



# Λογισμικά και Υπολογιστές στη Φυσική Υψηλής Ενέργειας (ΦΥΕ)

Γιαννης Αποστολακης  
**CERN**

v0.98.7  
2017.08.28

**[John.Apostolakis@cern.ch](mailto:John.Apostolakis@cern.ch)**

# Κυρια Θεματα



- z Η χρήση των Υπολογιστων
- z Τα Λογισμικα στη Φυσικη Υψηλης Ενεργειας (ΦΥΕ)
- z Το πλεγμα (GRID) – και μεγεθη
  - y Υπολογ. αναγκες, αλλες εφαρμογες

# Οι χρήσεις των Υπολογιστών και των Λογισμικών στη ΦΥΕ

- z Καταγραφή δεδομένων
  - y Data Acquisition (DAQ)
  - y Περιλαμβάνει την Επιλογή συγκρούσεων
- z Ανακατασκευή (reconstruction)
  - x Αμεσως (online) ή αργότερα (off-line)
- z Προσομοίωση (simulation)
- z Ανάλυση δεδομένων (data analysis)

# Πλano της ομιλιας

- z Κυρια (Ανοιχτα) Λογισμικα στη ΦΥΕ
  - y Ανακατασκευη (reconstruction)
    - x Αμεσως (online) ή αργοτερα (off-line)
  - y Προσομοιωση (simulation)
    - x Τι είναι και σε τι χρησιμευει
  - y Αναλυση δεδομενων (data analysis)
    - x Νεα σωματιδια, καλυτερη γνωση/ιδιοτητες αλλων
- z Το πλεγμα (GRID) – και μεγεθη
  - y Υπολογ. αναγκες, και αλλες εφαρμογες

# **Ανακατασκευή**



**Μια γρήγορη εισαγωγή**

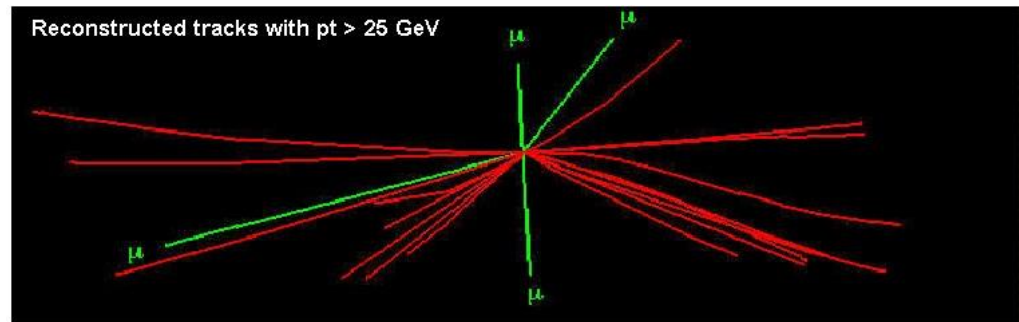
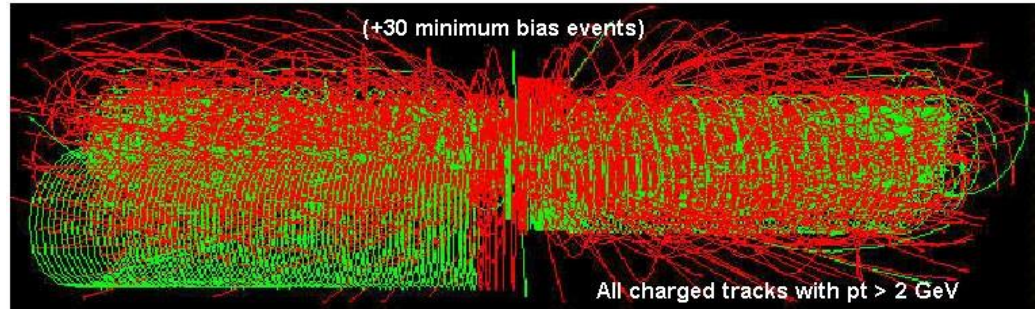
# Η δυσκολεια της Ανασυγκρότησης

Starting from  
this event

Μ'αυτη την  
αρχικη εικονα  
(αριστερο ημισυ)

Looking for  
this “signature”

Ψαχνουμε αυτη  
την περιληπτικη  
εικονα



→ **Selectivity: 1 in  $10^{13}$**

(Like looking for a needle in 20 million haystacks)

# Οι σημερινοί ανιχνευτές

## z Πολλα τμήματα

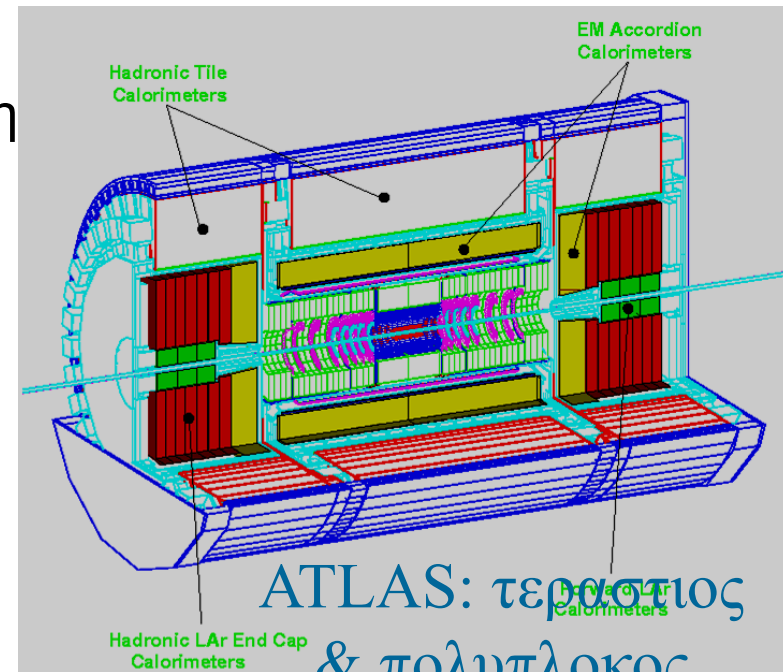
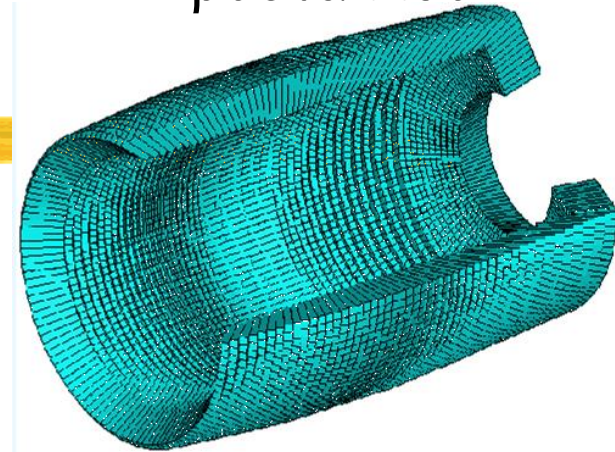
### y Διαφορετικές ανάγκες

- x Μετρηση θέσης (τρακερ - trackers)
- x Μετρηση ενέργειας (καλοριμετρα ή θερμιδομετρα)

## z Λογω της πολυ-πλοκοτητας

### y Όλες οι μελετες χρειαζονται υπολογιστικα εργαλεια

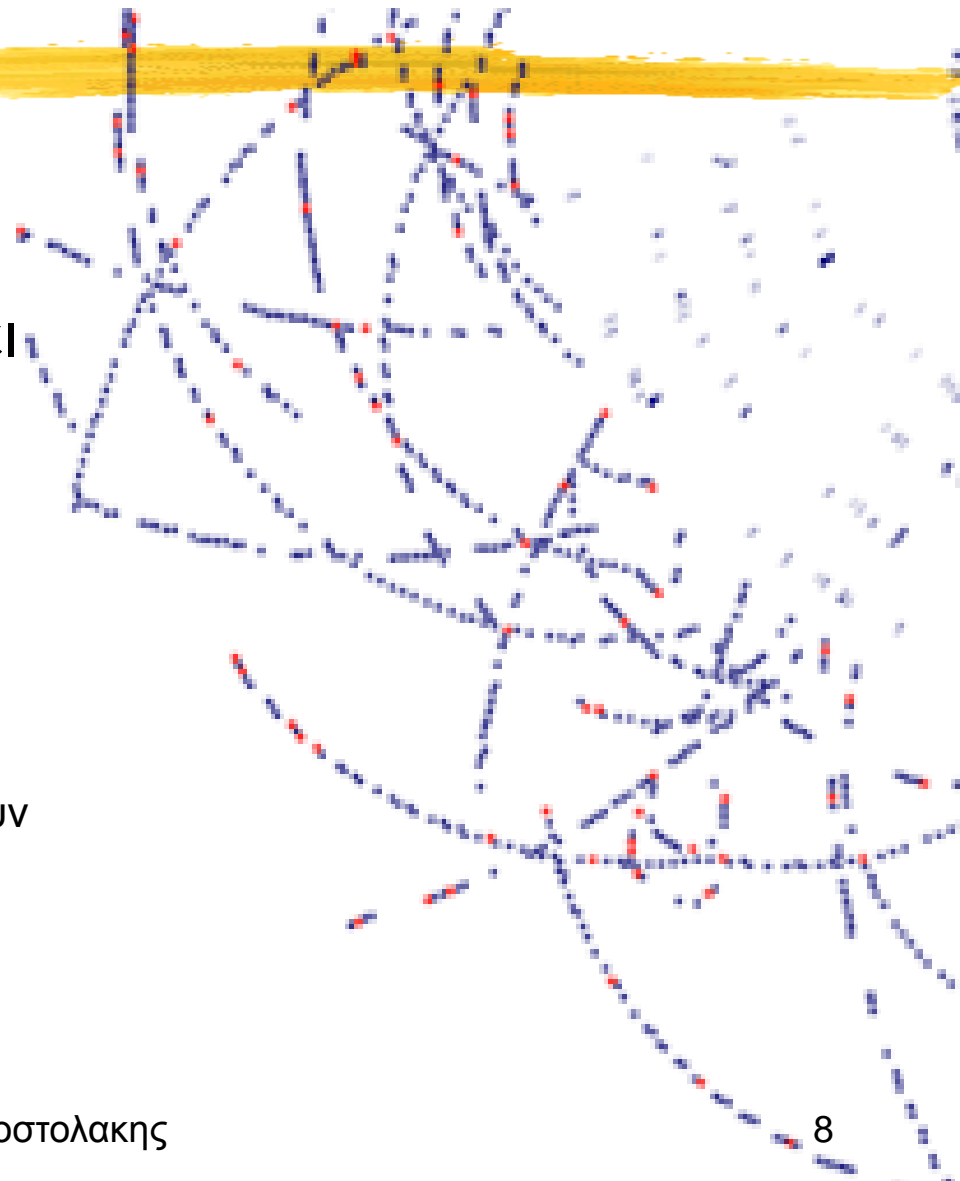
Καλοριμετρο  
κρυσταλλου



ATLAS: τεραστιος  
& πολυπλοκος

# Τι είναι η ανακατασκευή?

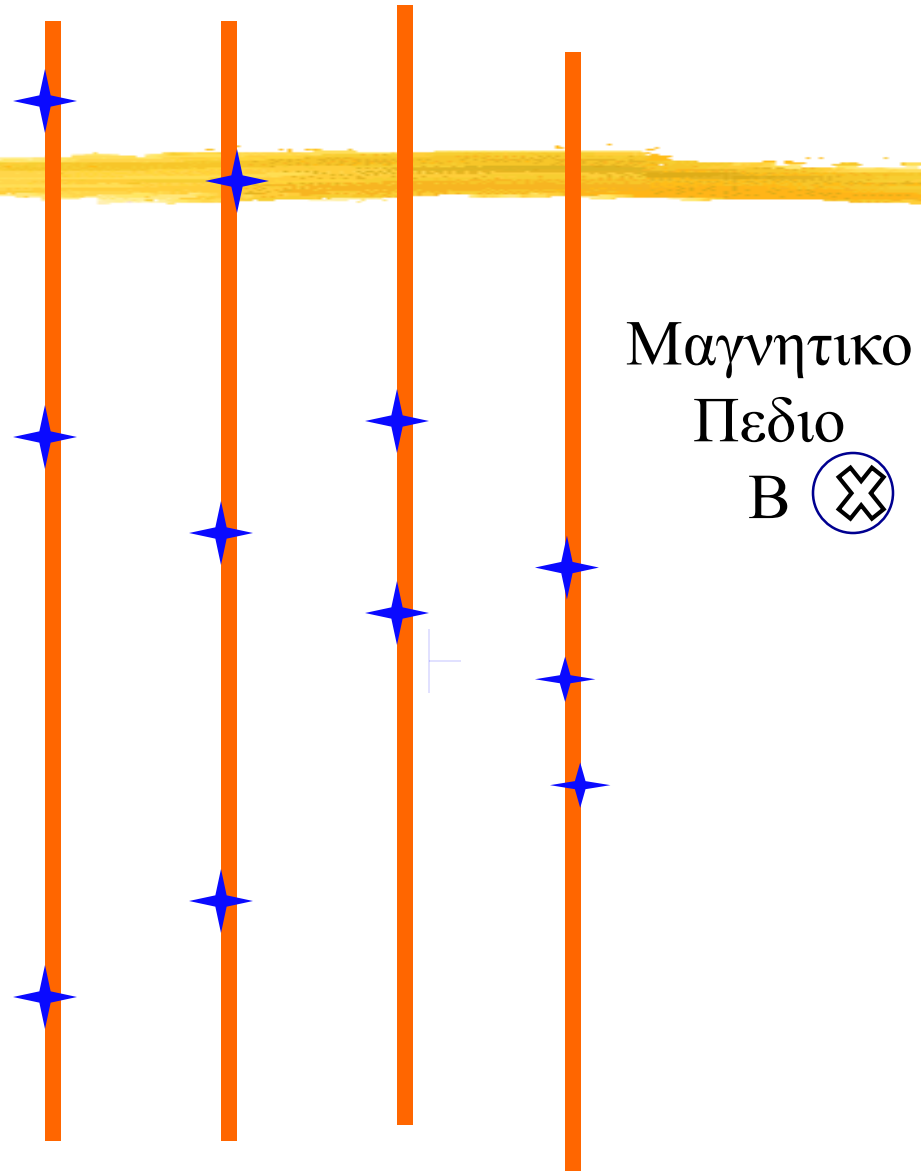
- z Οι μετρήσεις είναι σαν ένας γρίφος
  - y Τι τροχιές τις προκάλεσαν?
- z Καθε μέτρηση θέσης βοηθάει
  - y Υπάρχουν όμως 100-άδες ως χιλιάδες μετρήσεις
- z Η ανακατασκευή πρέπει να βρει τη λύση!
  - y Ξεροντας καλά το μαγνητικό πεδίο
    - x Βρισκουμε ποιες μετρήσεις ανοικουν σε ποιες τροχιές





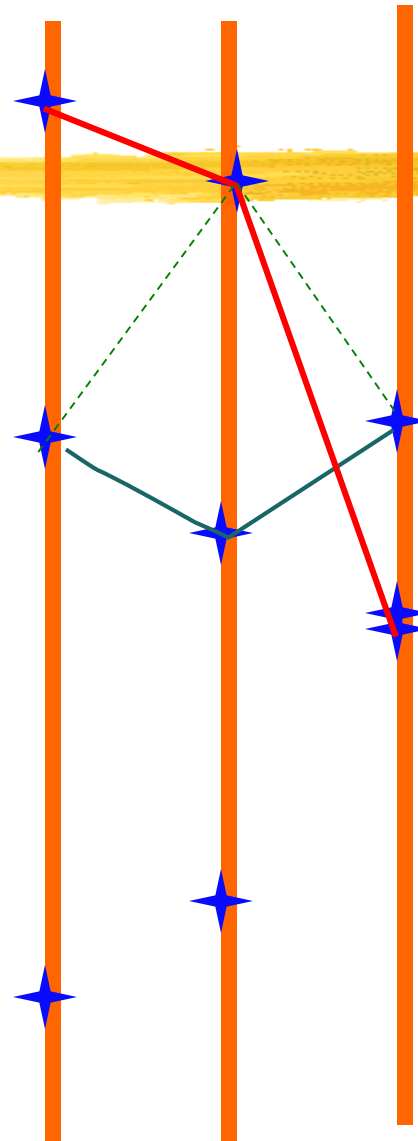
# Ανακατασκευή στην πράξη

- 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
- 



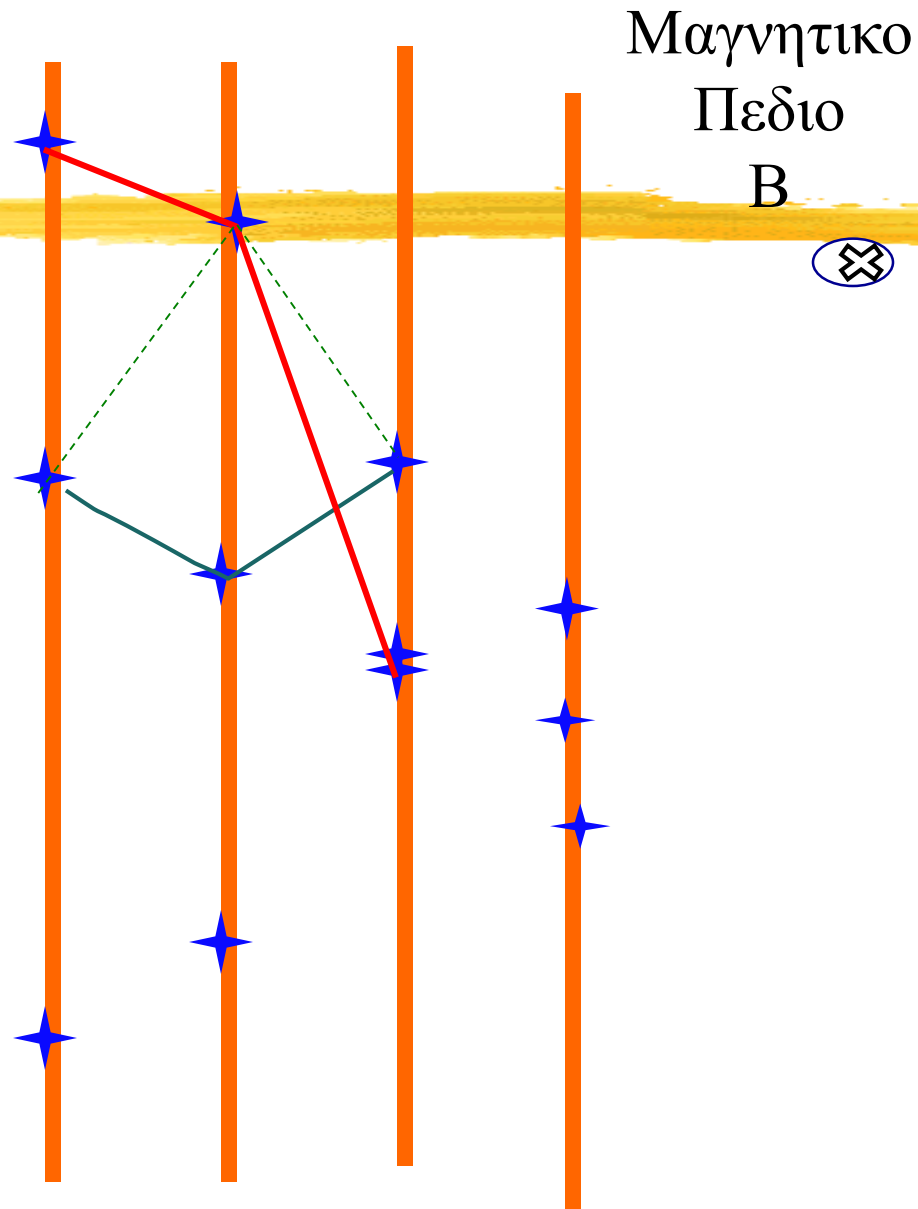
Μαγνητικό  
Πεδίο  
 $B$



Αλγοριθμ-ος/-οι  
Φίλτρο Καλμαν  
(Kalman filter)

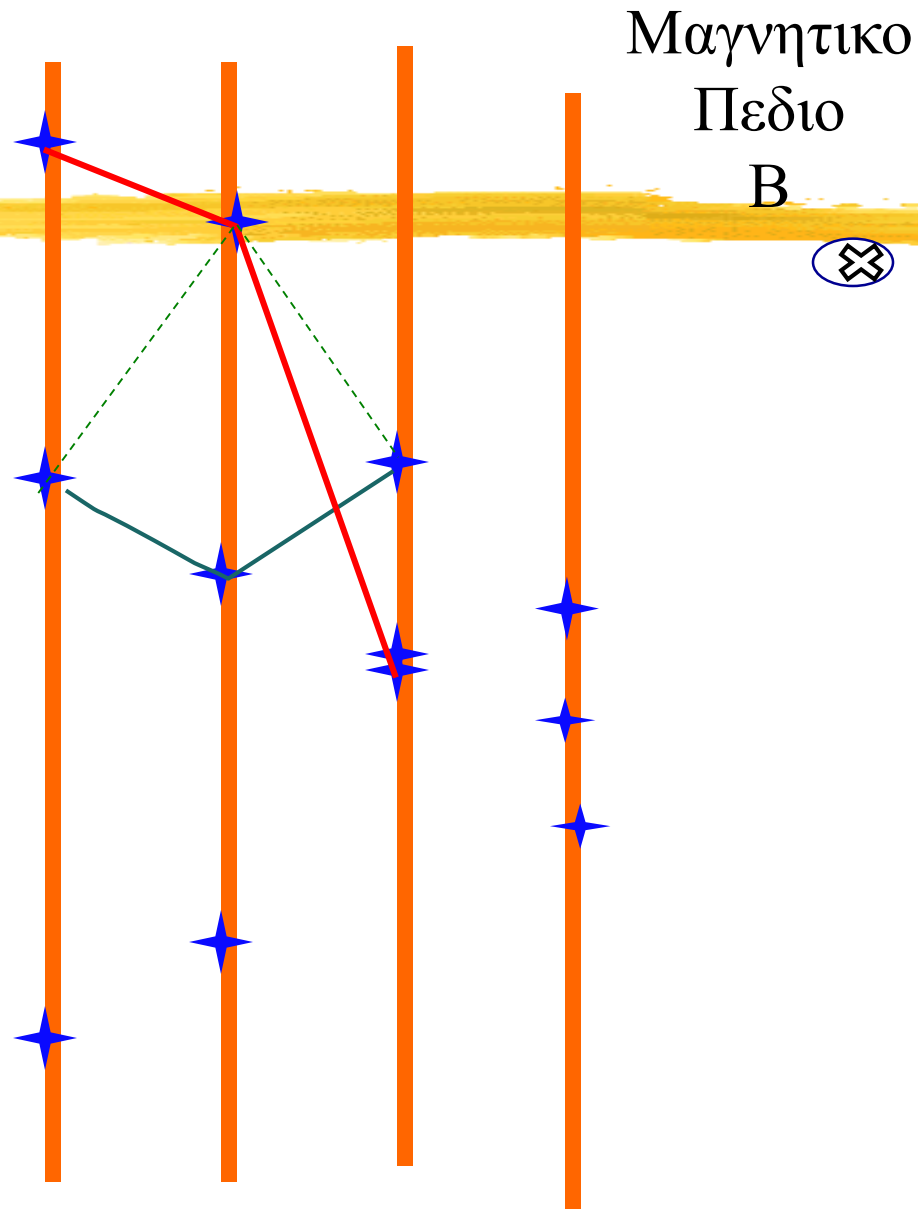
# Ανακατασκευή στην πράξη

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



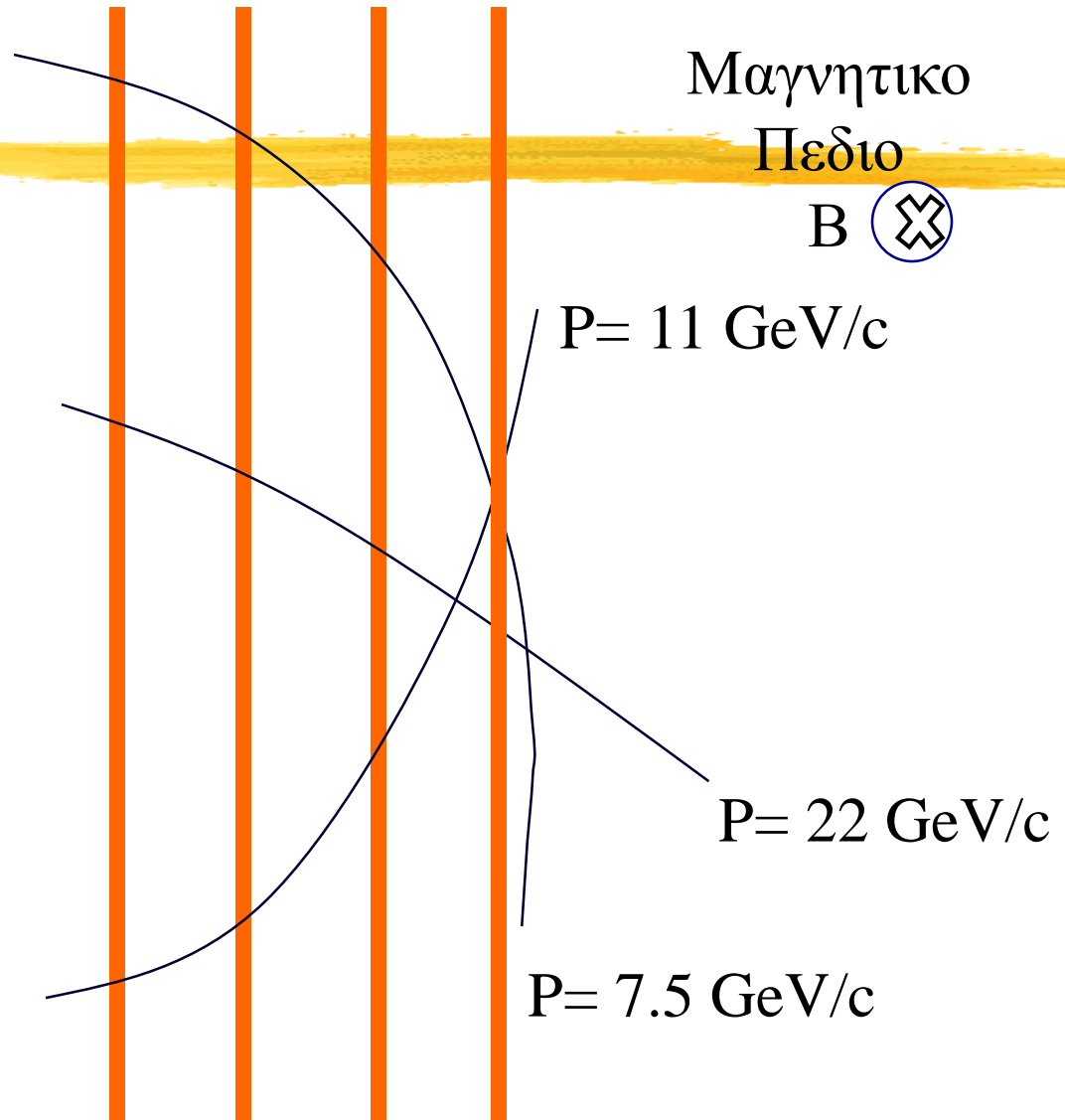
# Ανακατασκευή στην πράξη

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



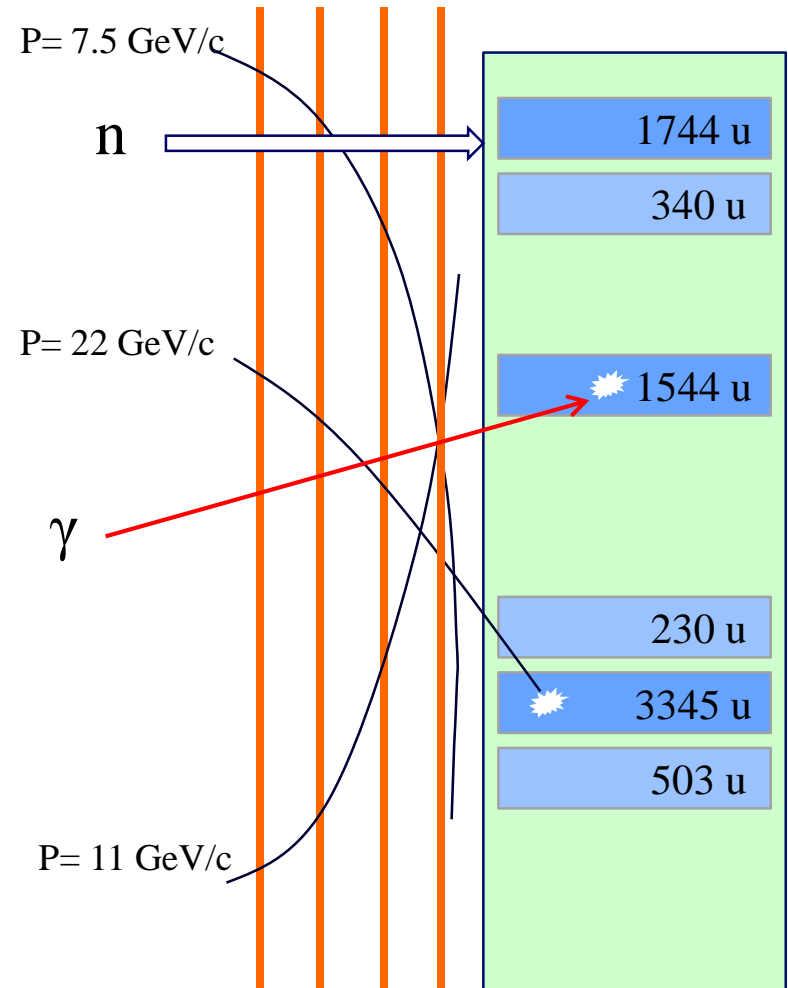
# Ανακατασκευή: πρώτο αποτέλεσμα

- 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

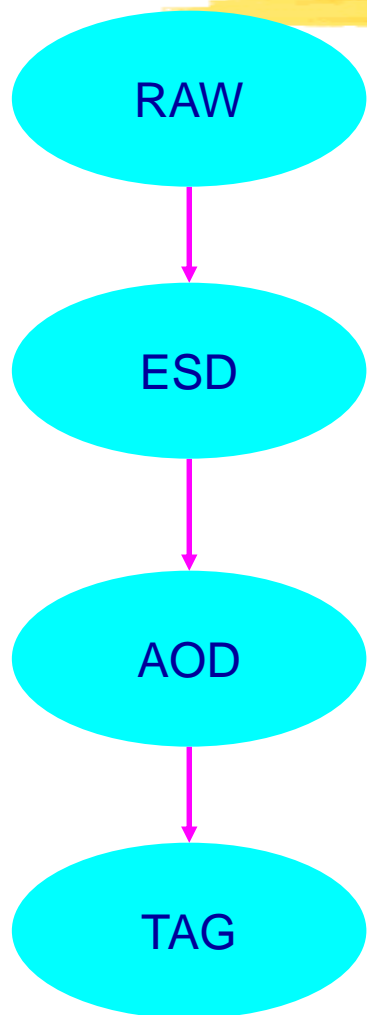


# Από μερική προς 'ολική' ανακατασκευή

- z Μετά από αυτά τα πρώτα ίχνη
- z Προσθέτουμε ανιχνευτές
  - y Ευαίσθητους, σε περισσότερα σωματίδια, άλλες αλληλεπιδράσεις
- z Βγάζουμε πληροφορίες για τα άλλα σωματίδια ( $\gamma$ ,  $\pi$ ,  $K$ ,  $\rho$ ,  $n$ , ... )
- z Συγκρύνουμε «ολόκληρη» την εικόνα της κρούσης



# Χρηση της ανακατασκευης



Οτι κατεγραψαν οι ανιχνευτες (DAQ)

Ο αρχικος ογκος πληροφοριων

RAW

Ανασυγκροτηθησα πληροφορια

Προτύπα αναγνώρισης:  
- Συμπλέγματα (clusters)  
- Υποψήφιες τροχιες

ESD

Συμπερασματα για την αναλυση

Πληροφορίες «για τη Φυσική»:  
Εγκάρσια ορμή,  
Συσχέτιση σωματιδίων, πιδακες (jets),  
(καλύτερη) αναγνώριση σωματιδίων

AOD

Πληροφορίες ταξινόμησης

Πληροφορίες σχετικές με τη γρήγορη επιλογή των κρουσεων

TAG

# Προσομοίωση και Ανιχνευτές



Τι είναι η προσομοίωση

Γιατί υπάρχει

Πώς γίνεται



# Τι είναι η προσομοίωση?

Φυσικό σύστημα

Μοντέλο = εξισώσεις

Εξελίξη συστήματος

Βγαζω αποτελέσματα



# ΓΙΑΤΙ ΚΑΝΟΥΜΕ ΠΡΟΣΟΜΙΩΣΗ?

- Ένα smartphone η μια εφαρμογή
- Ένα αυτοκίνητο
- Βηματοδότη
- Κυκλοφορίας πόλης
- Air «ΝΤΟΥΣ»
- Ηλεκτρονικά σε δορυφορο
- Ιατρικό ανιχνευτή
- ακτινοβολία

# Τι είναι προσομοίση;

z Φτιαχνουμε μοντελα

y Του ανιχνευτη

x Γεωμετρια

x Υλικα

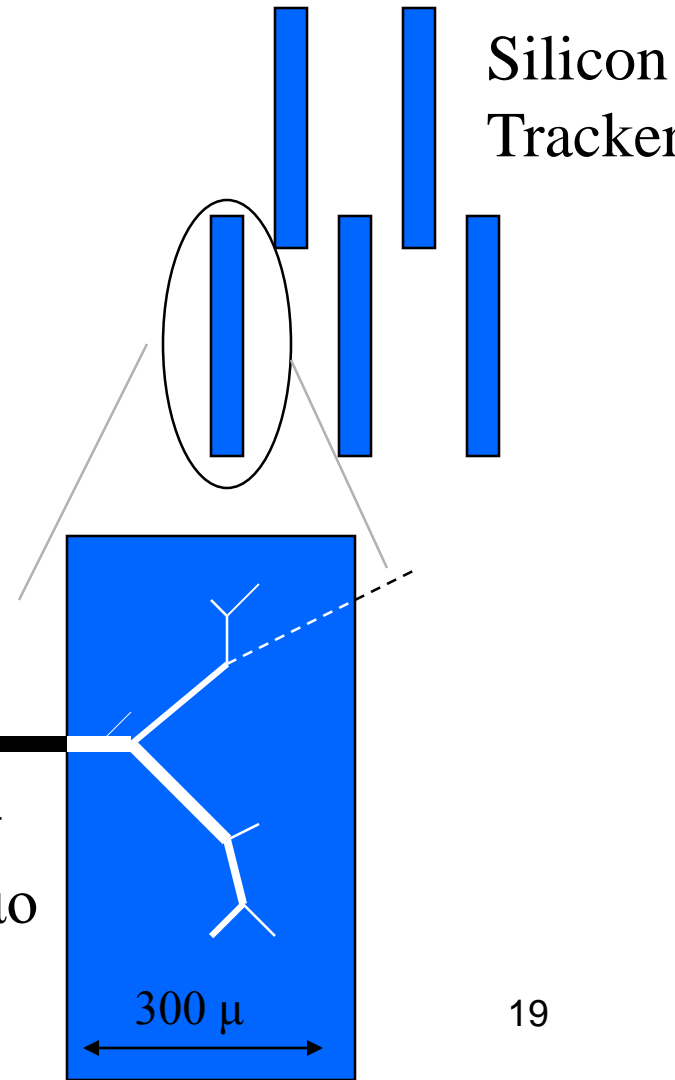
y Των αλληλεπιδρασεων

x Καθε γνωστου τυπου

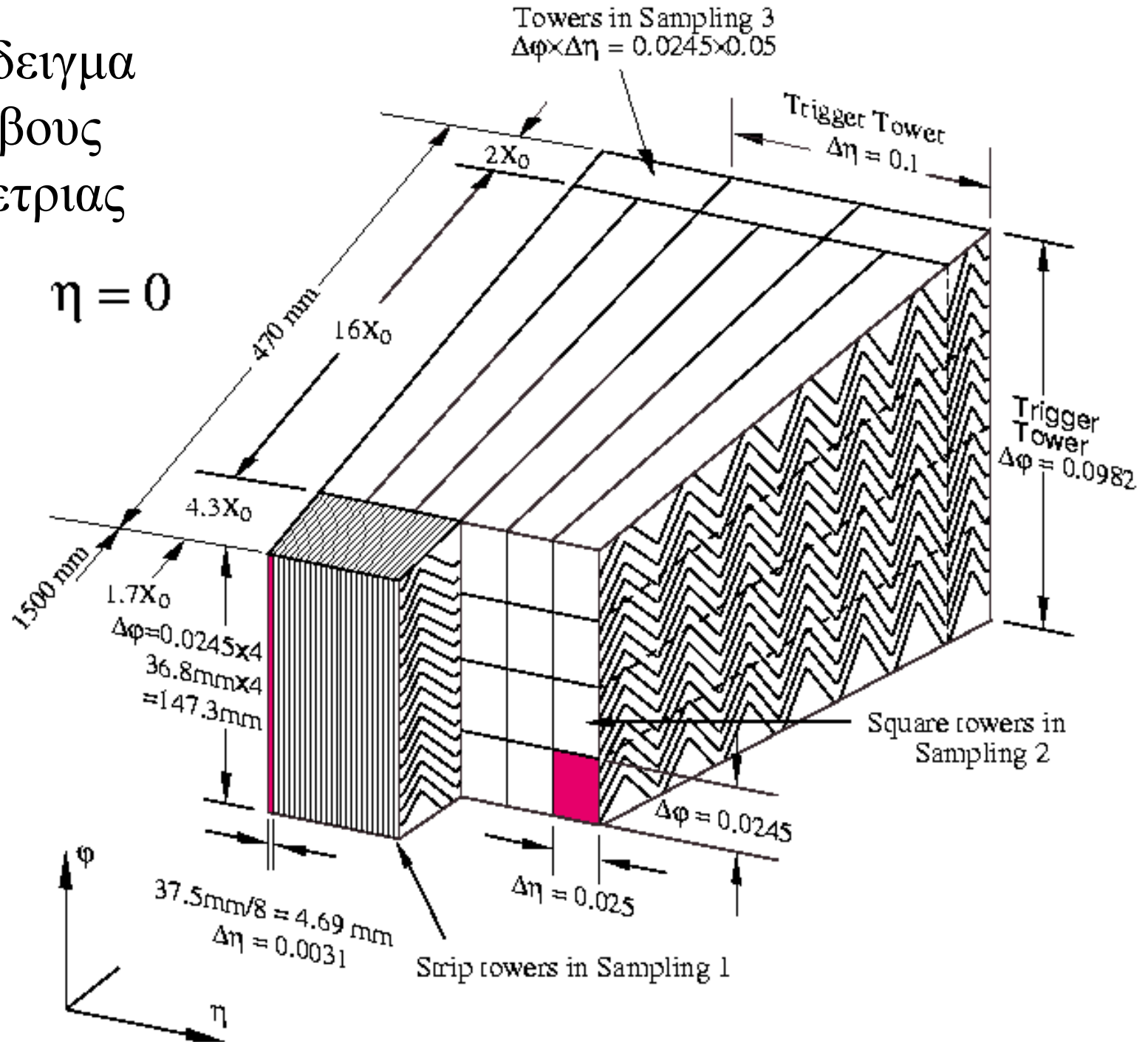
- Ηλεκτρομαγνητικου
- Υσχηρου πυρηνικου

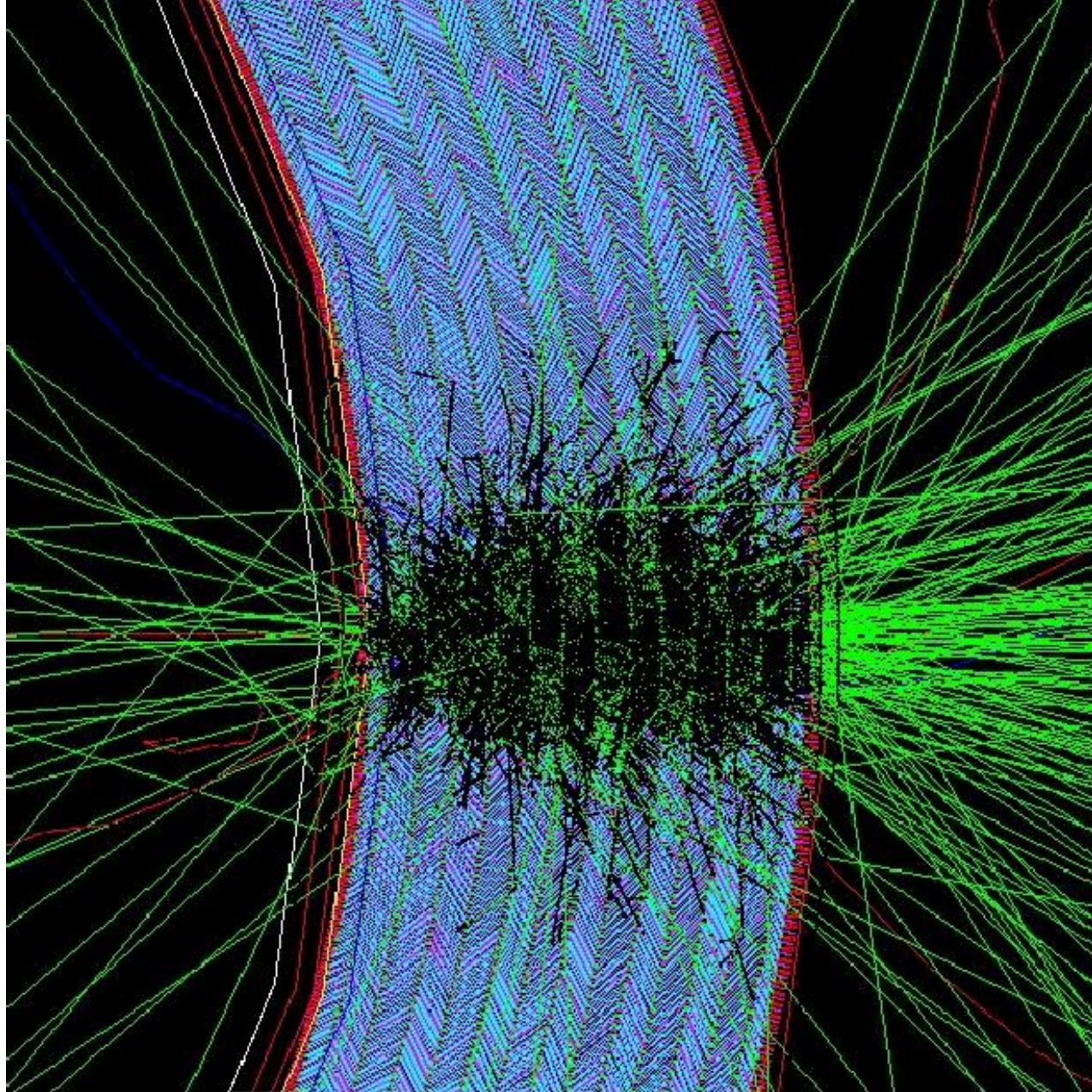
$$\sigma_{\text{συνολο}} = \sum \sigma_{\text{φαινομενου}}$$

2.5 MeV e<sup>-</sup>  
ηλεκτρονιο

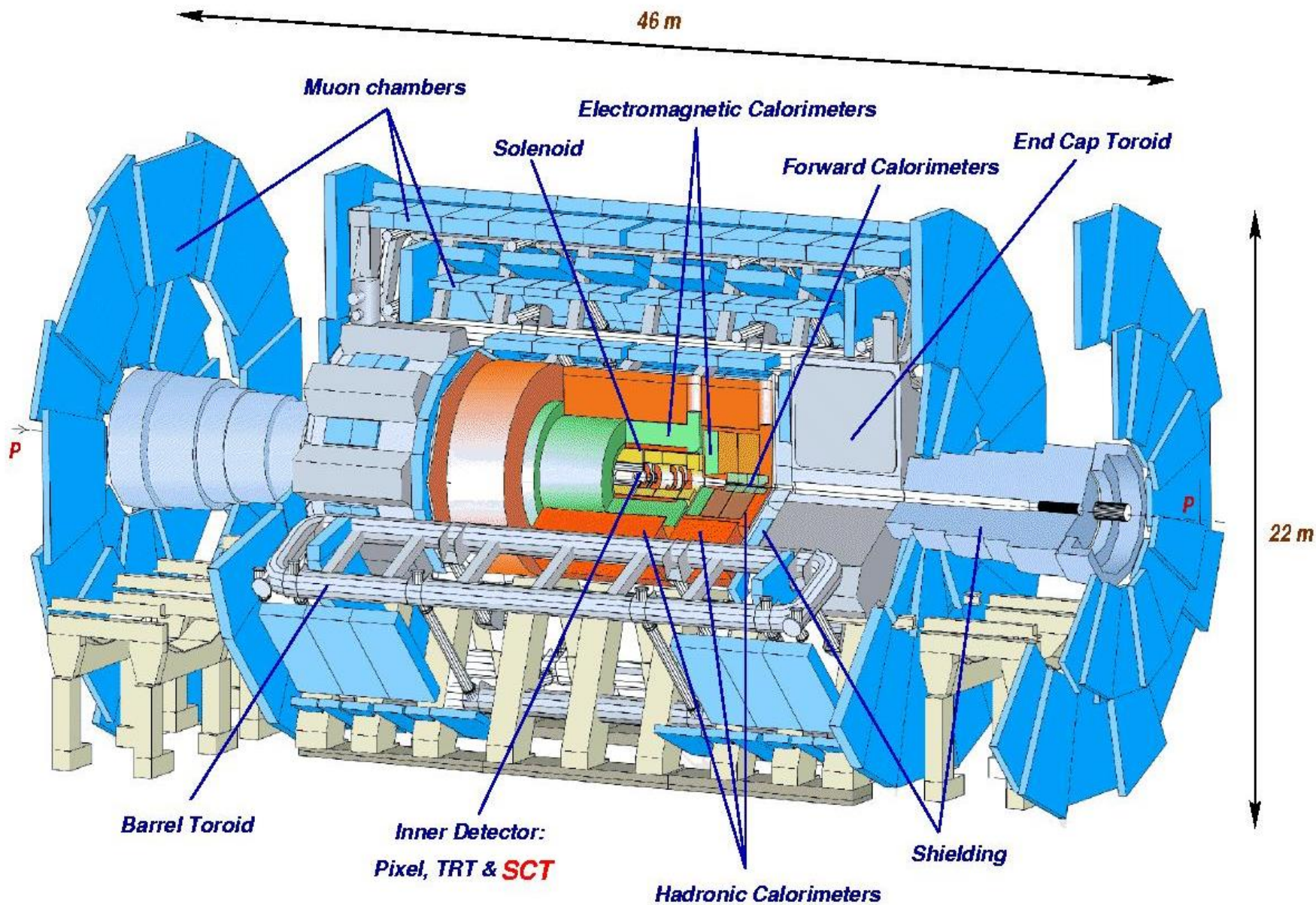


# Παραδειγμα ακριβους γεωμετριας





# Οι ανιχνευτές του ΑΤΛΑΣ





From:  
Cosmus  
Project,

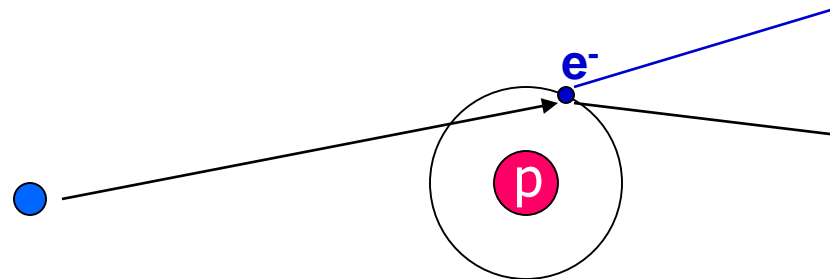
AIRES  
Project,

Sergio  
Sciutto

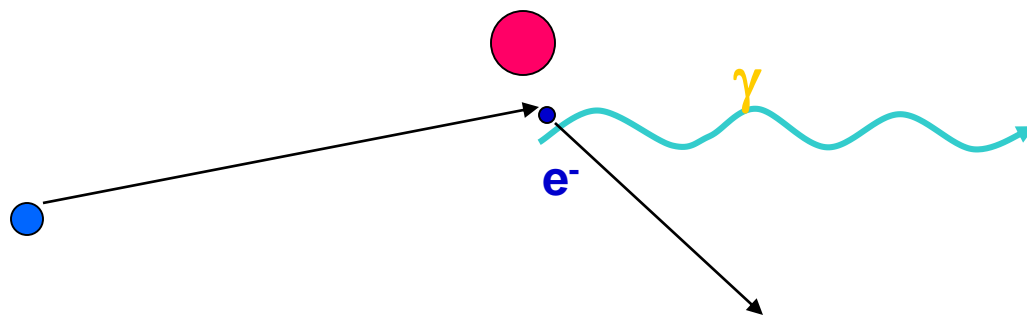
# Βασικές Αλληλεπιδρασεις

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

x Ιοντισμός

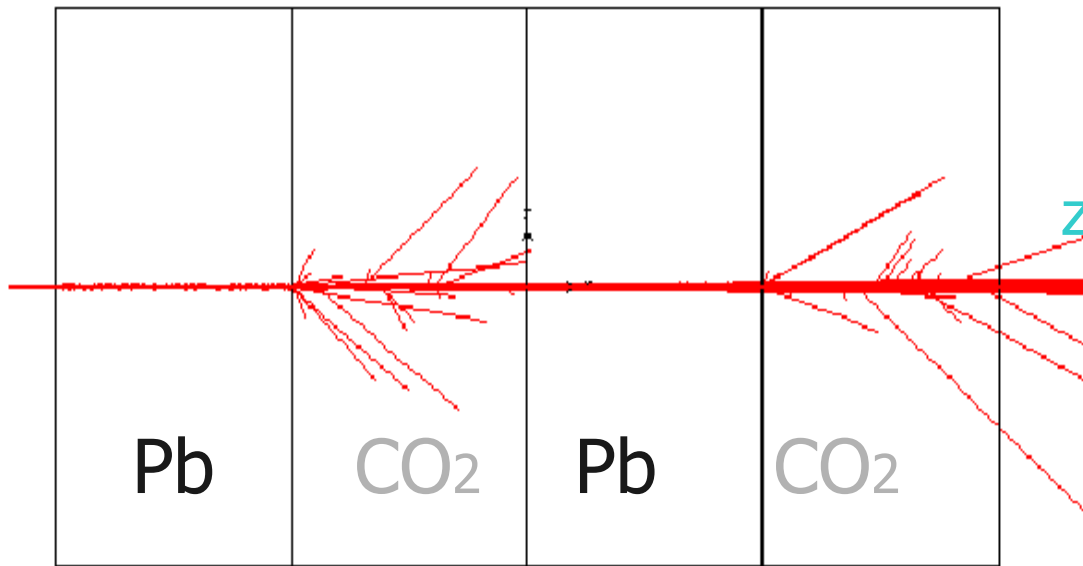


x Bremsstrahlung





# Ενα απλο παραδειγμα



z Στο μολυβδο  
παραγονται πολλα  
δευτερευοντα σωματιδια

y Τα περισοτερα μενου  
κοντα,

y Μερικα ξεφευγουν.

z Το διοξειδιο του  
ανθρακα, σαν αεριο, εχει  
μικρη πυκνοτητα

y Οσα σωματιδια φτανουν η  
παραγωνται, πανε μακρια

y Παραγονται λιγοτερα

## GEANT 3

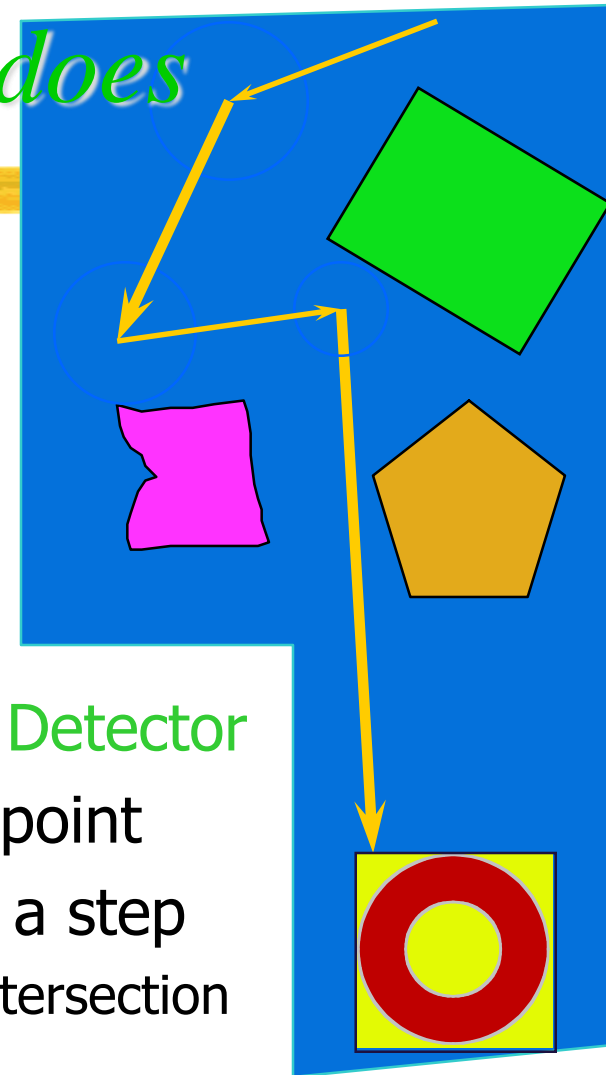
# Geant4 geometry: what it does

## Describes a Detector

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

## Navigates in Detector

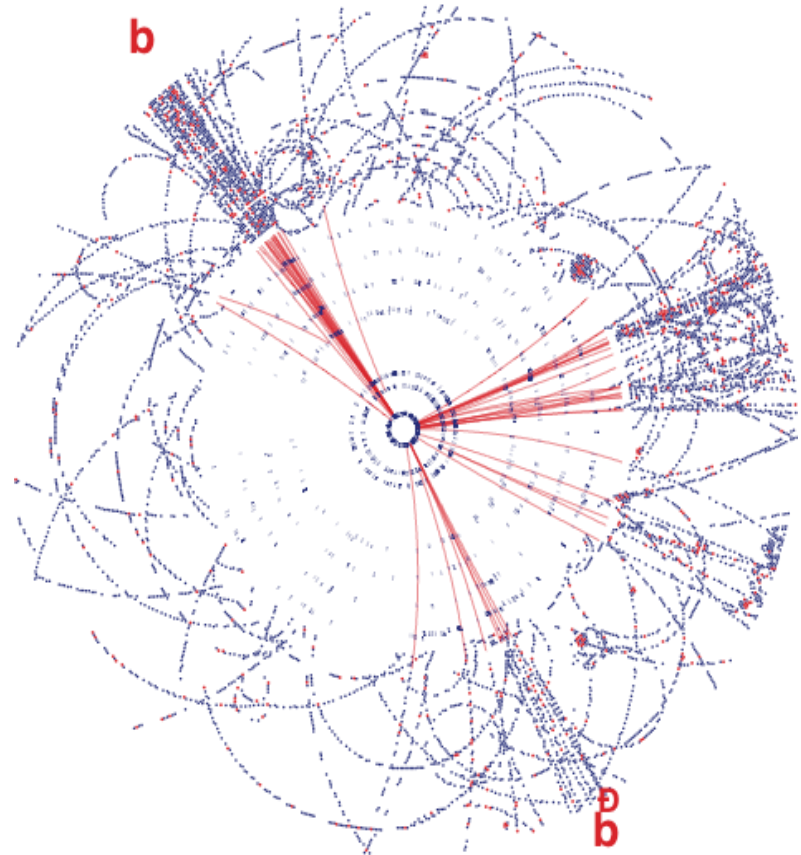
- z Locates a point
- z Computes a step
  - y Linear intersection



# Atlas : Physics Signatures and Event Rates

- ❑ Οι δεσμες διασταυρονονται 40 MHz
- ❑  $\sigma_{inelastic} = 80 \text{ mb}$ 
  - Σε καθε περασμα πολλές 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  
 $H \rightarrow b\bar{b}$



# Γιατι προσομοιωση ?

- z Για να σχεδιασουμε τους ανιχνευτες
- z Για να ετοιμασουμε τις μεθοδους ανακατασκευης
- z Να χαρακτηρισουμε τον ανιχνευτη
- z Να αξιολογήσουμε την σημαντικότητα των γεγονοτων
- z Απο πρωτη ιδεα, ως την τελευταια λεπτομερεια
- z Πριν αρχισει να τρεχει το πειραμα
  - y Για να ειναι ετοιμο για την αρχη – π.χ. την πρωτη δεσμη
- z Συνεχως
  - y ως το τελος των πειραματων (μεγαλη και διαρκης αναγκη)

# Παραδειγμα σύγκρισης

z Δειγματοληπτικό θερμιδομετρο

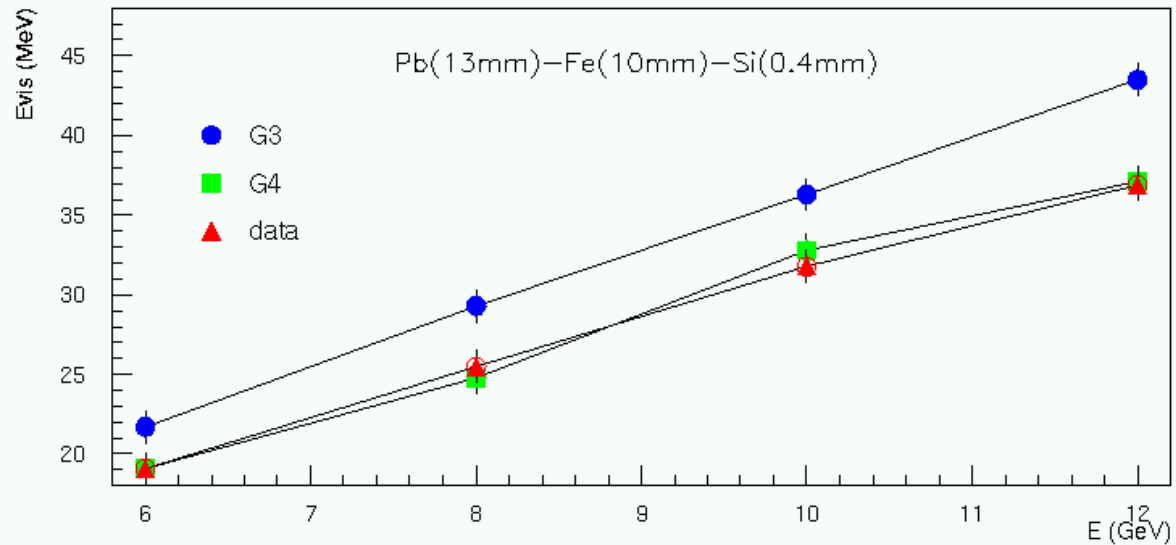
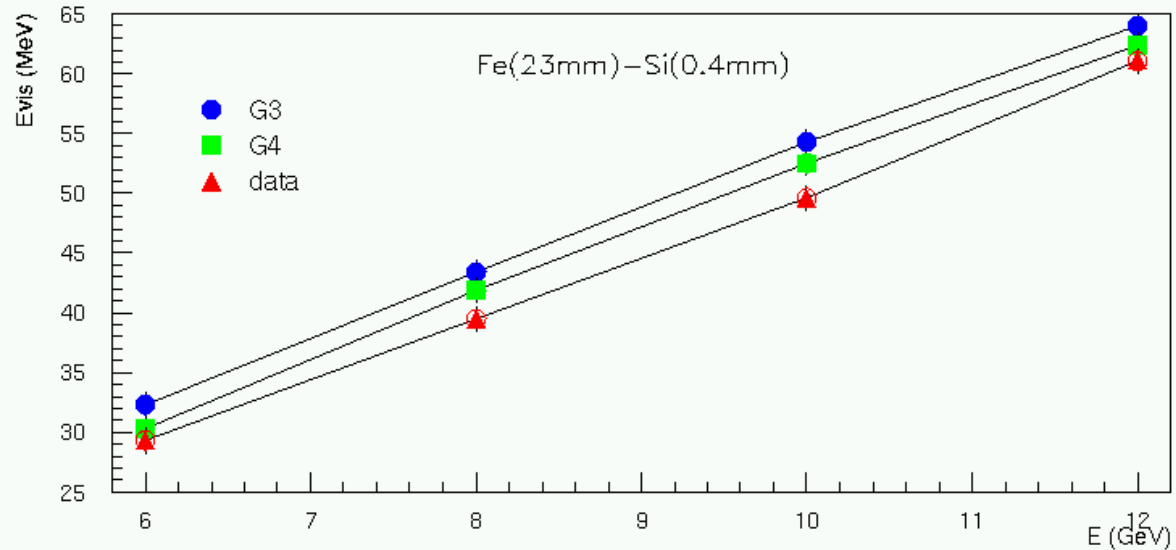
y Ορατή ενέργεια

z Δοκιμες

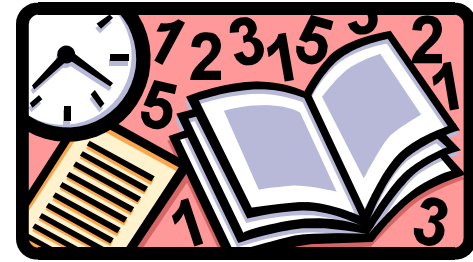
y Για όλες τις ηλεκτρομαγνητικές αλληλεπιδρασεις, ηλεκτρονίων (e-) φωτονίων

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

Pb-Fe-Si e- 6 8 10 12 GeV 100 evts



# One area: Tracking



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

y Determine the **step length**

- x Corresponding to the applicable physics processes
- x Checking if it crosses a geometrical boundary

y Model the **final state** of the track,

- x Advancing it, potentially in an EM field,
- x Applying the actions of the physics processes,
  - which can create **secondary** particles.

y **Deposit** energy in current position ( 'hit' ).

# Υπαρχει απλη λυση ?

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

y Για μερικες απλες ερευνες, Ναι

y Για πολλες (τις περισσοτερες) χρειαζεται ολη η εικονα

x Υπαρχει κατανομή τιμων

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

z Με ποιο σφαλμα ξερετε την ταδε ενεργεια?

# GEANT 4



- z Detector simulation **tool-kit** for HEP
  - y offers alternatives, allows for tailoring
- z Software Engineering and OO technology
  - y provide the method for building, maintaining it.
- z **Requirements** from LHC/HEP and beyond:
  - y medical and space science applications
- z **World-wide collaboration** (120 physicists & engineers, >10 countries)
- z **Open source** – anyone can use & improve it!



# Άλλες χρήσεις προσομοίωσης



## z Πλανητες / δωρυφωροι

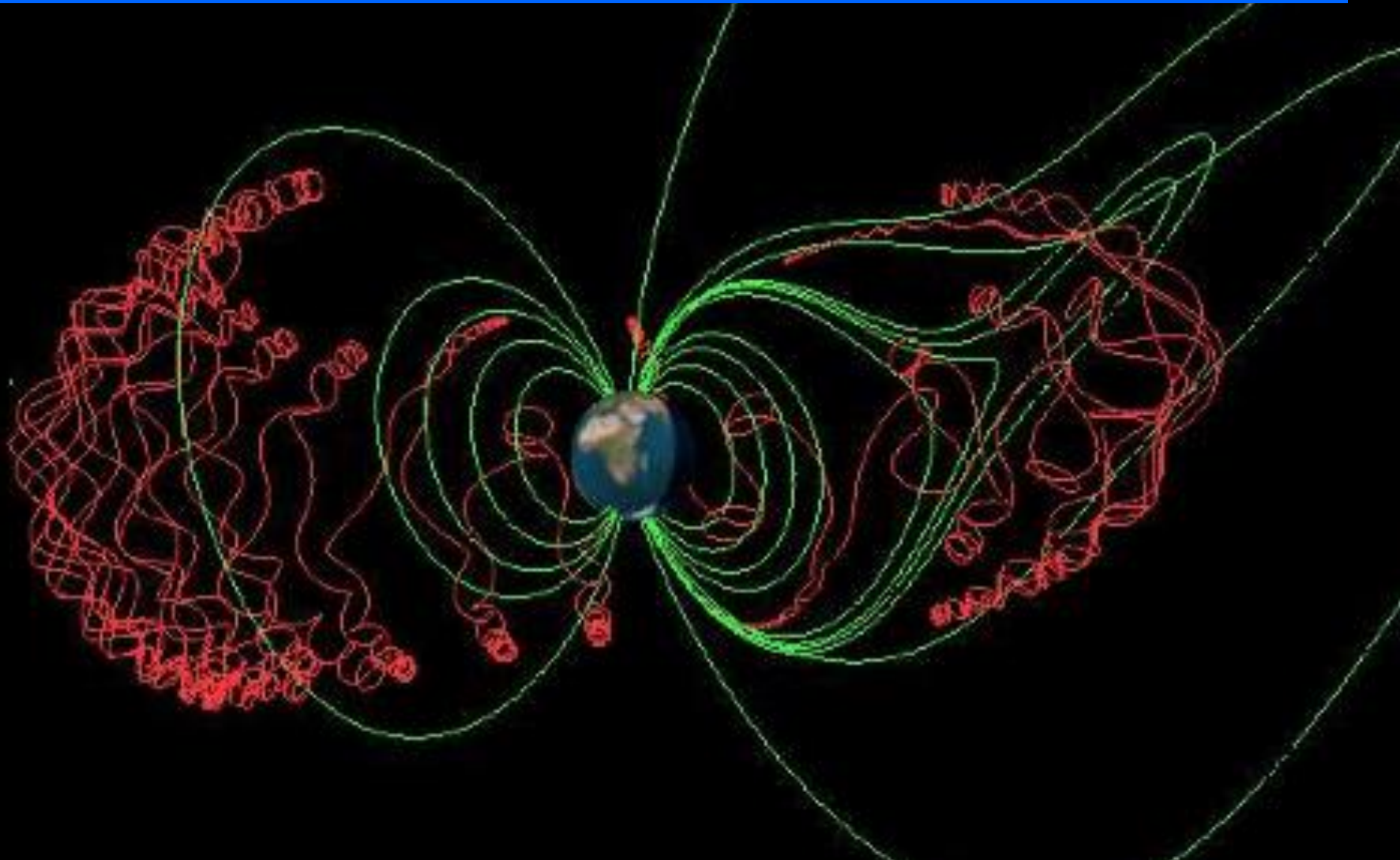
- y Αλληλεπίδραση ηλεκτρονίων/προτονίων με ατμοσφαιρα και μαγνητικο παιδιο της Γης

## z Ιατρικη

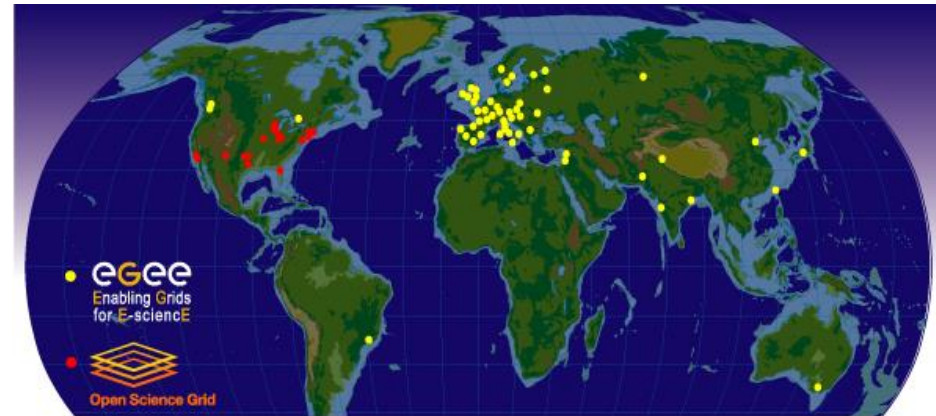
- y Για ιατρικη διαγνωση

- y Ραδιο-ιατρικη – θεραπεια με πρωτονια

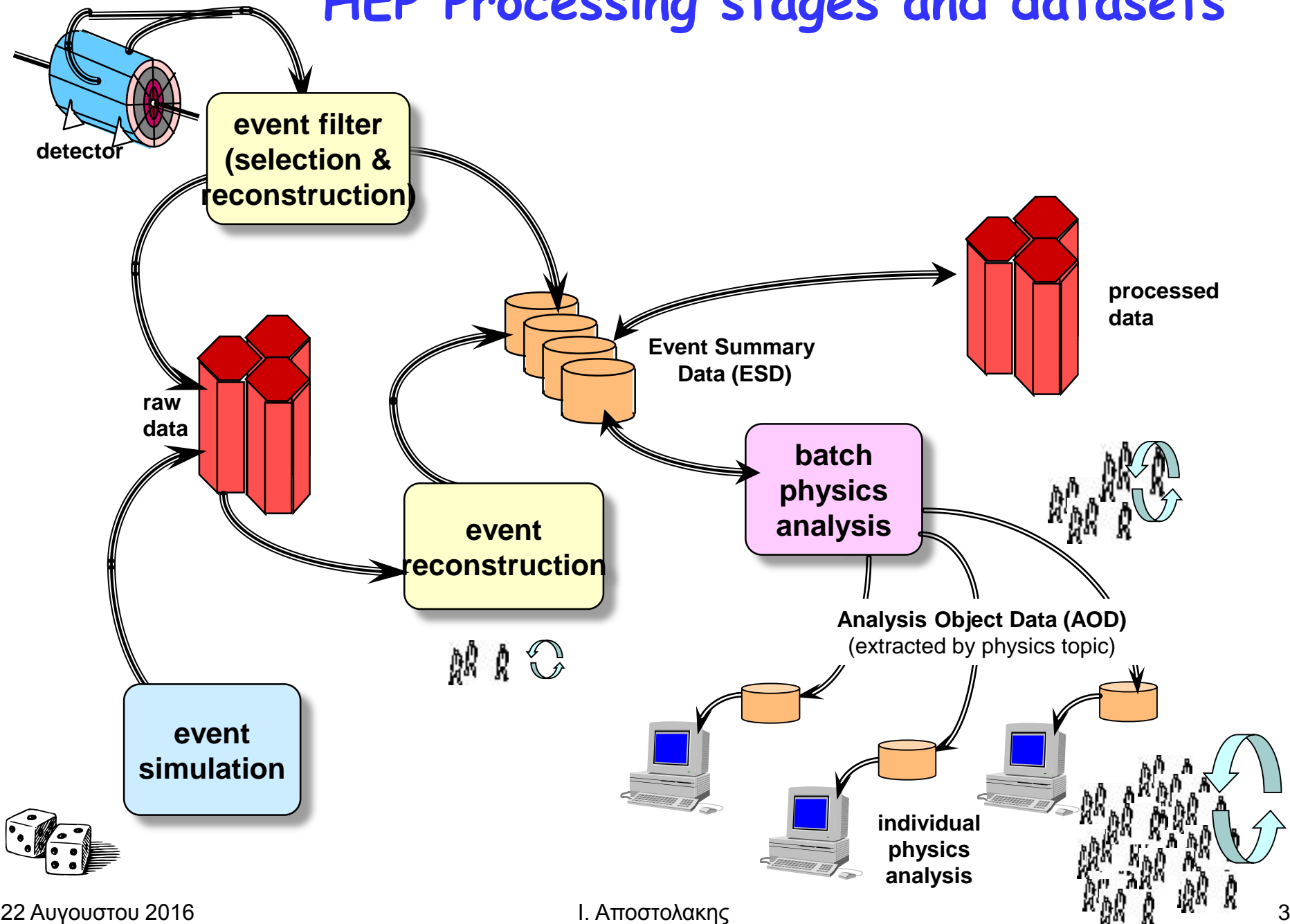
Trapped particles in Earth's Magnetic field



# Αναλυση δεδομενων



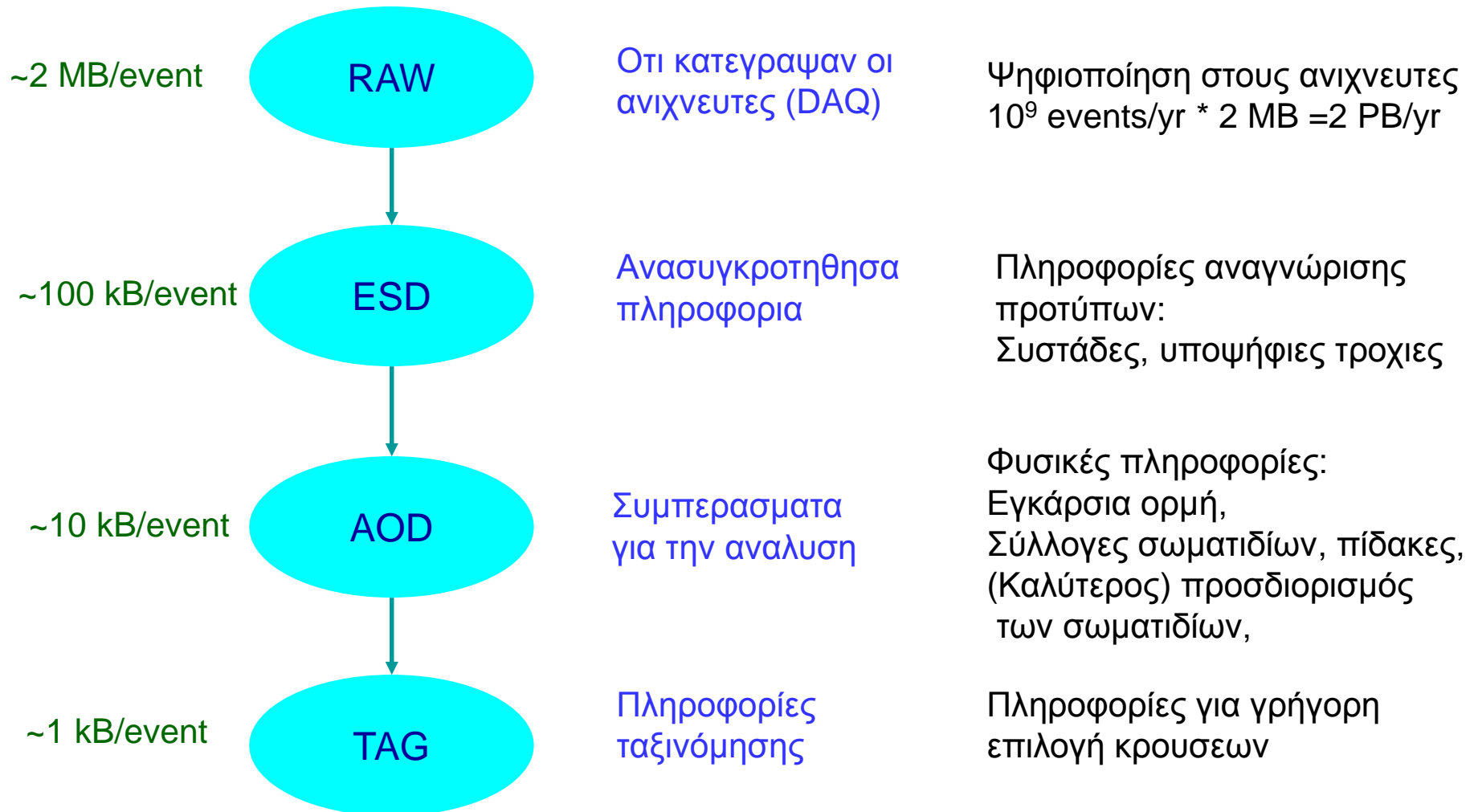
# HEP Processing stages and datasets

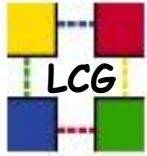


# Αναλυση:

- ❑ Επειτα απο την ανακατασκευη
- ❑ Χρησιμοποιει τα 'δεδομενα' (Data) προιοντα της ανακατασκευης
  - Ιεραρχεια 'δεδομενων' απο Ολικα (RAW), ως συνοπτικα/συμπερασματικα (AOD)
- ❑ Οι ομαδες ενος πειραματος χρησιμοποιουν τα δεδομενα
  - Στο CERN (δυσκολο) η που ? ... στο Πλεγμα (GRID)
- ❑ Hyratia: μικρο κοματι αναλυσης πειραματος στο σχολειο
  - [Εισαγωγη](#), Κοσμος/[Portal](#) Εφαρμογη - Δ.Φασουλιωτης (Εθν.Καπ. Παν. Αθηνas), Α. Αλεξοπουλος(CERN)
  - <http://hyratia.iasa.gr/en/index.html>
  - <http://indico.cern.ch/conferenceDisplay.py?confId=257353#2013-07-08>

# Ιεραρχία Δεδομενων (Data)





# Distribution of Computing Services

## Summary of Con

All experiments - 2008

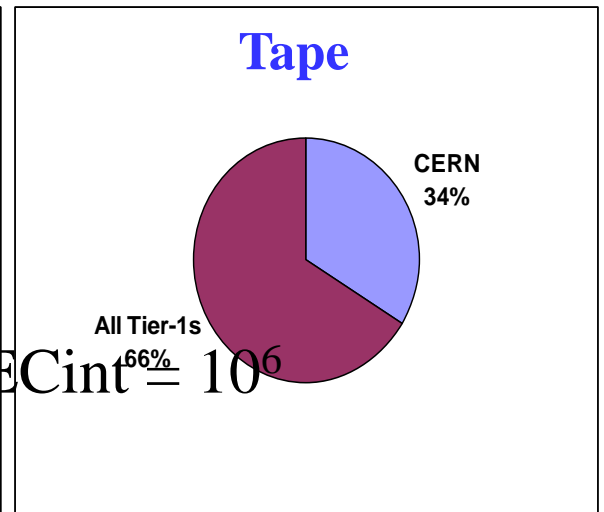
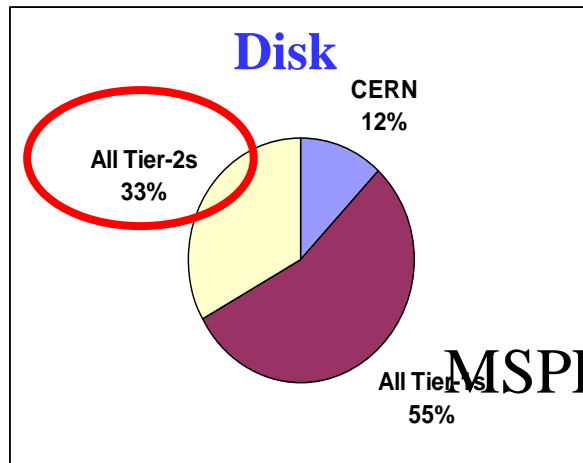
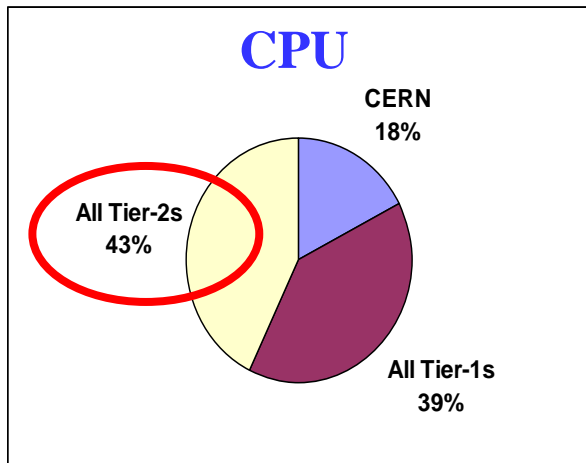
From LCG TDR - June 2008

Τα νέα δεδομένα θα  
μεγαλώνουν 15 PetaBytes

καθε χρόνο – με δυο αντιγραφα

Περίπου 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



Μεγάλο ποσοστό των υπολογιστών και  
δίσκων είναι διαδεδωμένα

σε 120 υπολογιστικά κέντρα

MSPECint = 106



# Λυση : το Πλεγμα (Grid)

- Χρησιμοποιούμε 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

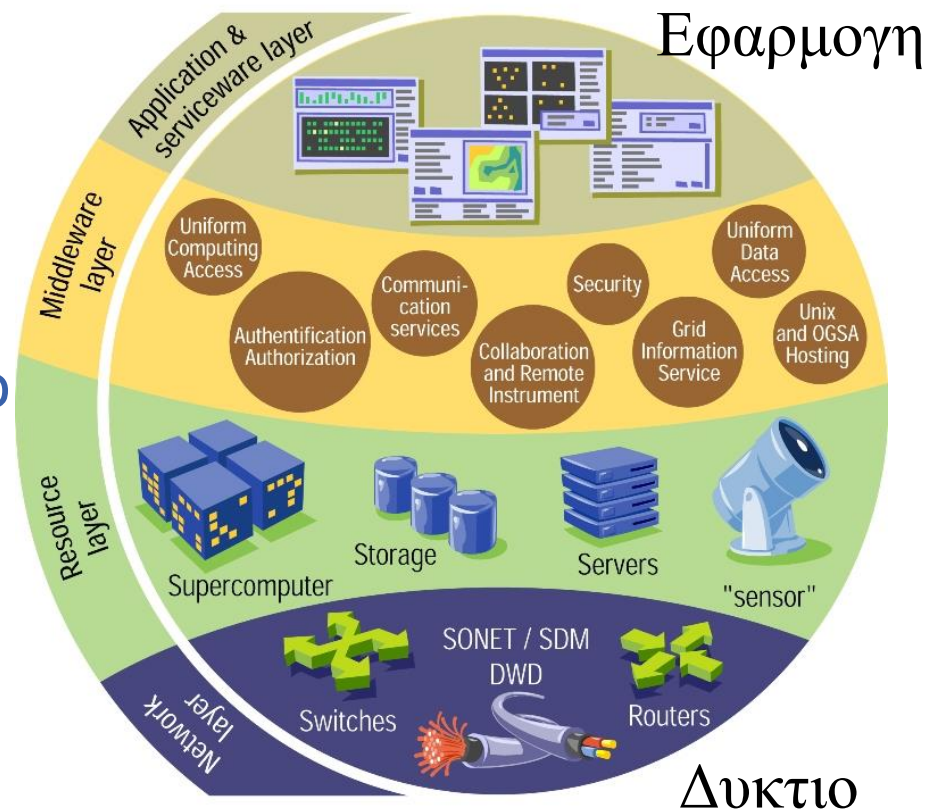
The **Grid** είναι μια υποδομή που παρέχει αδιάλειπτη πρόσβαση σε υπολογιστική ισχύ και χωρητικότητα αποθήκευσης δεδομένων καταναεμημένη σε όλη την υφήλιο



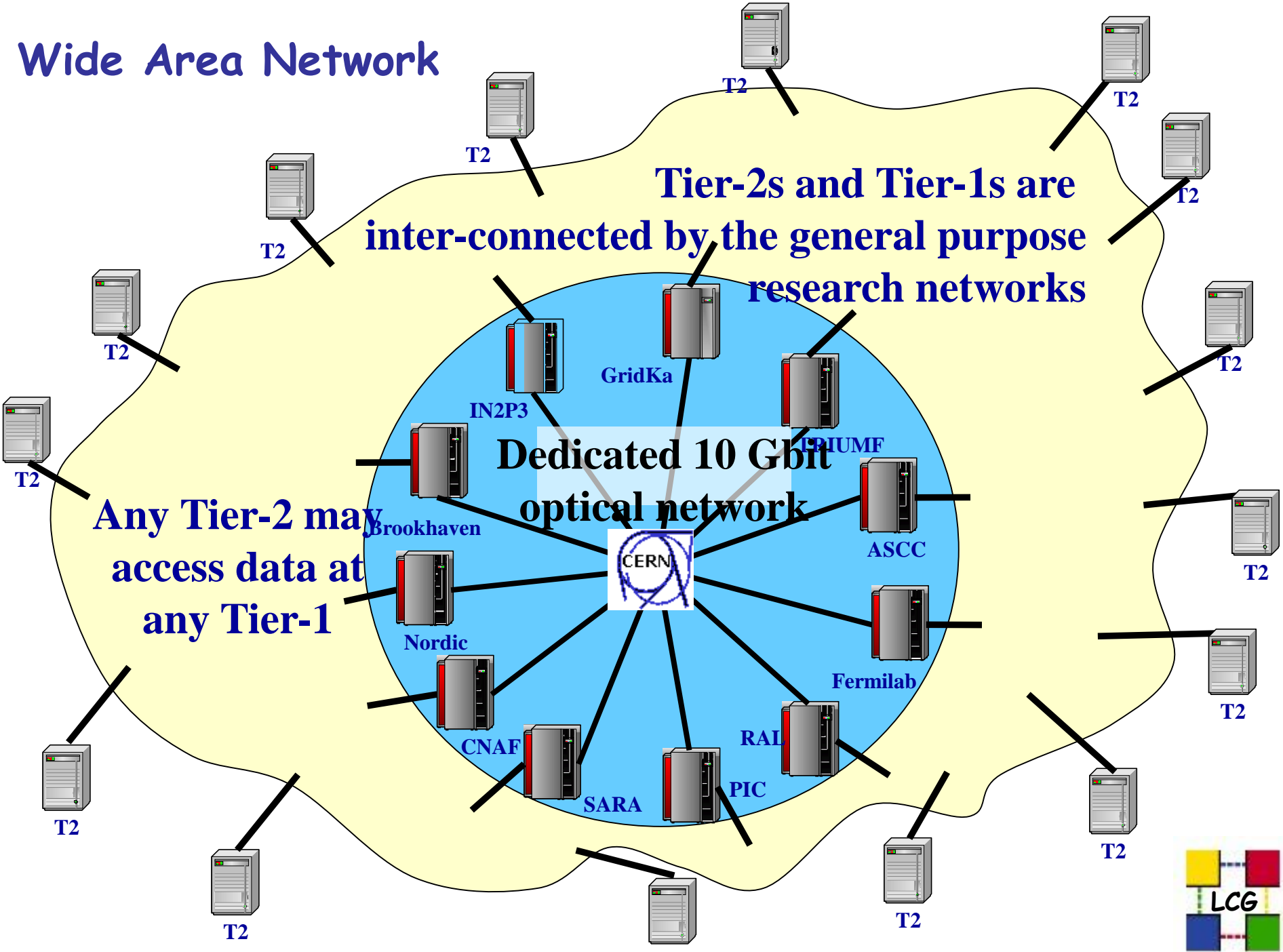


# Πως δουλεει το Πλεγμα (Grid)

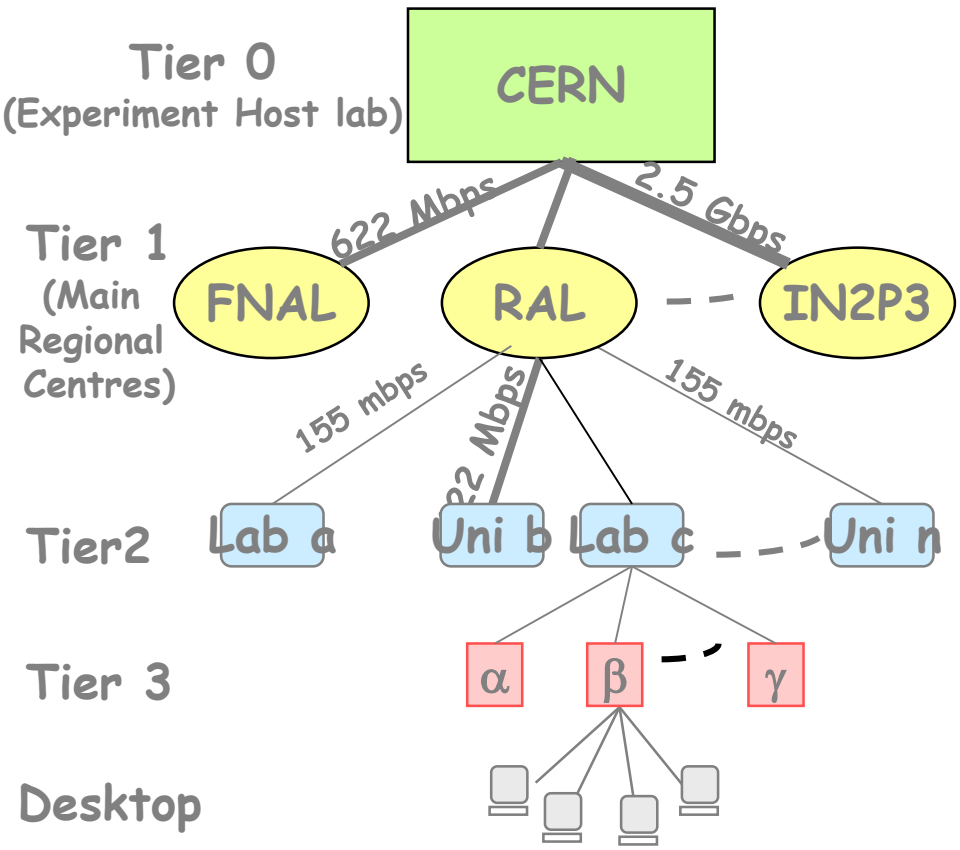
- Βασίζεται σε ειδικευμενο λογισμικο, το λεγομενο **μεσο-λογισμικο** (middleware).
- Η ιδεα του Middleware ειναι να βρει αυτοματα τα **στοιχεια** που χρειαζεται ο/η ερευνητης, και το **υπολογιστικο δυναμικο** να τα επεξεργαστει.
- Middleware εξισοροπει το φορτιο σε διαφορους πορους και εγκαταστασεις.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



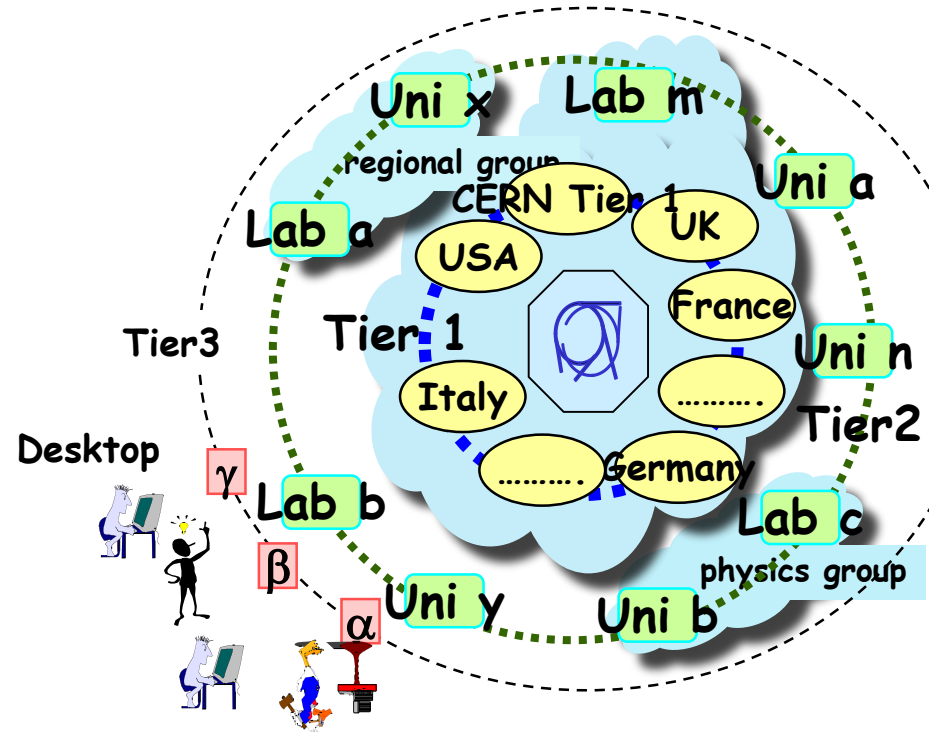
# Wide Area Network



# A Multi-Tier Computing Model

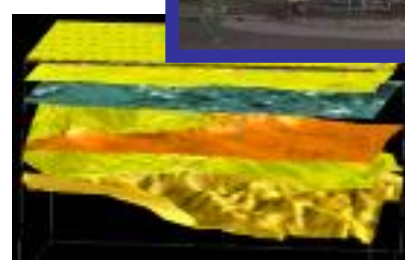
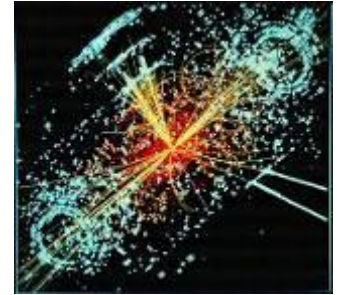


Manager View



User View

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

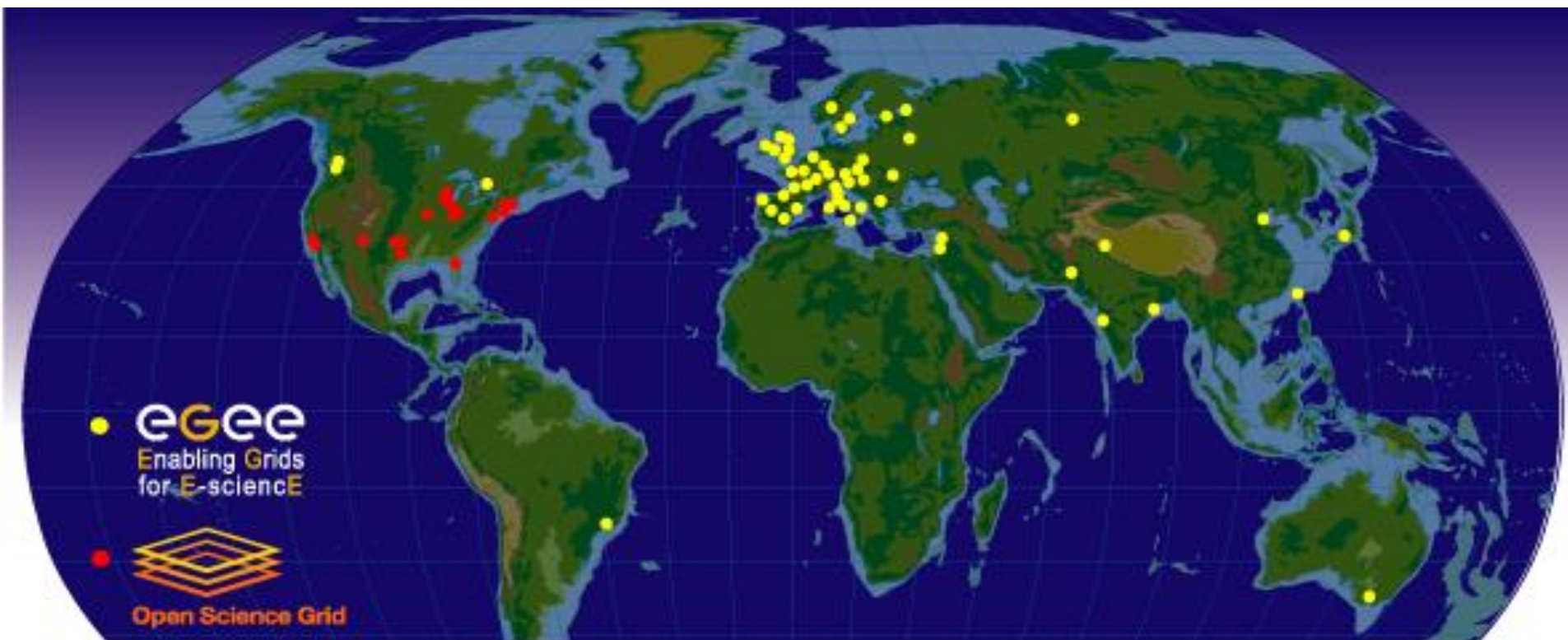




# LCG depends on two major science grid infrastructures ...

**EGEE** - Enabling Grids for E-Science

**OSG** - US Open Science Grid



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

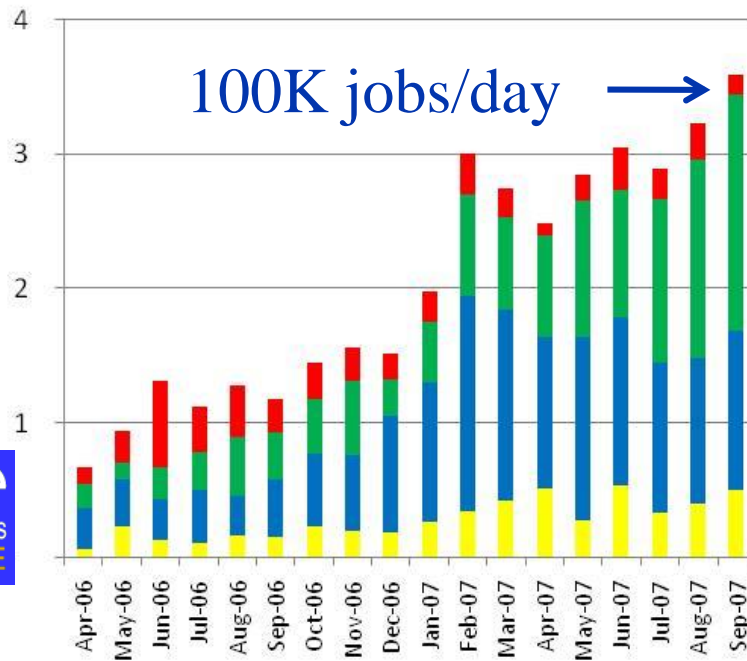


# Grid Activity

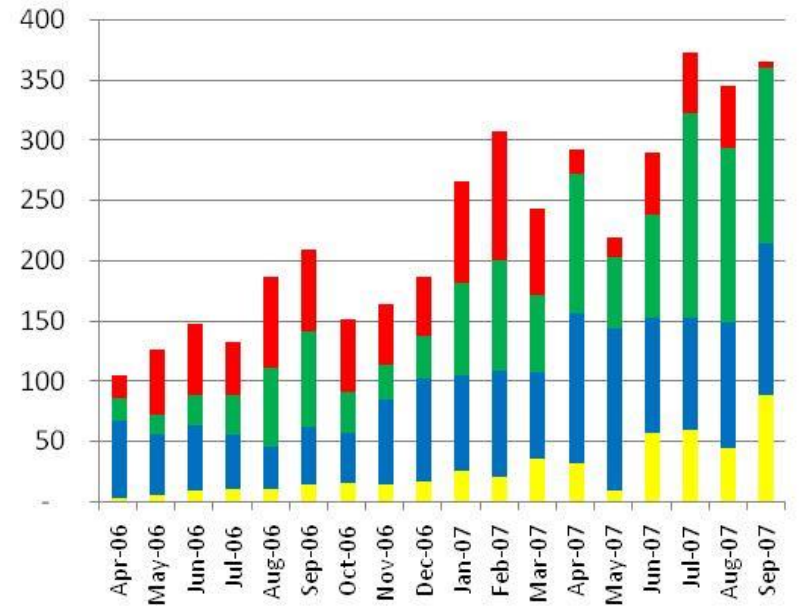
- Συνεχης αύξηση στη χρήση του EGEE και OSG πλεγμάτων(grids)
- All sites reporting accounting data (CERN, Tier-1, -2, -3)
- Συνεχης αυξηση χρησης π.χ. 18 μηνες (ως Σεπ 2007) - επι 5 ο αριθμος των jobs και επι 3.5 ο χρονος χρησης

- lhcb
- cms
- atlas
- alice

LHC Jobs Run (millions)

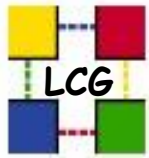


LHC CPU Usage (M SI2K-days)



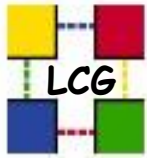
# Οι φυσικοί προγραμματίζουν

- z Ολη η ανάλυση μετρήσεων γίνεται με υπολογιστές
- z Οι φυσικοί ΥΕ χρησιμοποιούν ειδικευμένα προγράμματα
  - y Μερικοί γράφουν μεγάλες ρουτίνες (routines/Fortran, methods/C++)
  - y Οι πιο πολλοί κάνουν μικρές ρουτίνες, για τις δικές τους ανάγκες
- z Όλοι θα χρησιμοποιήσω τα 'εργαλεία' να δουν τις περιλήψεις των μετρήσεων

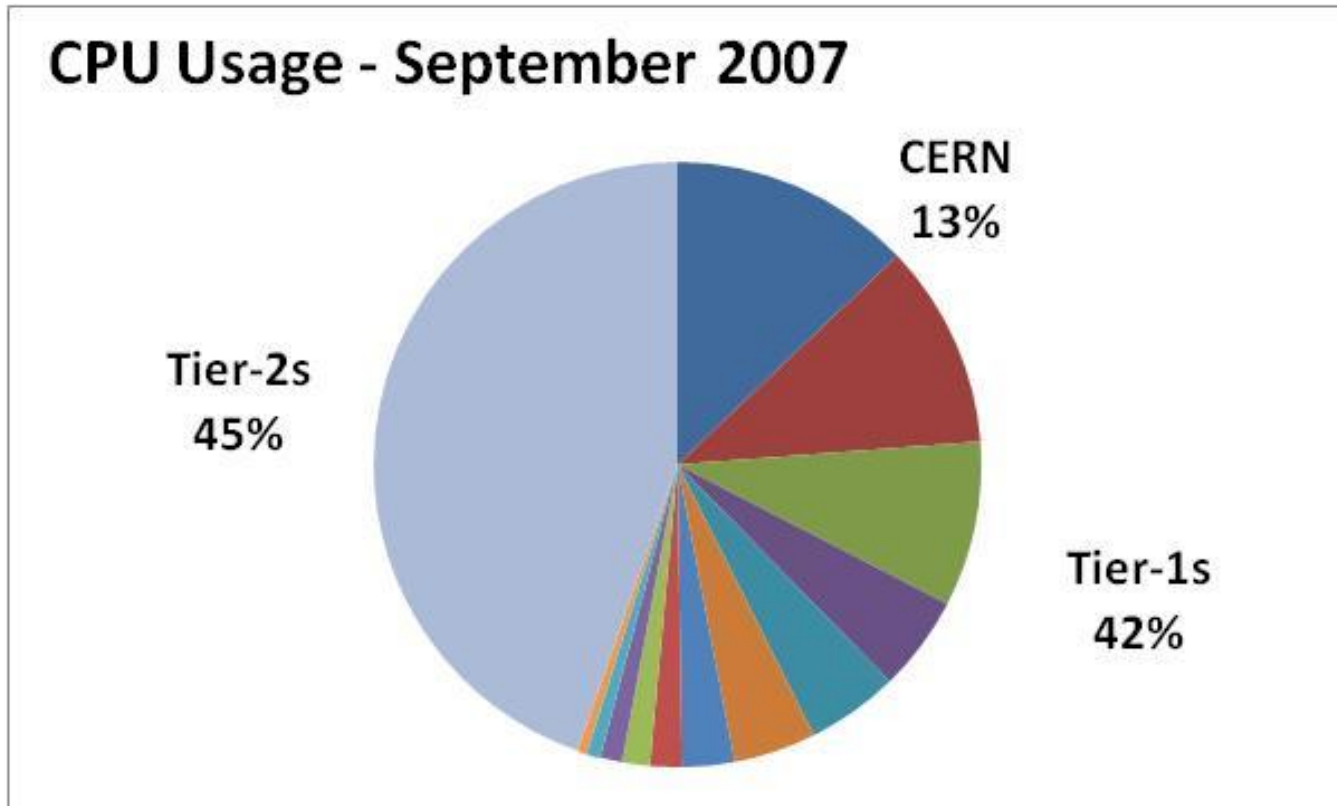


# Backup





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

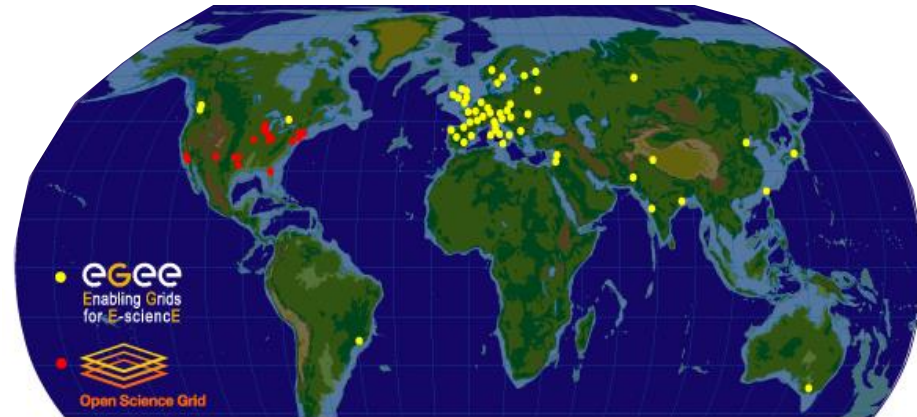


- > 80% of CPU Usage is external to CERN



# 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

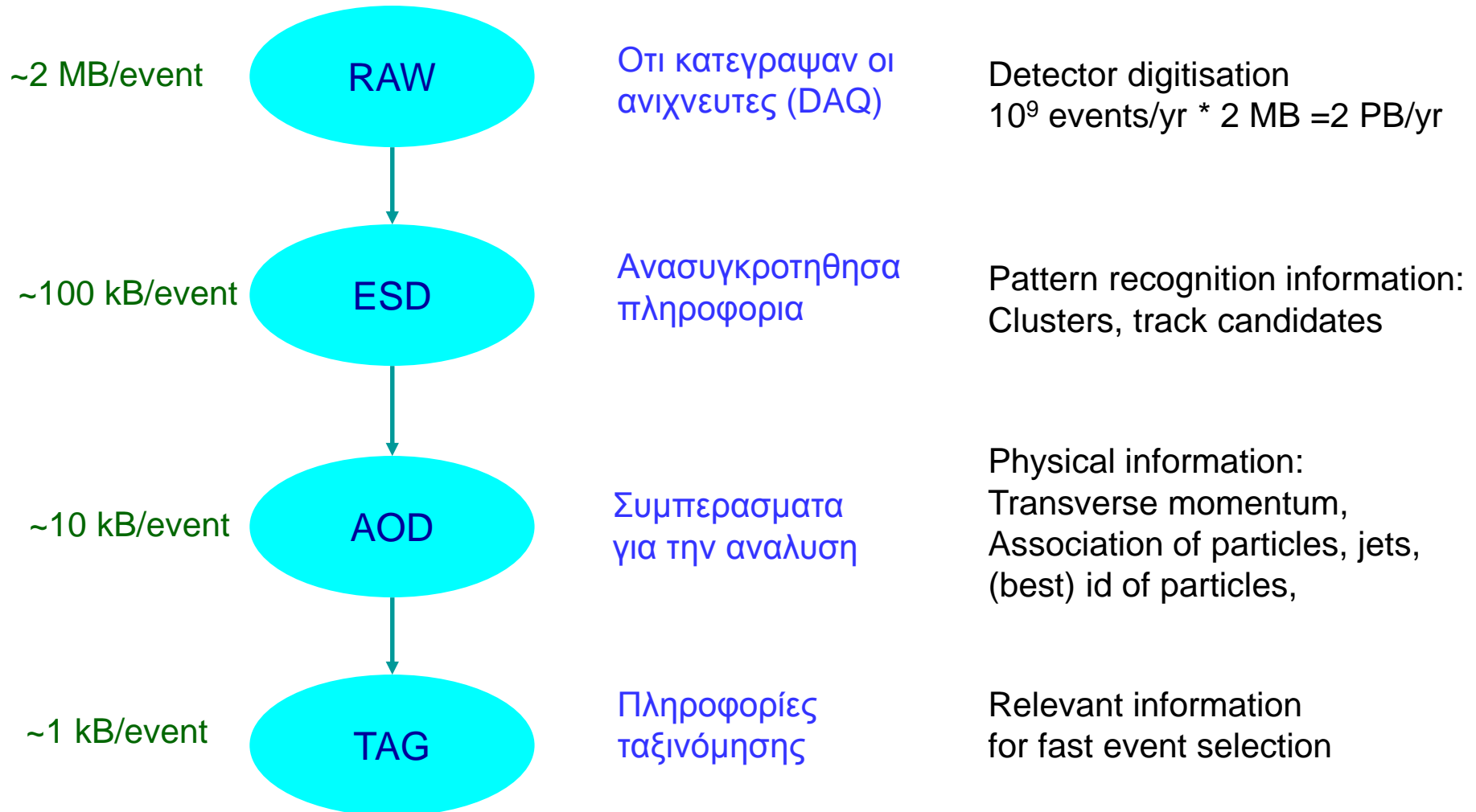


# Backup

A thick, horizontal yellow brushstroke with a textured, painterly appearance, extending across the width of the slide below the 'Backup' title.

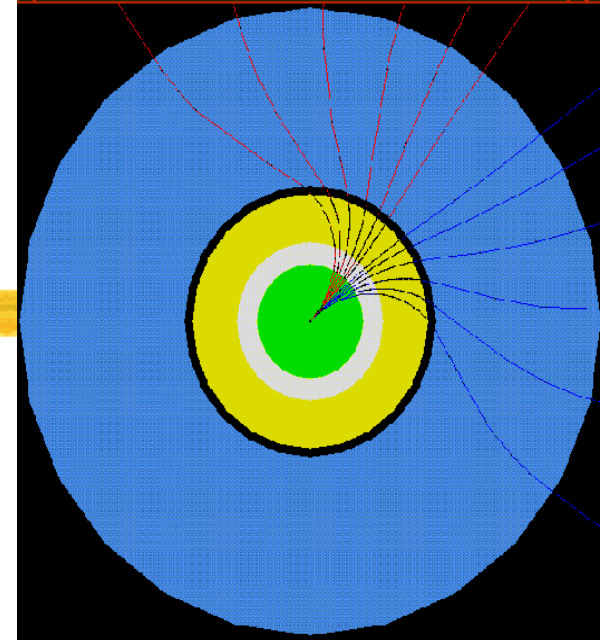
**More on simulation**

# Ιεραρχία Δεδομένων (Data)



# Propagating in a field

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



- z It is possible to tailor
  - y the accuracy of the splitting of the curve into linear segments,
  - y the accuracy in intersecting each volume boundaries.
- z These can be set now to different values for a single volume or for a hierarchy.

# Electromagnetic physics

## z Gammas:

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

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

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

## z Photons:

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

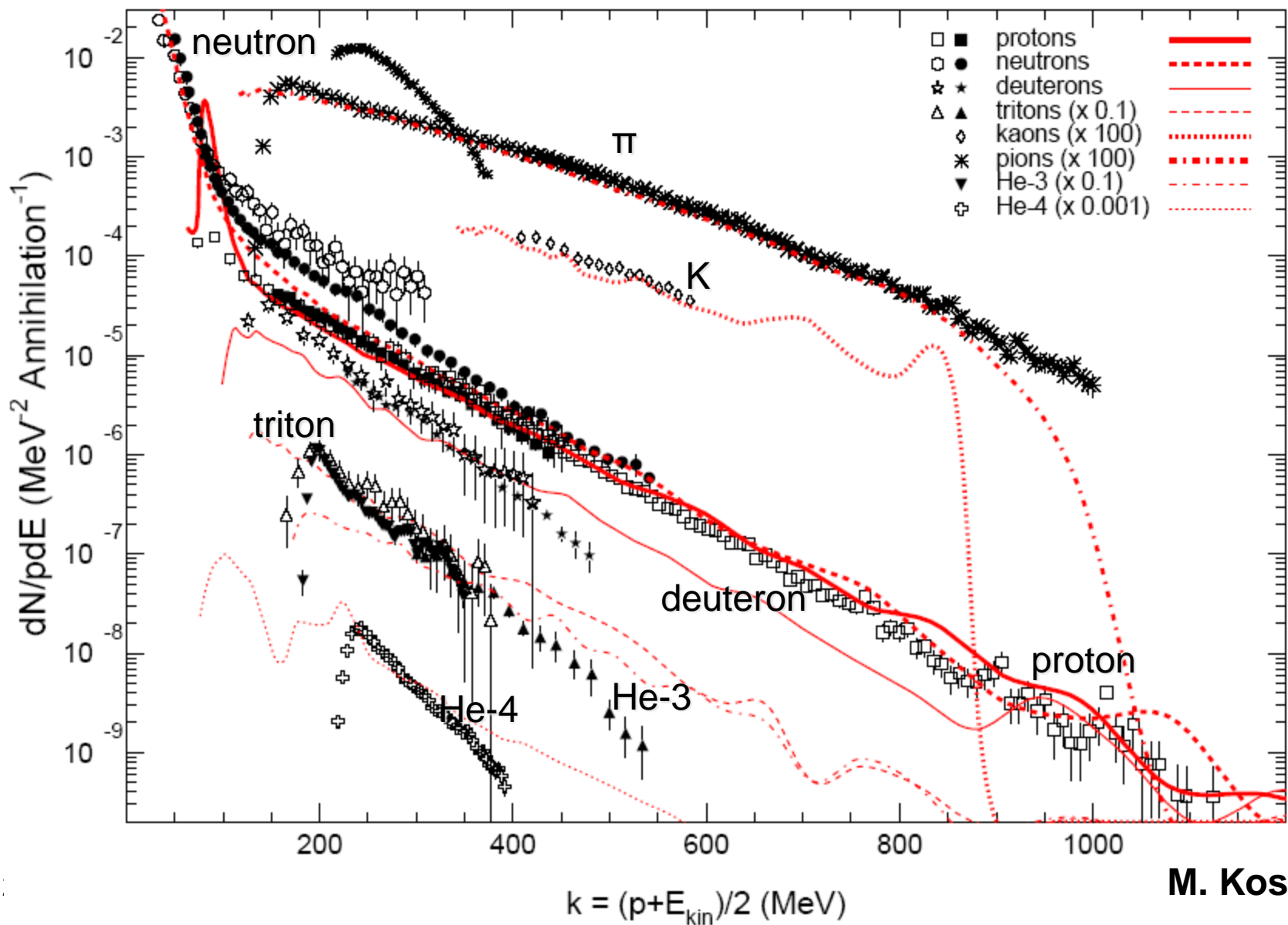
## z High energy muons and lepton-hadron interactions

## z Alternative implementation ("low energy")

- y for applications that need to go below 1 KeV

# Antiproton annihilation - CHIPS Model

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



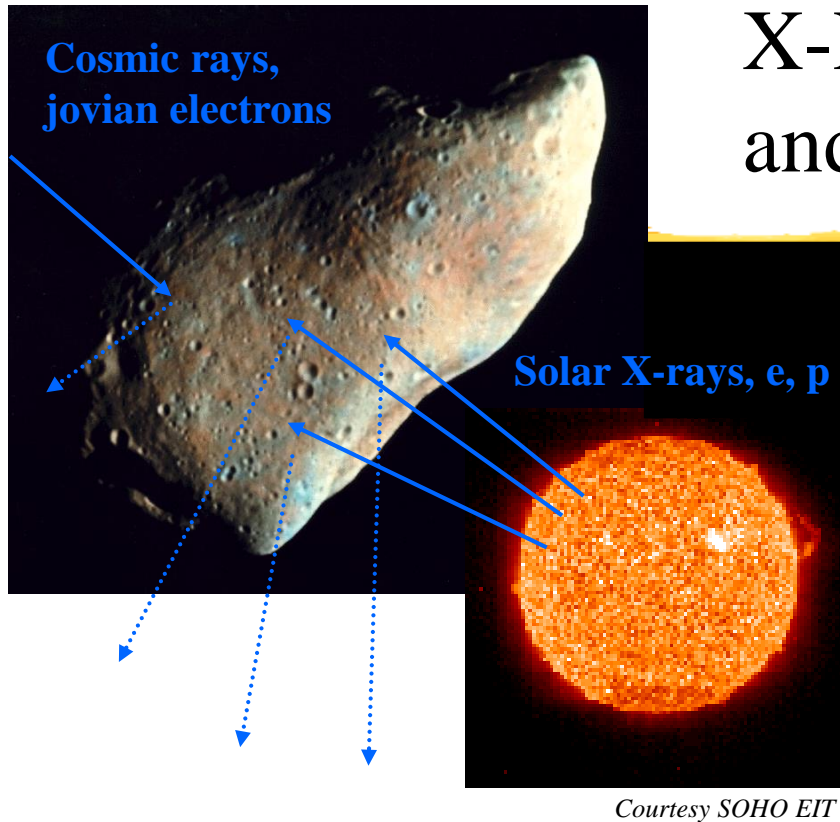
# Simulation ‘packages’



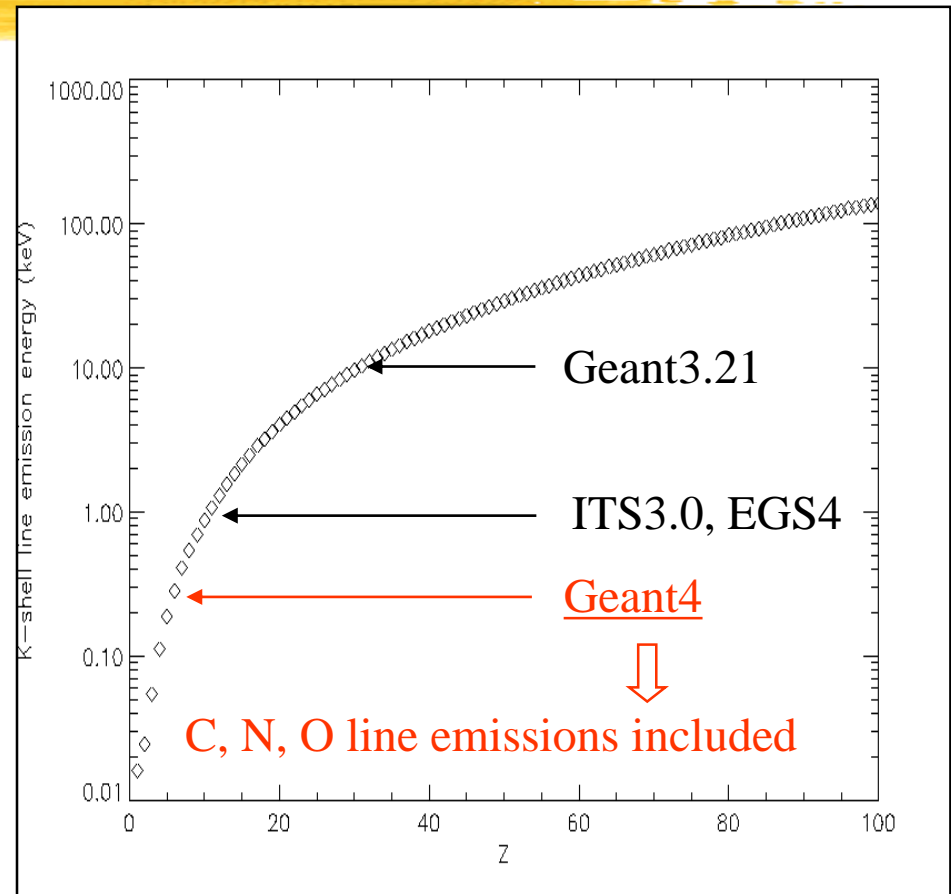
- z Provides the means to simulate
  - y the **physical processes** and
  - y **detector response** of an experiment.
- z As was realised by many in the past,
  - y **most of the parts** needed can be **common** between experiments (eg physics, geometry blocks) .
- z So it makes eminent sense to create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



# X-Ray Surveys of Asteroids and Moons

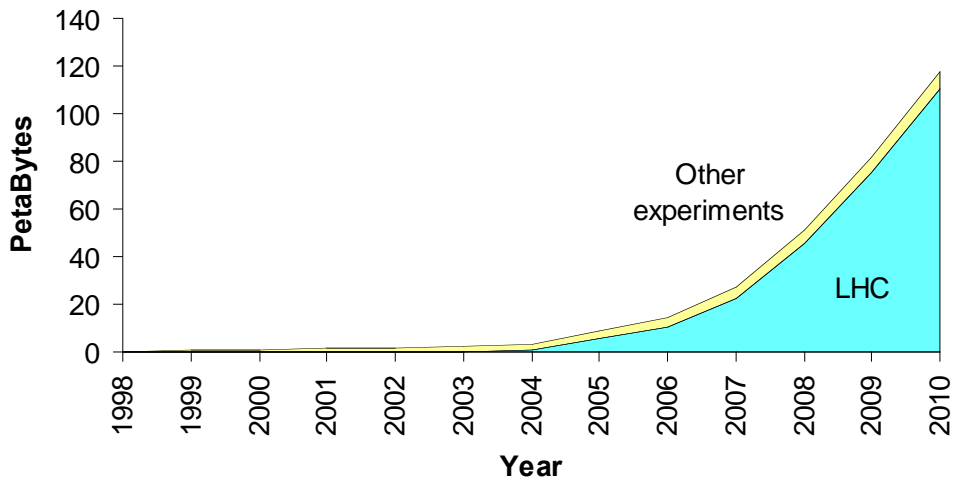


Induced X-ray line emission:  
indicator of target composition  
(~100  $\mu\text{m}$  surface layer)

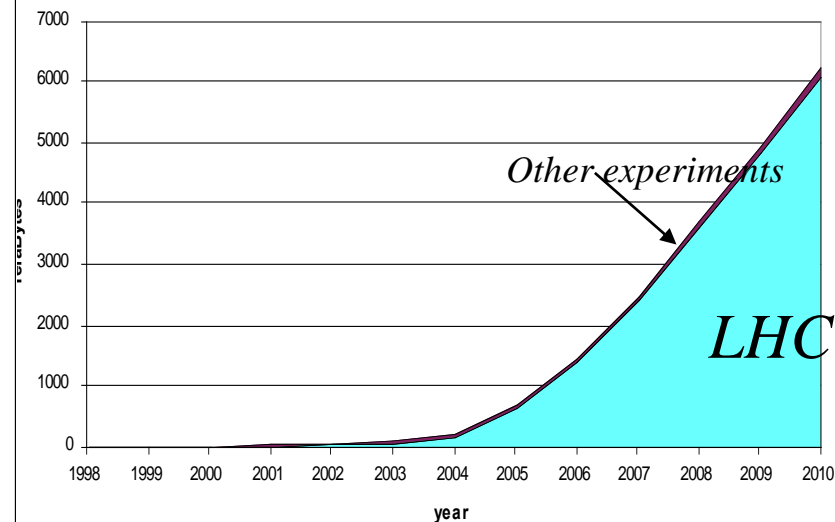


# CERN Centre Capacity Requirements for all

Estimated Mass Storage at CERN



Estimated DISK Capacity at CERN



processing

K 25,000

3 700

8 200

10 100

25,000

34,000

disk

PE

5.0

6.7

tape media

PE

36

48

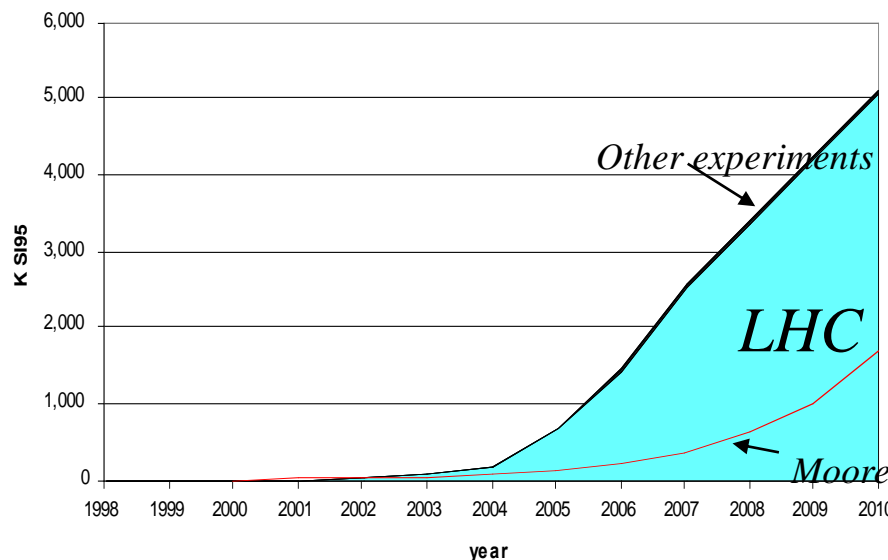
tape I/O

GI

39

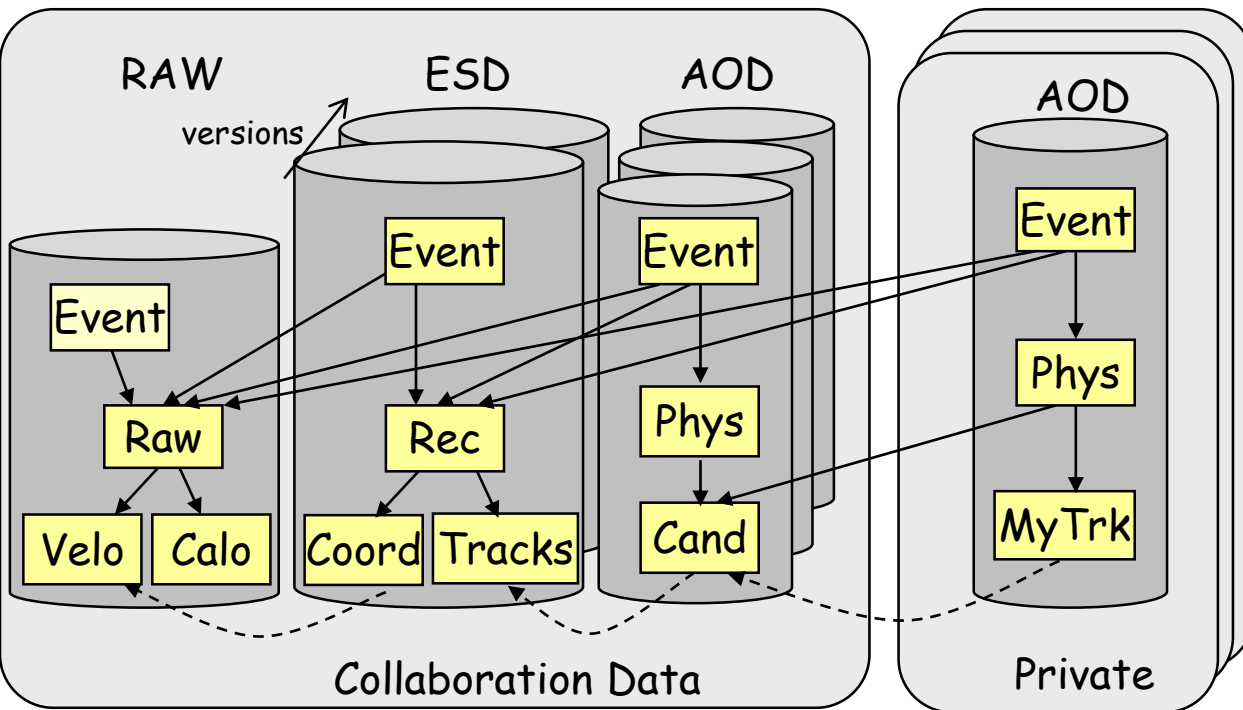
39

Estimated CPU Capacity at CERN



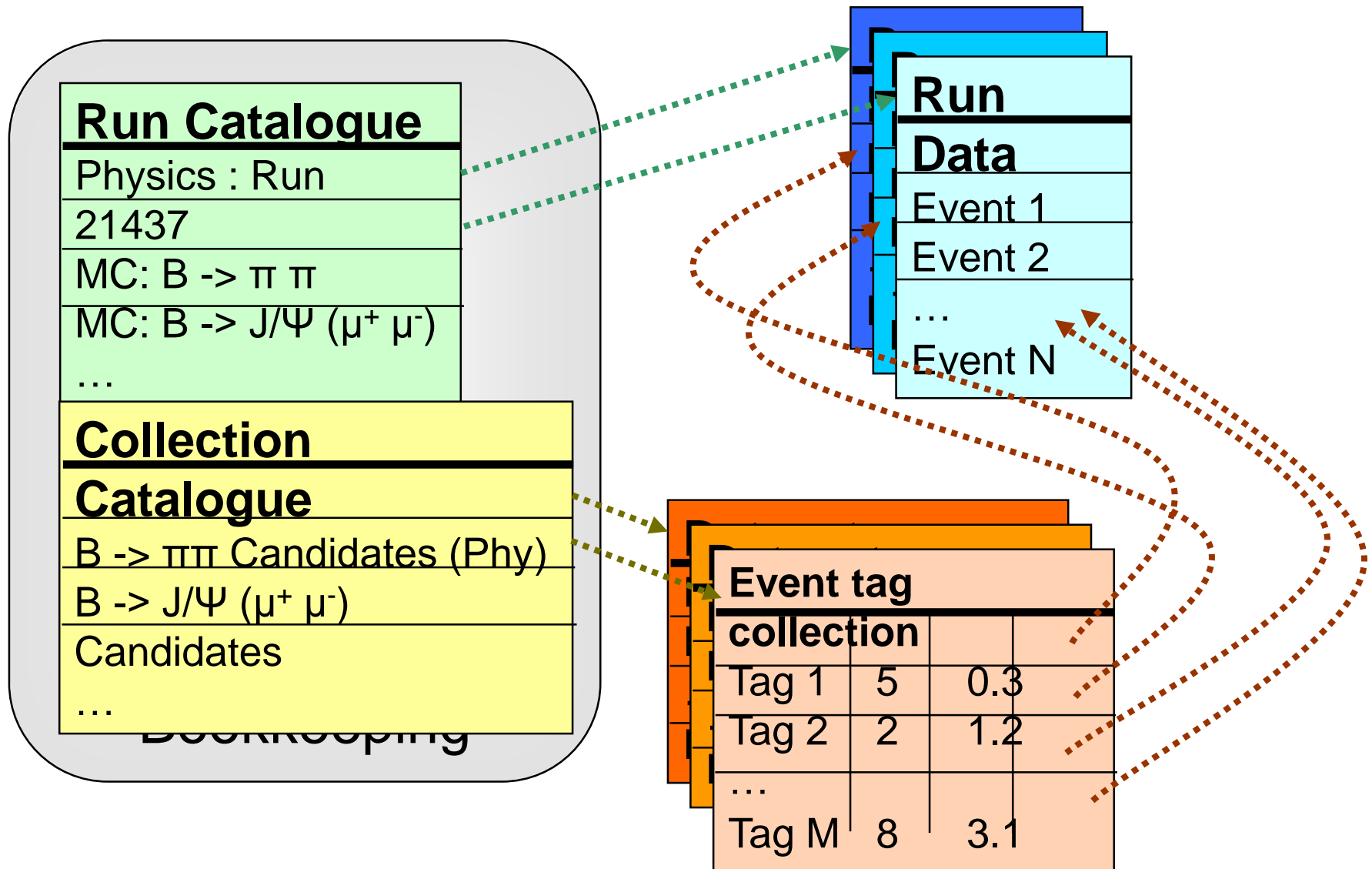
Moore's law

# Event Data



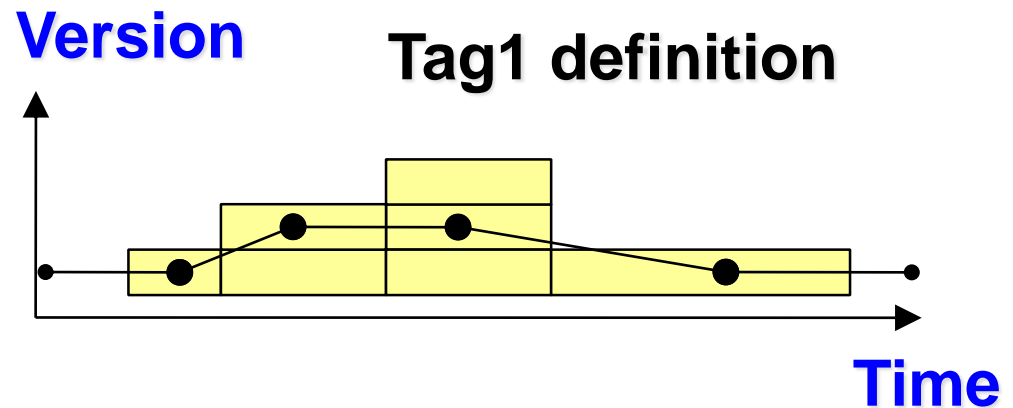
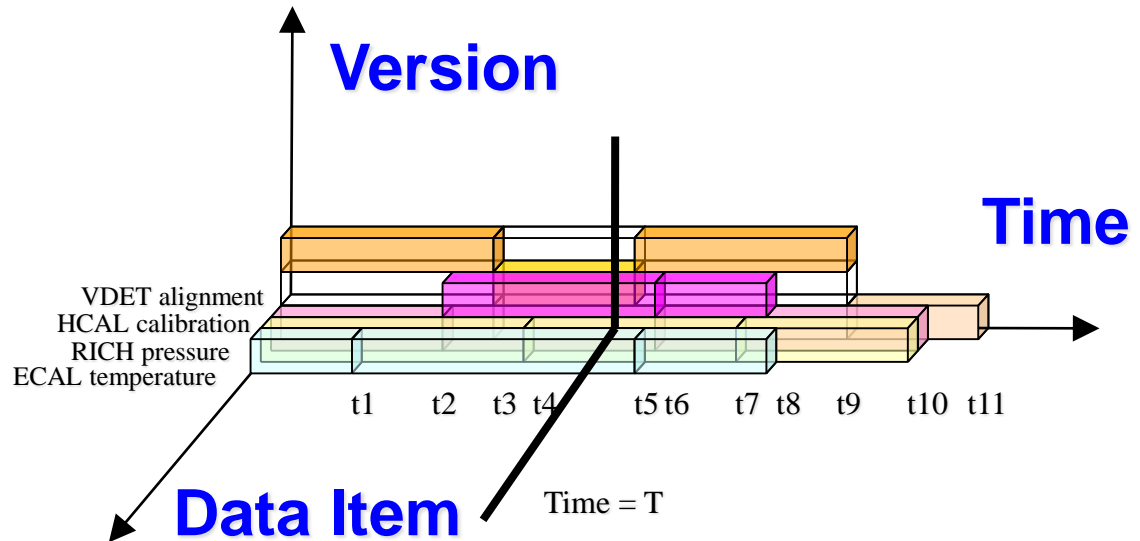
- ❑ Complex data models
  - ~500 structure types
- ❑ References to describe relationships between event objects
  - unidirectional
- ❑ Need to support transparent navigation
- ❑ Need ultimate resolution on selected events
  - need to run specialised algorithms
  - work interactively
- ❑ Not affordable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



# 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

# Actions during a Step

## z During each step

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

# Actions during a Step (cont)

## z During a step (continued)

y An (optional) user-written 'action' is called,

x Which can be used eg to create histograms or tallies.

y If the current volume contains a sensitive detector, that is addressed, allowing it eg

x to record the energy deposited,

x to record the exact position

in general to create a 'hit' that store all information that is relevant for that detector .



# Actions during a Step (cont)

## z During a step (continued)

y A parametrisation can be triggered (Geant4)

x Taking over from 'detailed' simulation

x 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.

