όλη η ανάλυση μετρησεων γίνεται με υπολογιστες
- ⌘ Οι φυσικοί ΥΕ χρησιμοποιουν ειδικευμένα προγραμματα
 - └─► Μερικοί γραφουν μεγαλες ρουτινες (routines/Fortran, methods/C++)
 - └─► Οι πιο πολλοι κανουν μικρες ρουτινες, για τις δικες τους αναγκες
- ⌘ Ολοι θα χροισημοποιουν τα 'εργαλεια' να δουν τις περιληψεις των μετρησεων



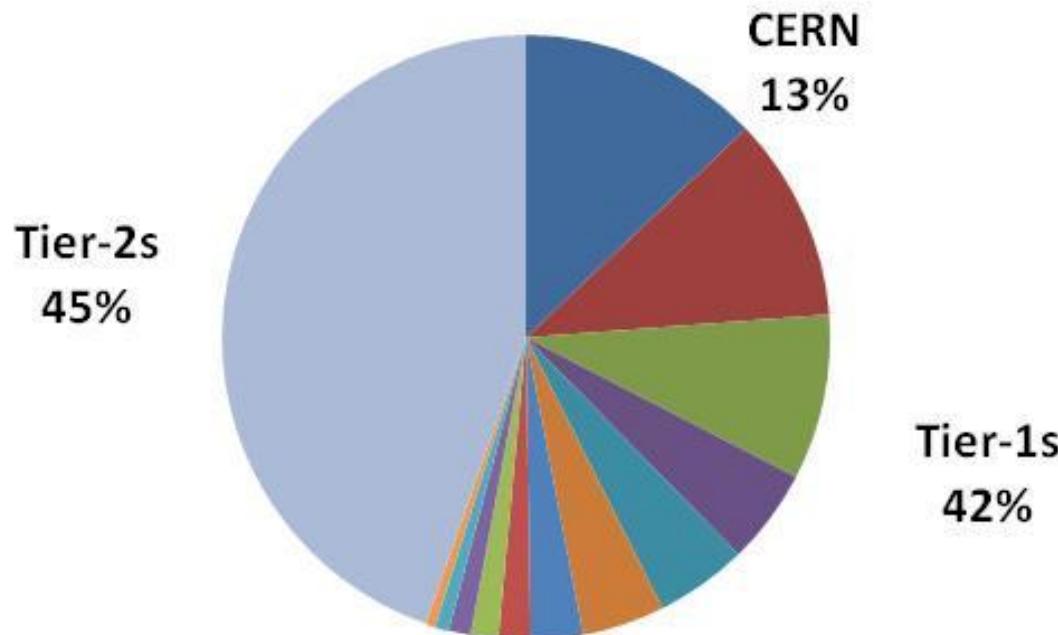
Backup





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

CPU Usage - September 2007



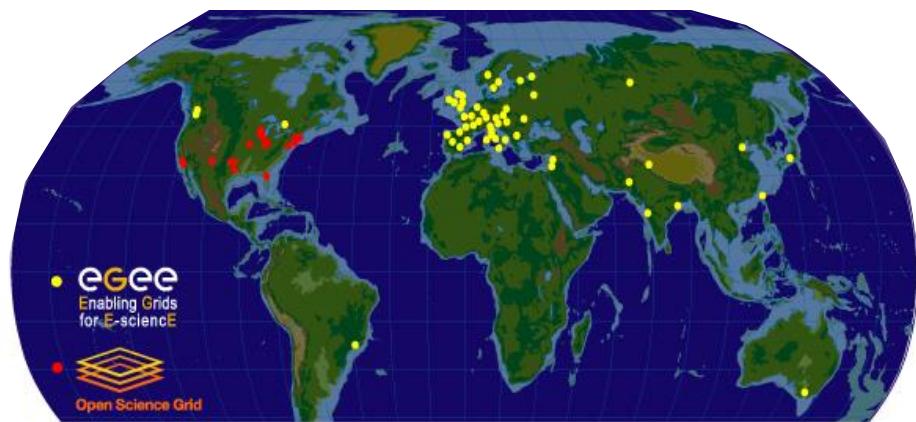
- > 80% of CPU Usage is external to CERN





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



Backup



More on simulation

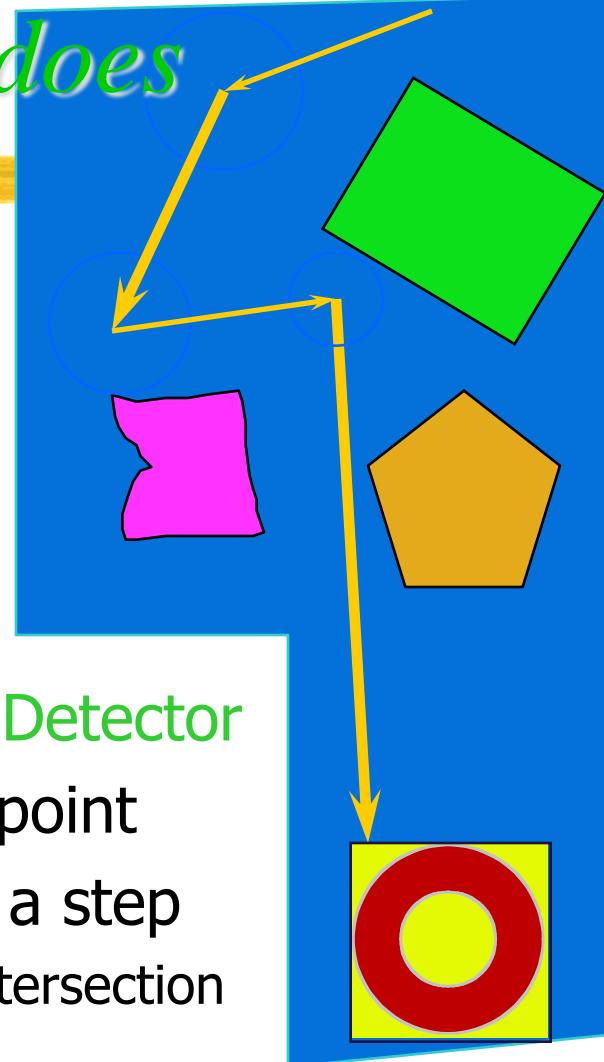
Geant4 geometry: what it does

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

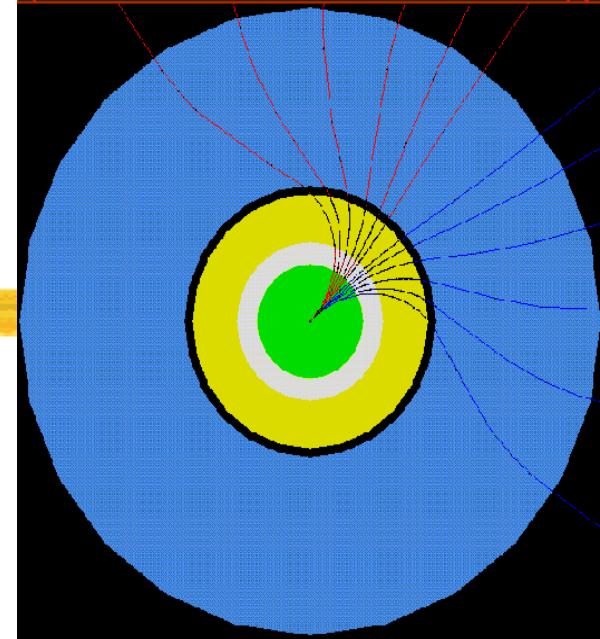
Navigates in Detector

- ⌘ 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:

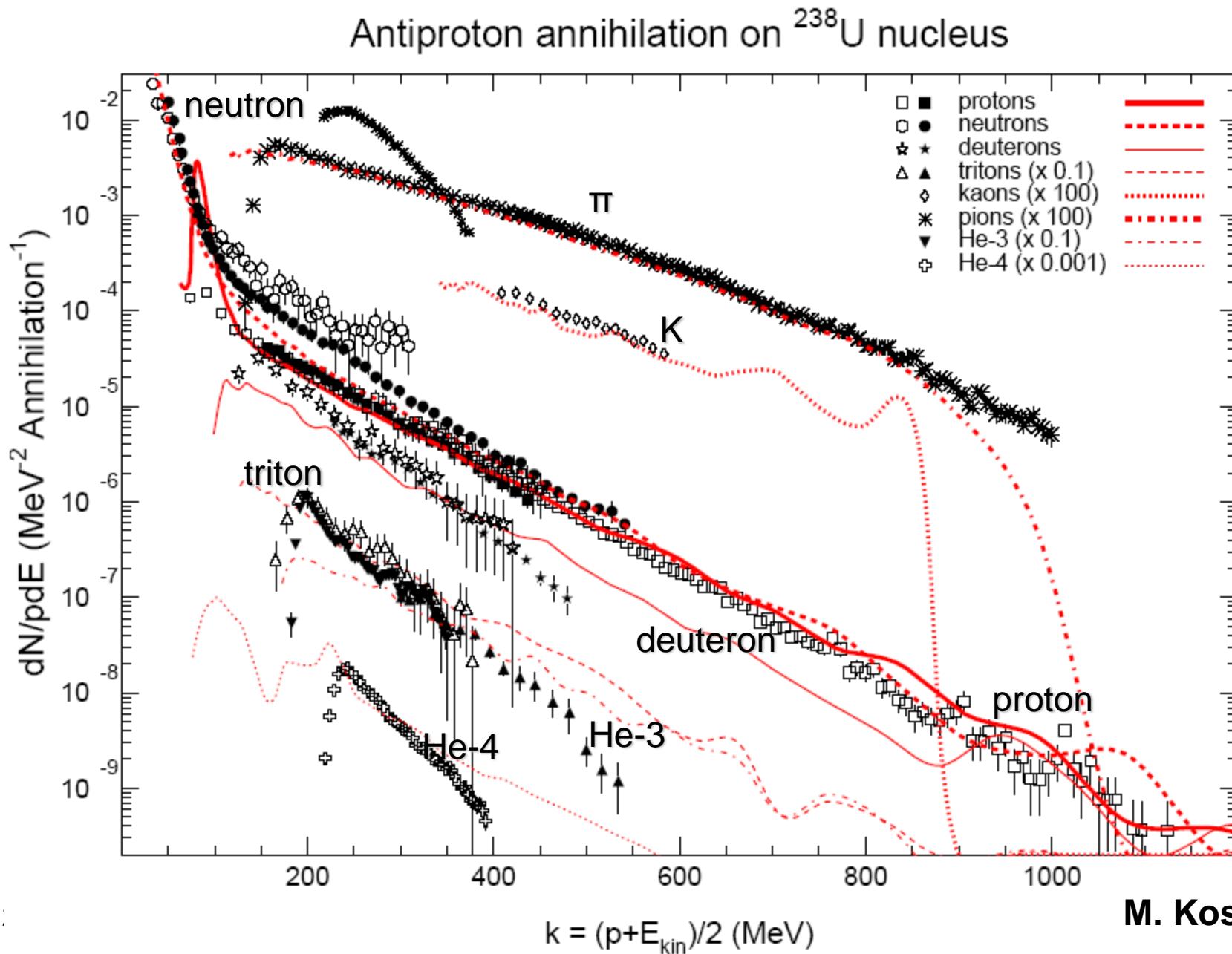
- ◻ Cerenkov, Rayleigh, Reflection, Refraction, Absorption, Scintillation

⌘ High energy muons and lepton-hadron interactions

⌘ Alternative implementation ("low energy")

- ◻ for applications that need to go below 1 KeV

Antiproton annihilation - CHIPS Model

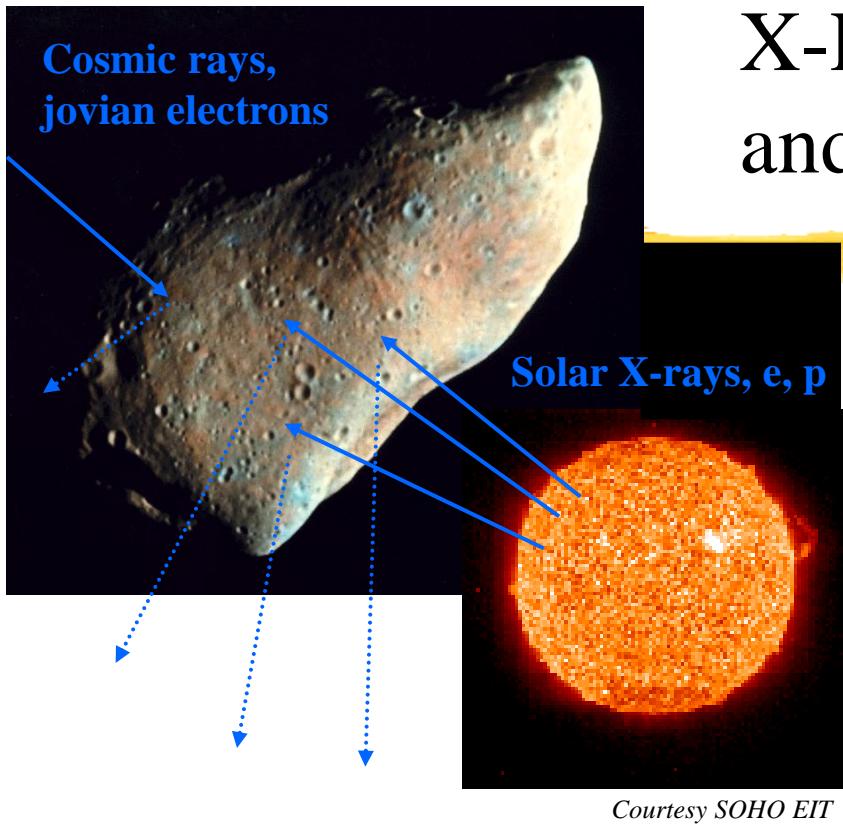


Simulation ‘packages’

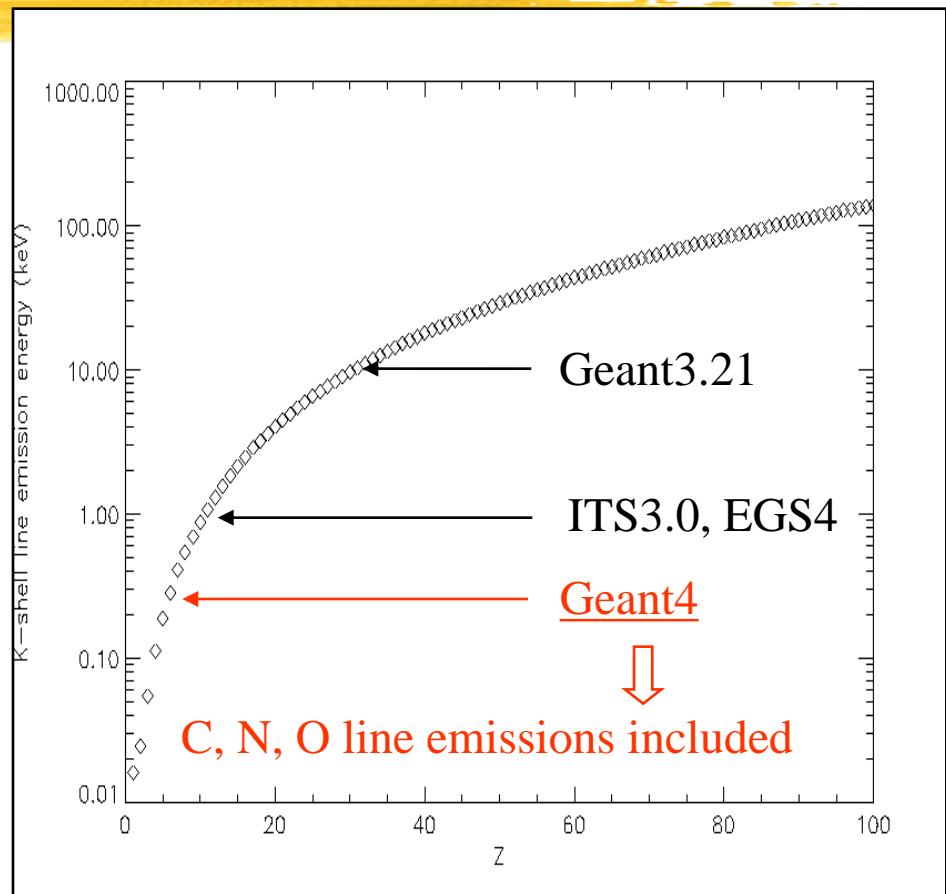


- ⌘ Provides the means to simulate
 - ↗ the **physical processes** and
 - ↗ **detector response** of an experiment.
- ⌘ As was realised by many in the past,
 - ↗ most of the parts needed can be **common** between experiments (eg physics, geometry blocks) .
- ⌘ So it makes eminent sense to create and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

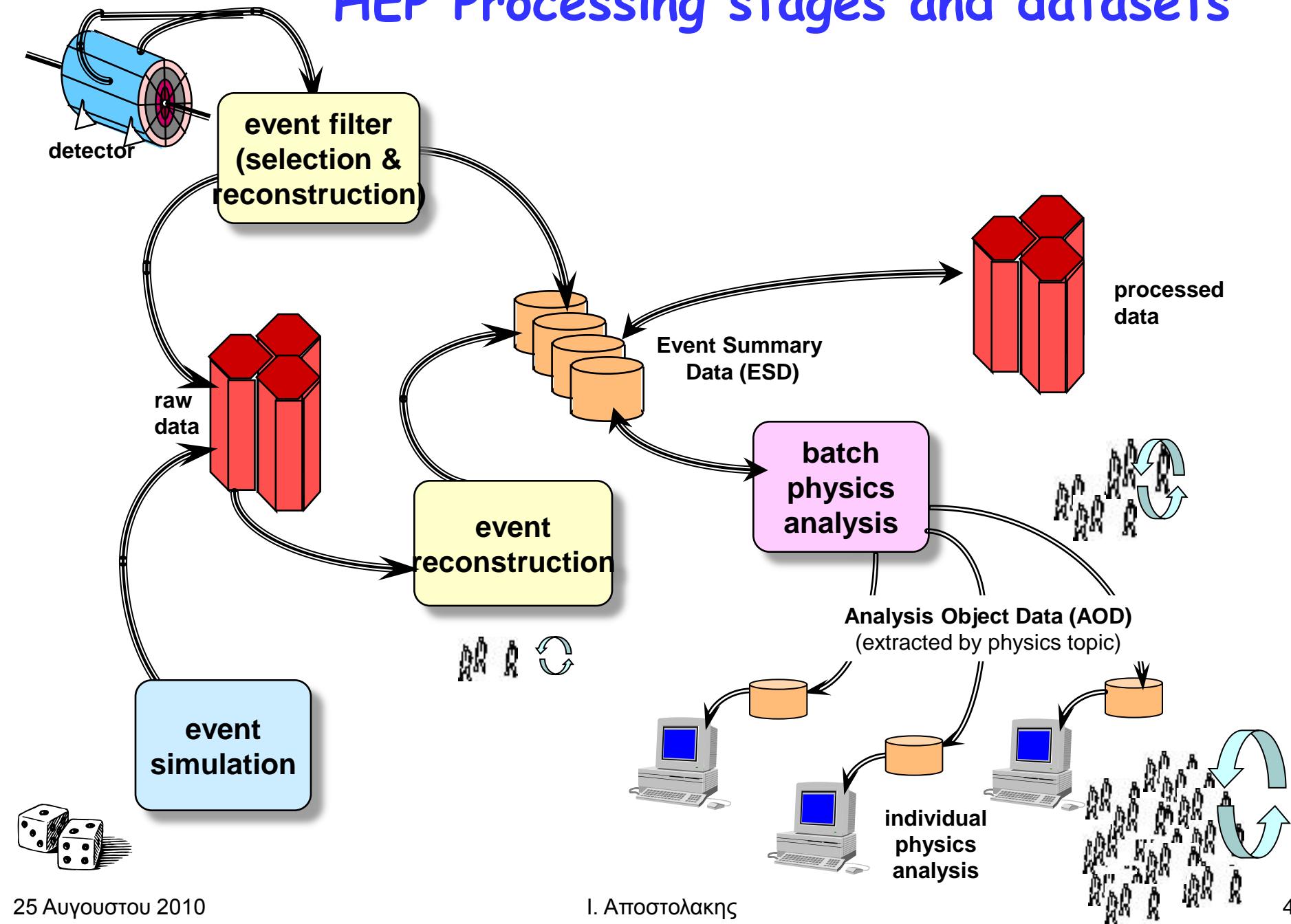
X-Ray Surveys of Asteroids and Moons



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

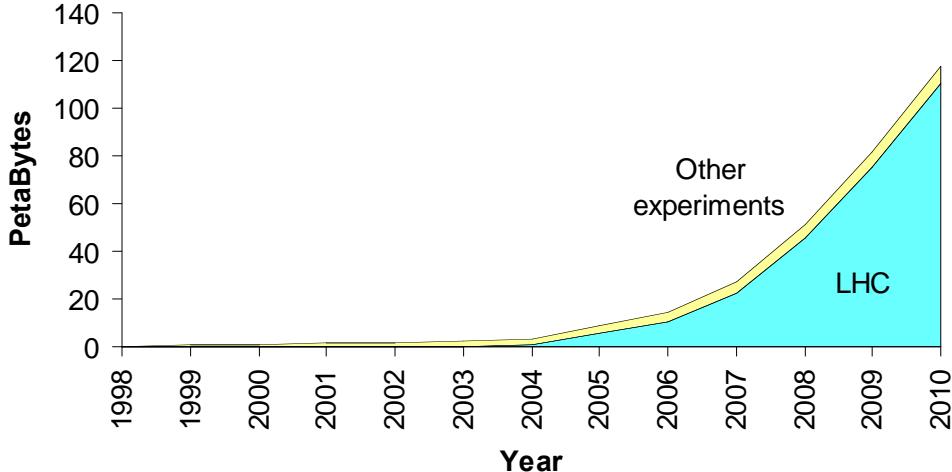


HEP Processing stages and datasets

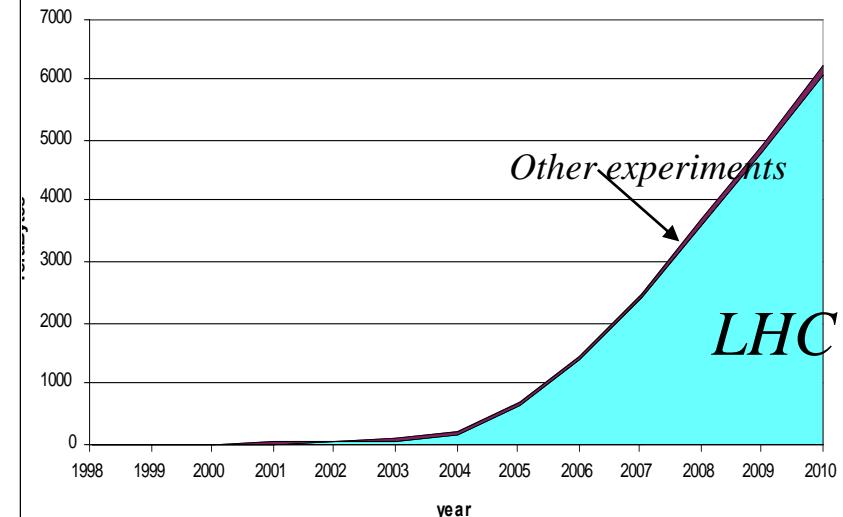


CERN Centre Capacity Requirements for all

Estimated Mass Storage at CERN



Estimated DISK Capacity at CERN



processing

K CT2000

3 700

8 200

10 100

25,000

34,000

disk

PE

5.0

6.7

tape media

PE

36

48

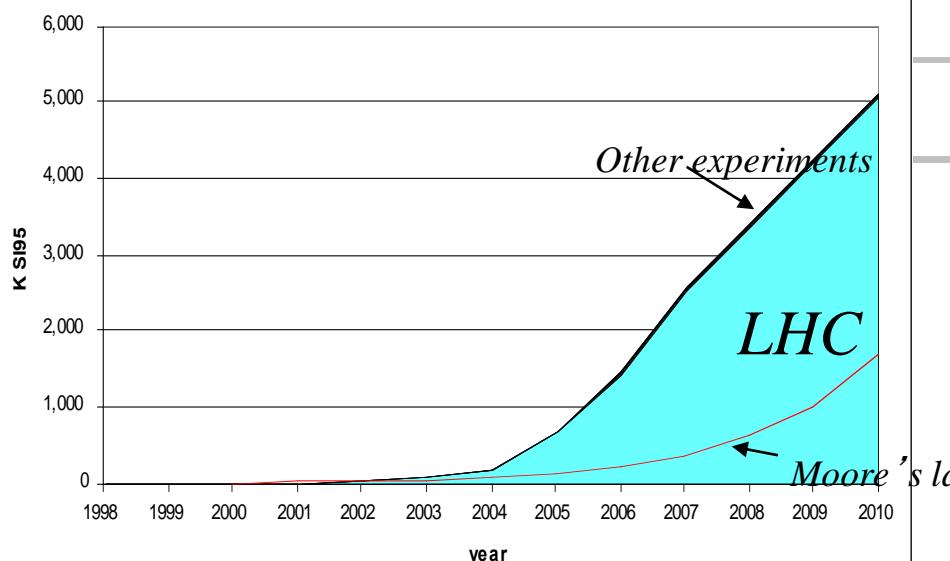
tape T/O

CT

39

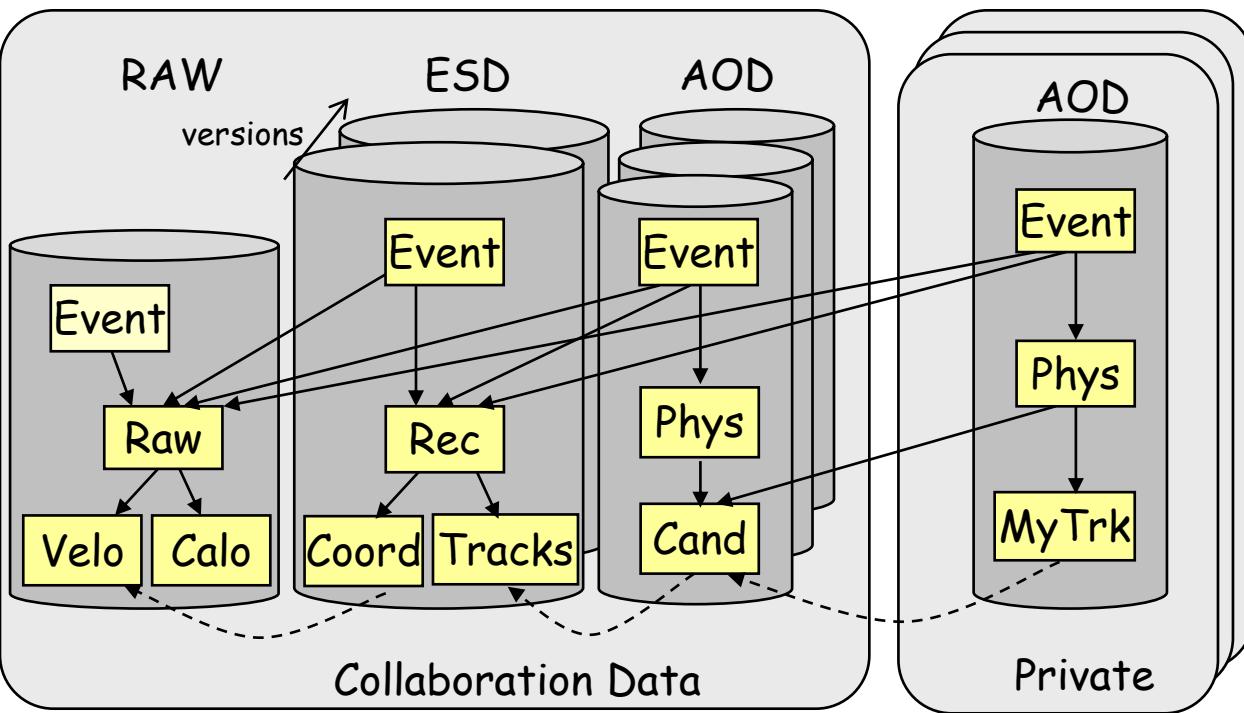
39

Estimated CPU Capacity at CERN



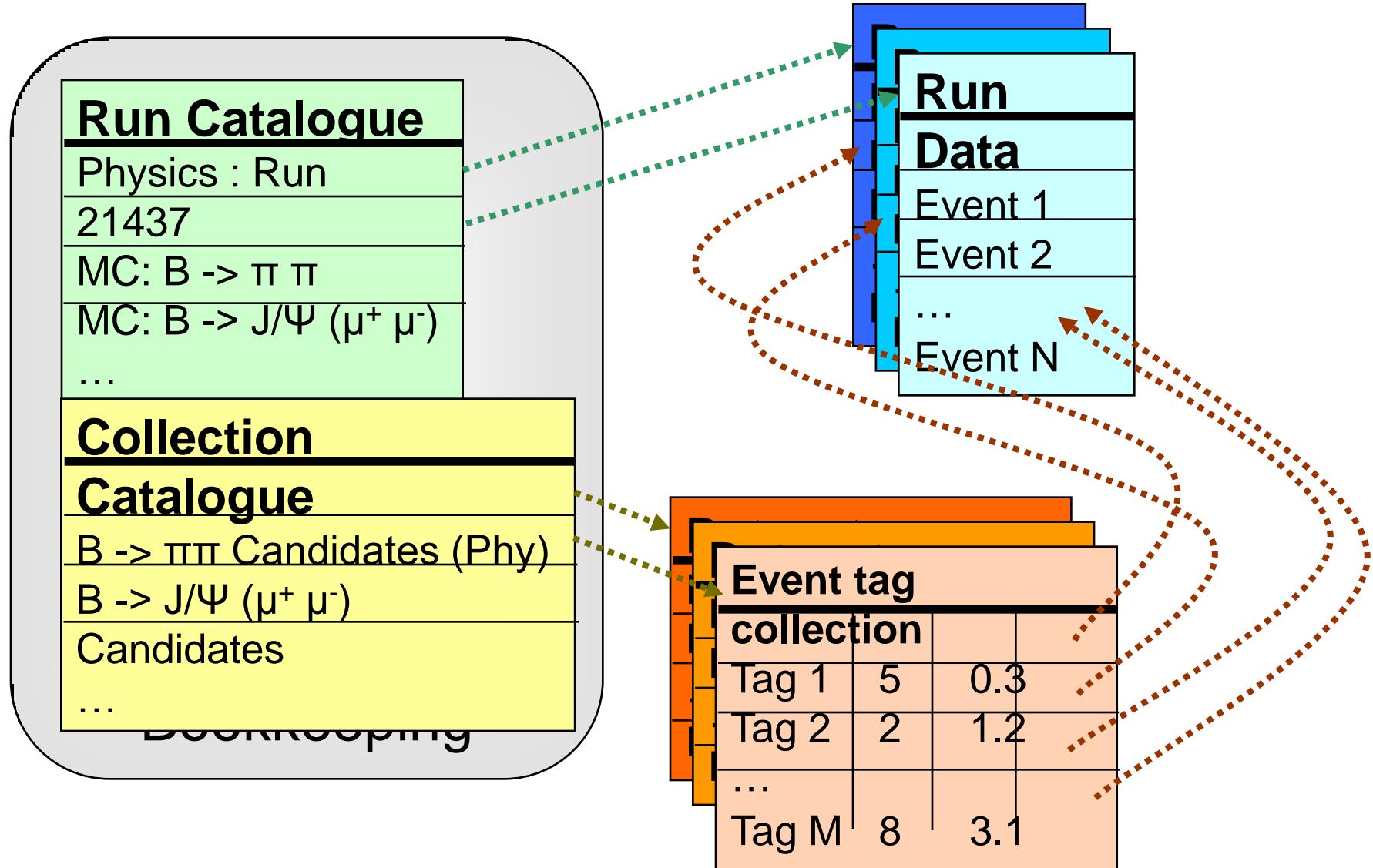
Moore's law

Event Data



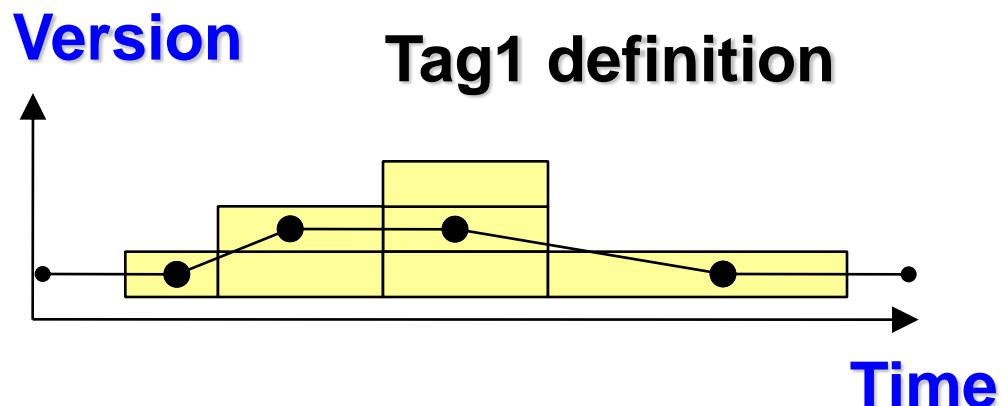
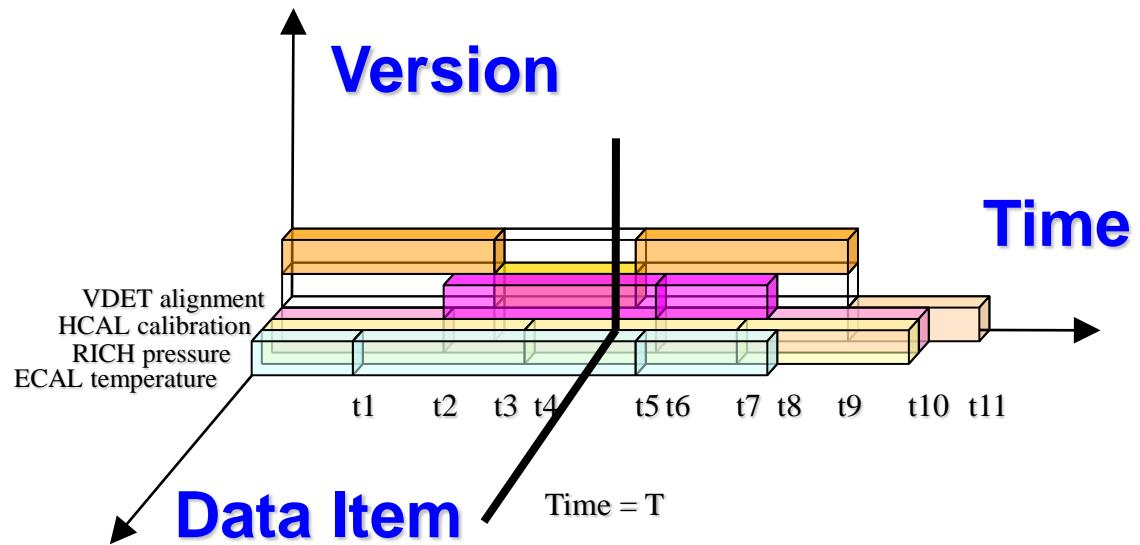
- ❑ Complex data models
 - ~500 structure types
- ❑ References to describe relationships between event objects
 - unidirectional
- ❑ Need to support transparent navigation
- ❑ Need ultimate resolution on selected events
 - need to run specialised algorithms
 - work interactively
- ❑ Not affordable if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

HEP Metadata - Event Collections

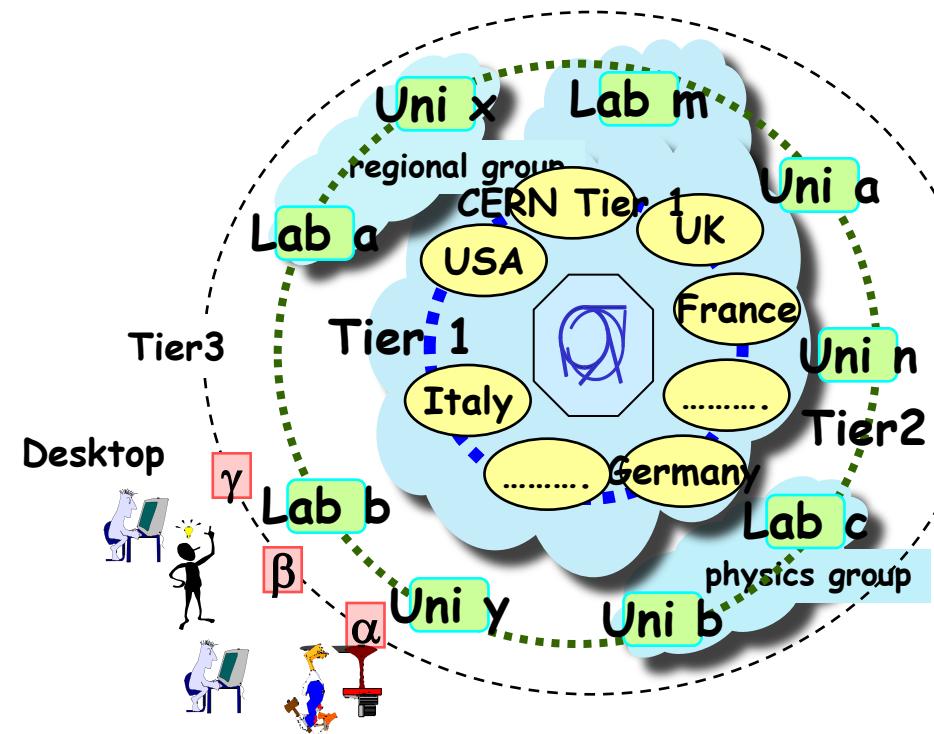
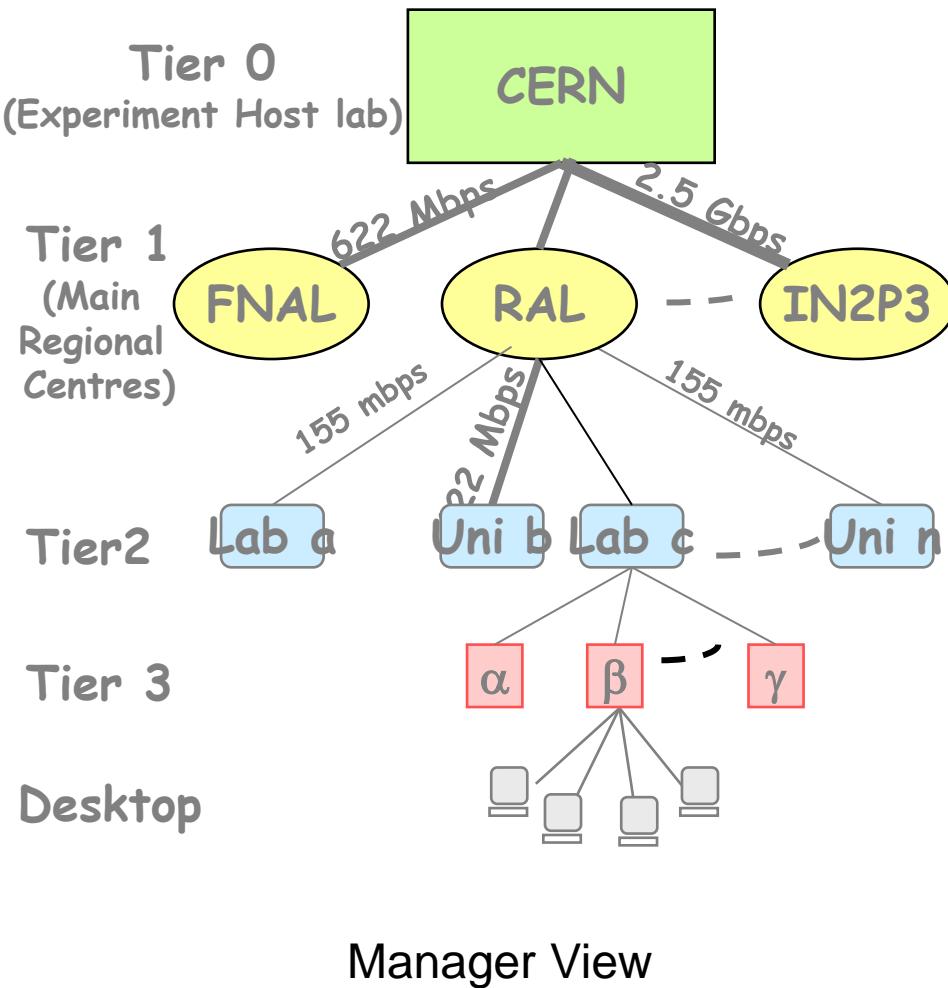


Detector Conditions Data

- 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



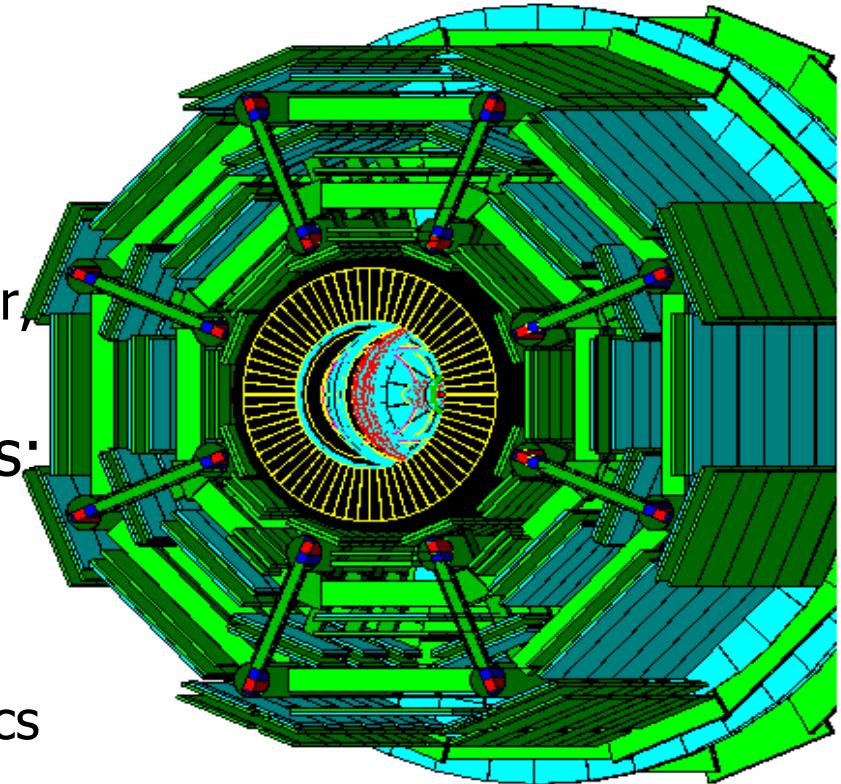
User View

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 ('discrete').

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)

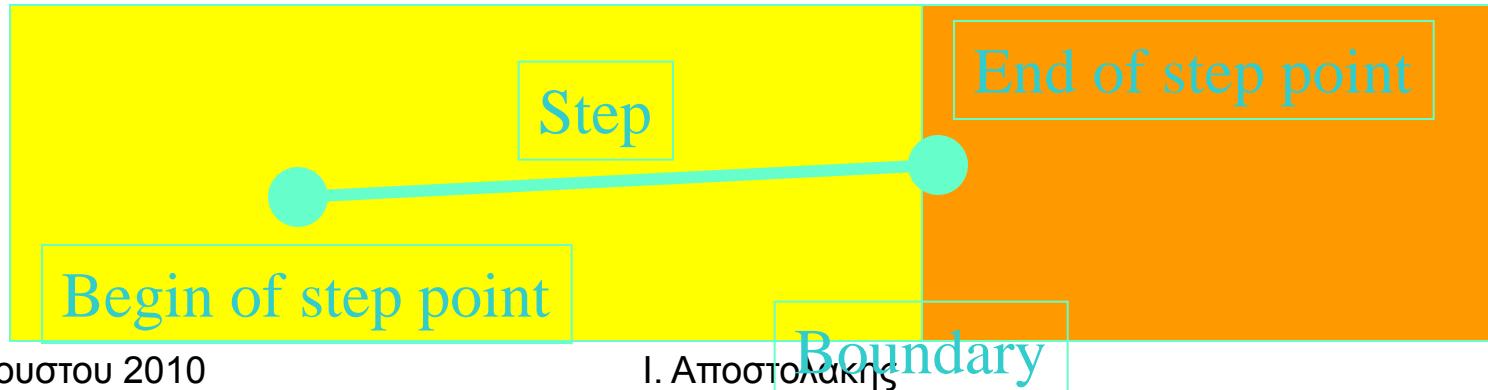
❖ During a step (continued)

❑ 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.

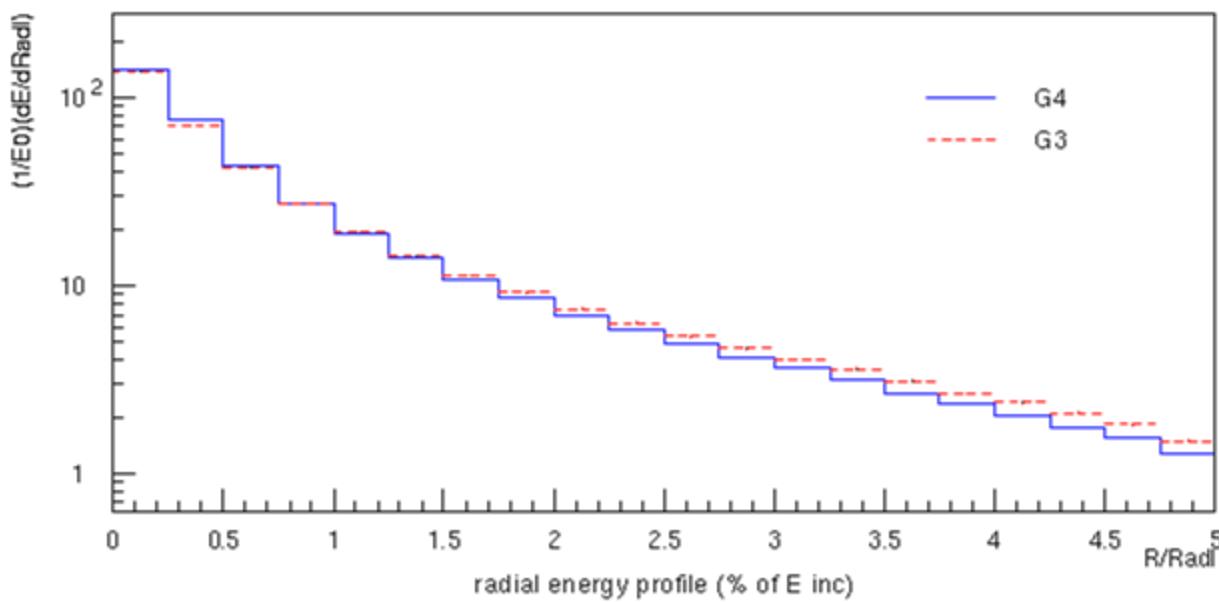
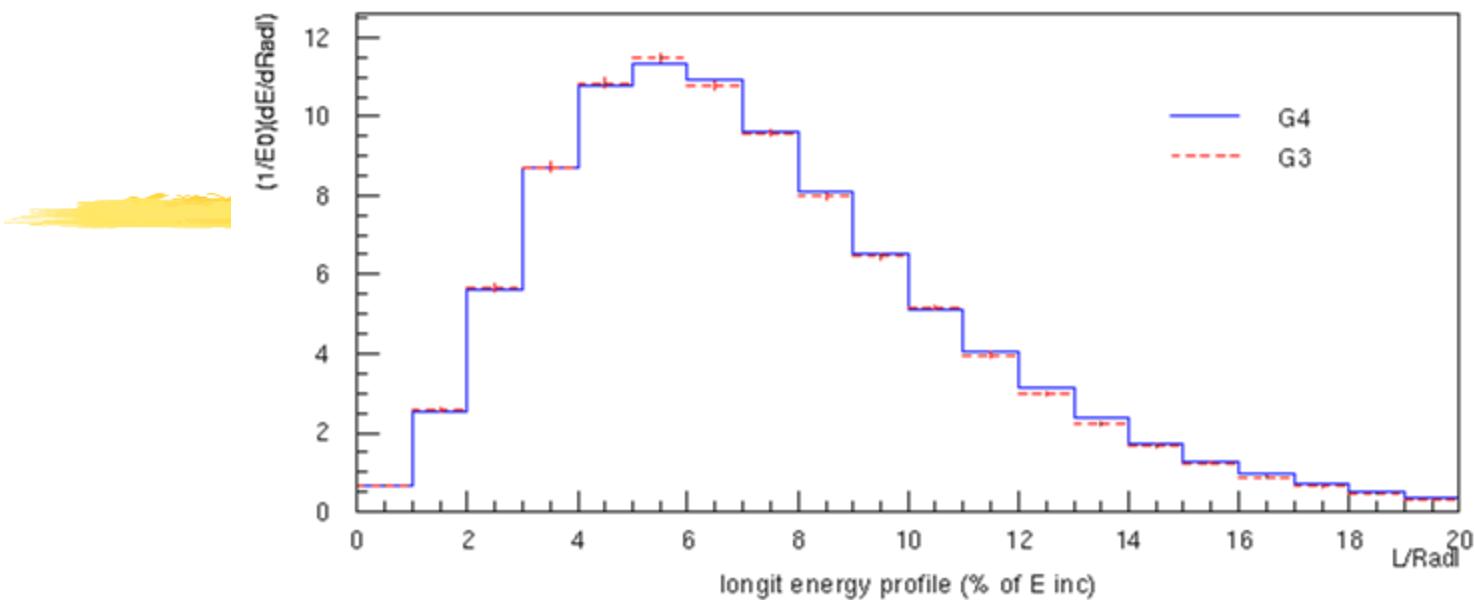


GEANT 4



- ⌘ 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



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.