



Οι Υπολογιστές στη Φυσική Υψηλών Ενέργειών (ΦΥΕ)

Ιωάννης Αποστολάκης
CERN

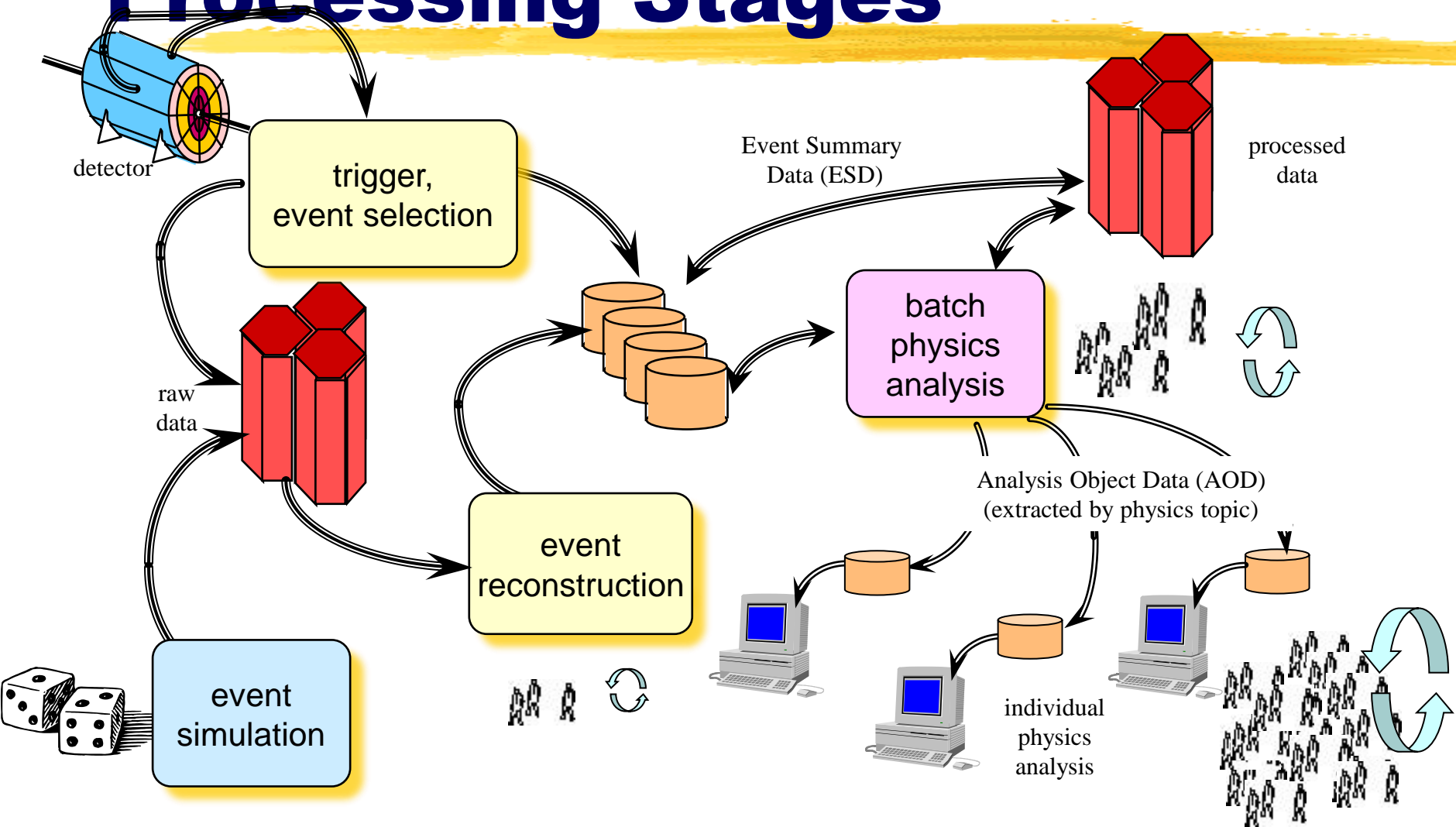
v0.98.8
2023.08.29

John.Apostolakis@cern.ch

Πλano της ομιλιας

- z Η χρηση των Υπολογιστών
- z Ανακατασκευή (reconstruction)
 - y Αμέσως (online) ή αργότερα (off-line)
- z Προσομοίωση (simulation)
- z Ανάλυση δεδομένων (data analysis)
- z Το πλέγμα (GRID) – και μεγέθη
 - y Υπολογ. ανάγκες, άλλες εφαρμογες

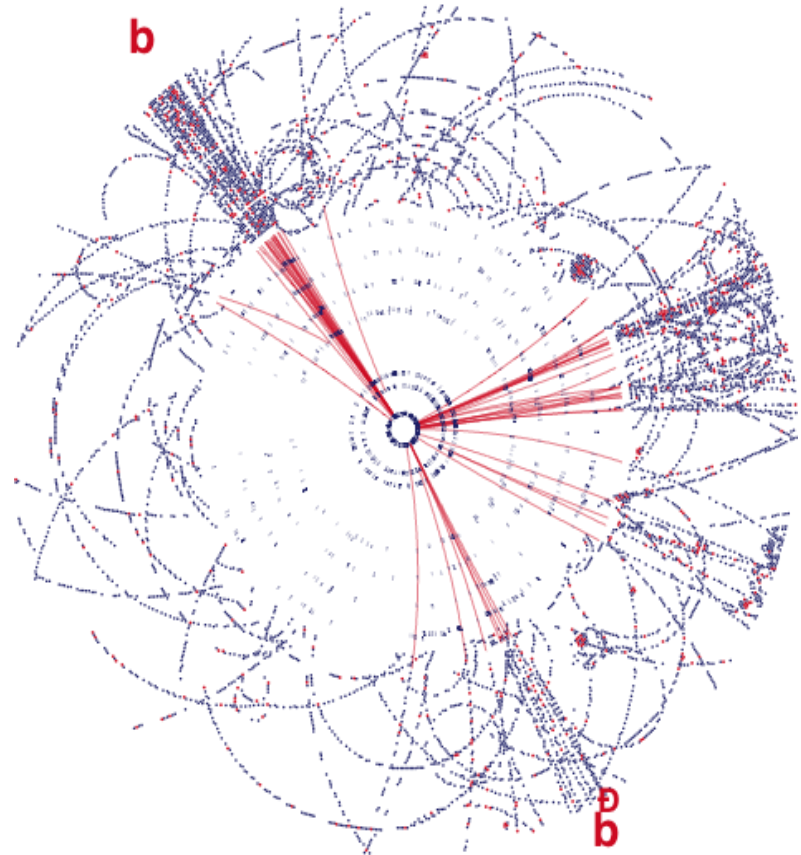
Processing Stages



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
 $H \rightarrow b\bar{b}$



Χρήσεις Υπολογιστών στη ΦΥΕ



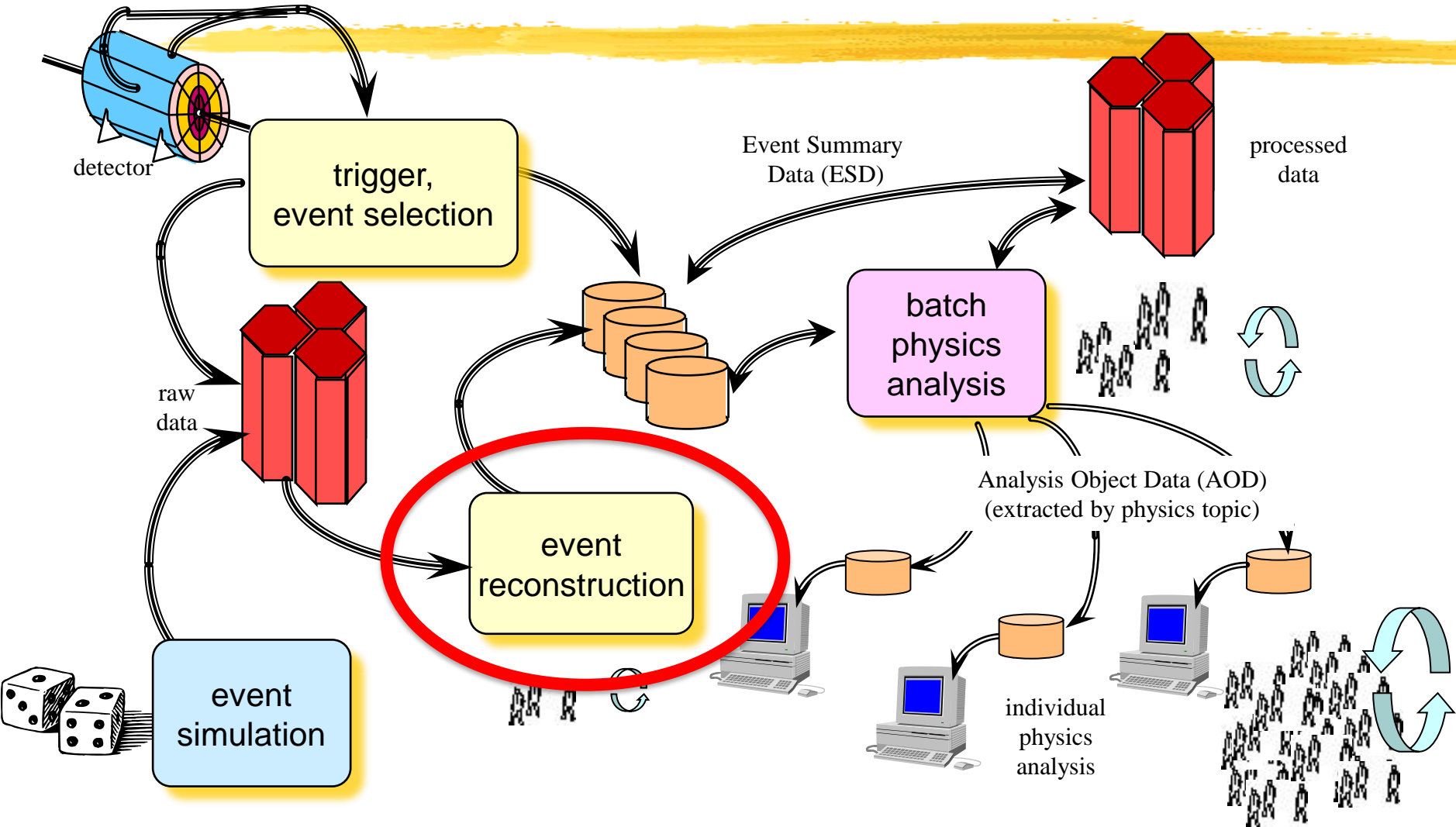
- z Καταγραφή δεδομένων
 - y Data Acquisition (DAQ)
 - y Περιλαμβάνει την επιλογή συγκρούσεων
- z Ανακατασκευή (reconstruction)
 - x Αμέσως (online) ή αργότερα (off-line)
- z Προσομοίωση (simulation)
- z Ανάλυση δεδομένων (data analysis)

Ανακατασκευή



Μια γρήγορη εισαγωγή

Στάδια Επεξεργασίας – Ανακατασκευή (Reconstruction)



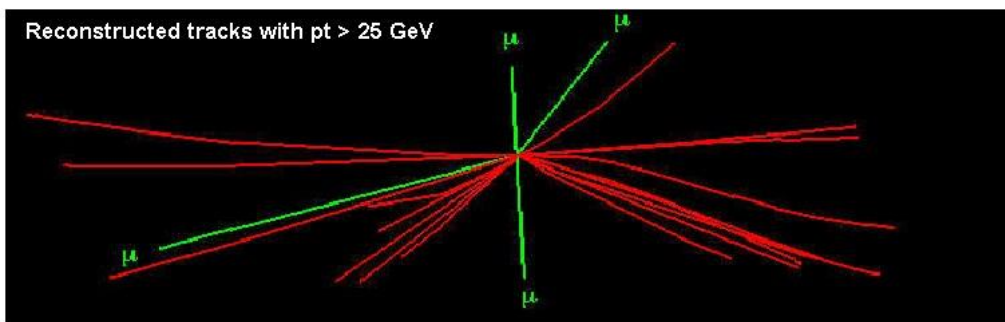
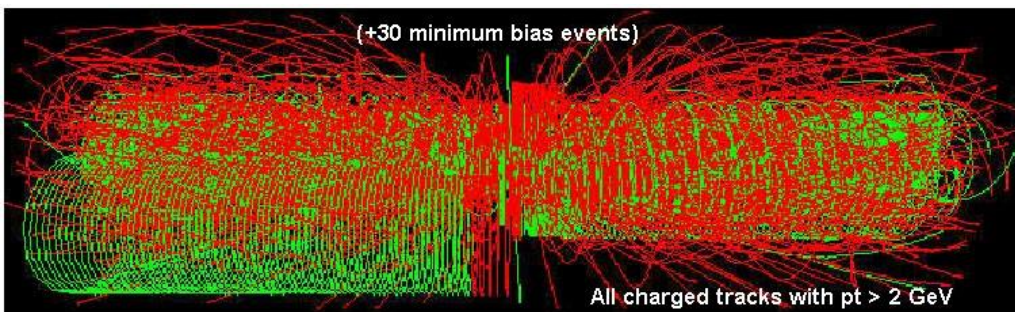
Η δυσκολία της Ανακατασκευής

Starting from
this event

Μ'αυτη την
αρχική εικόνα
(αριστερό ημισυ)

Looking for
this “signature”

Ψάχνουμε αυτή
την περιληπτική
εικόνα



→ **Selectivity: 1 in 10^{13}**

(Like looking for a needle in 20 million haystacks)

Οι σημερινοί ανιχνευτές

z Πολλα τμήματα

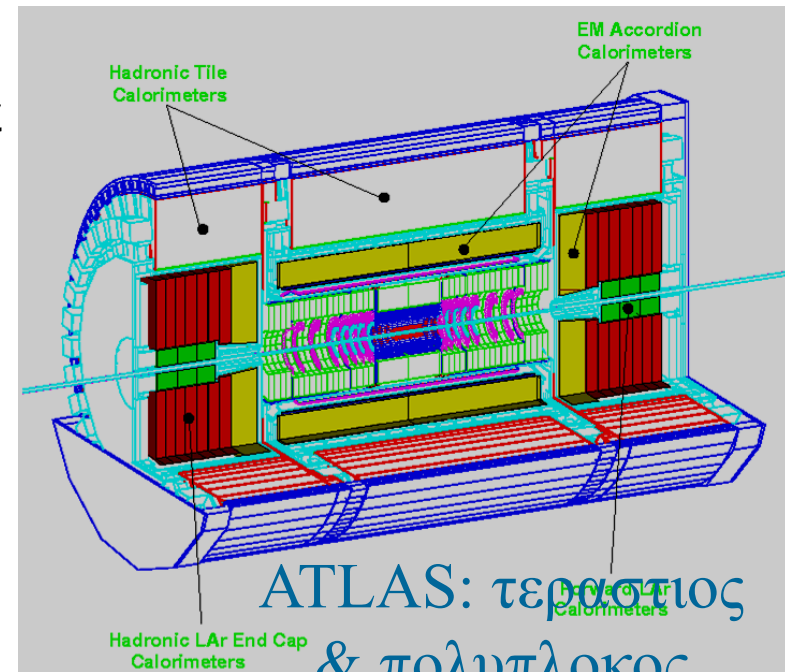
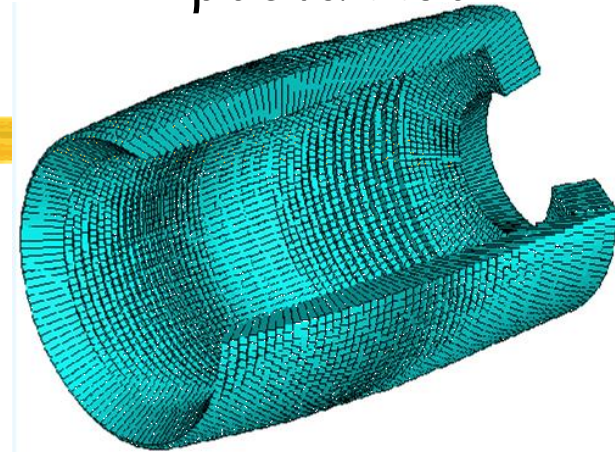
γ Διαφορετικές ανάγκες

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

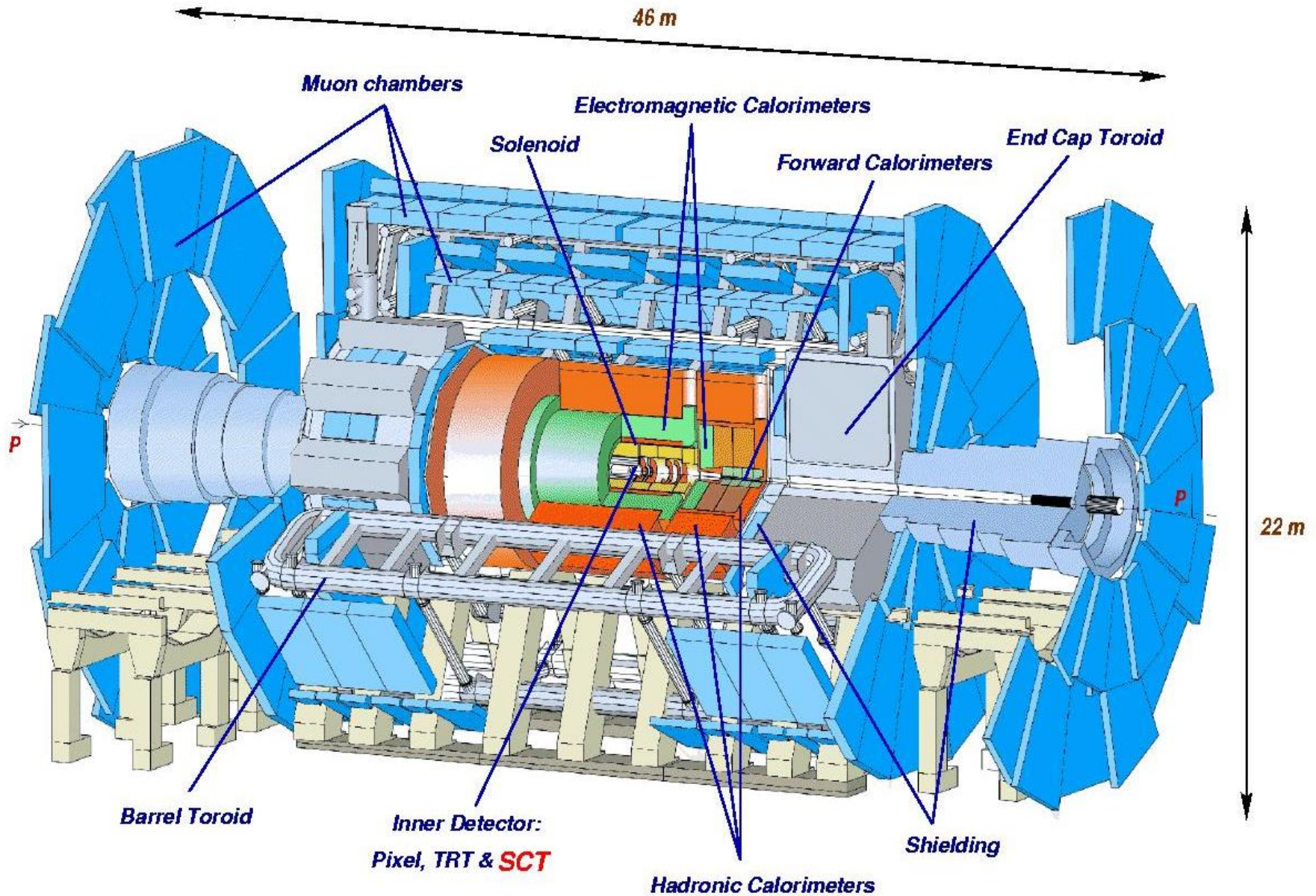
z Λογω της πολυ-πλοκοτητας

γ Όλες οι μελετες χρειαζονται υπολογιστικα εργαλεια

Καλοριμετρο
κρυσταλλου

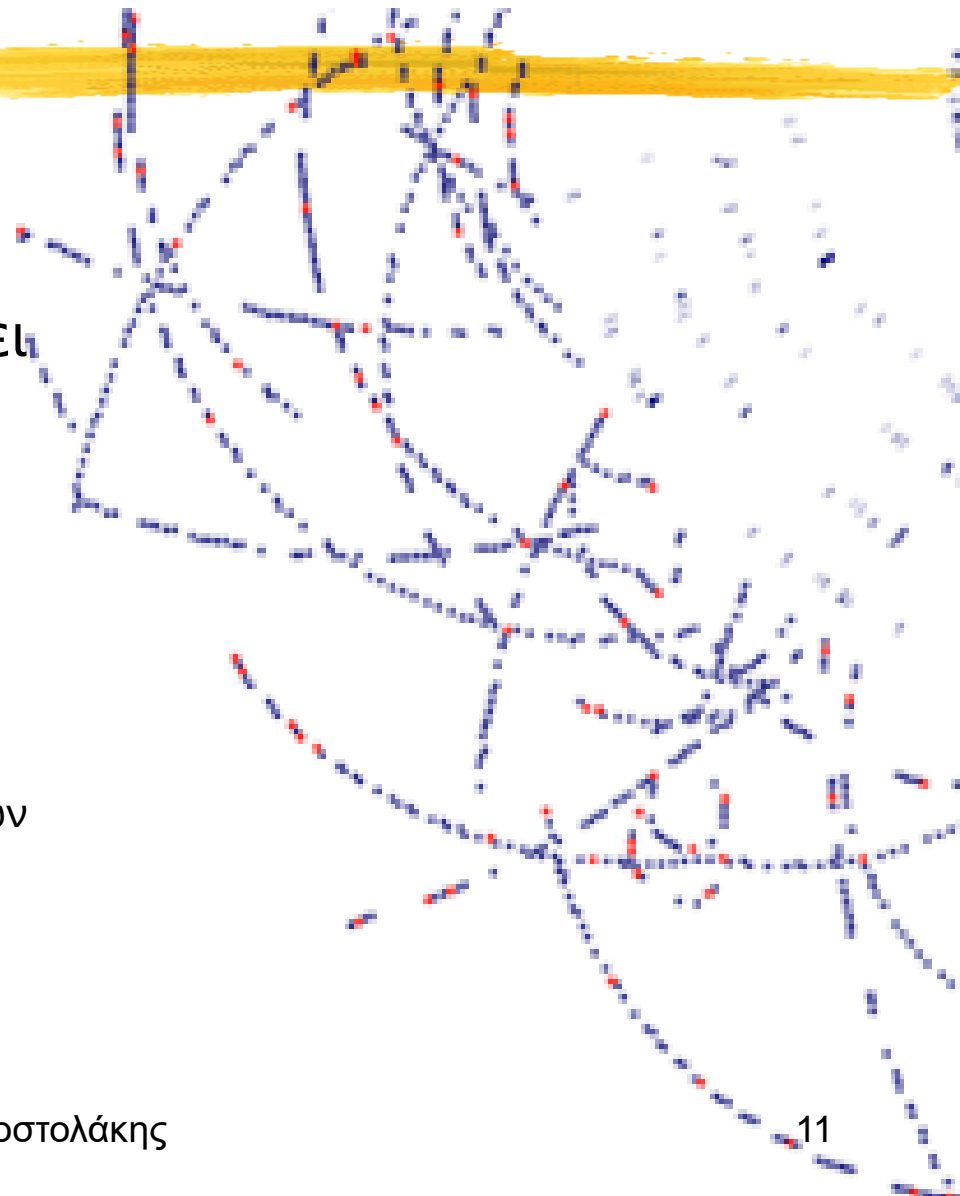


Οι ανιχνευτές του ΑΤΛΑΣ

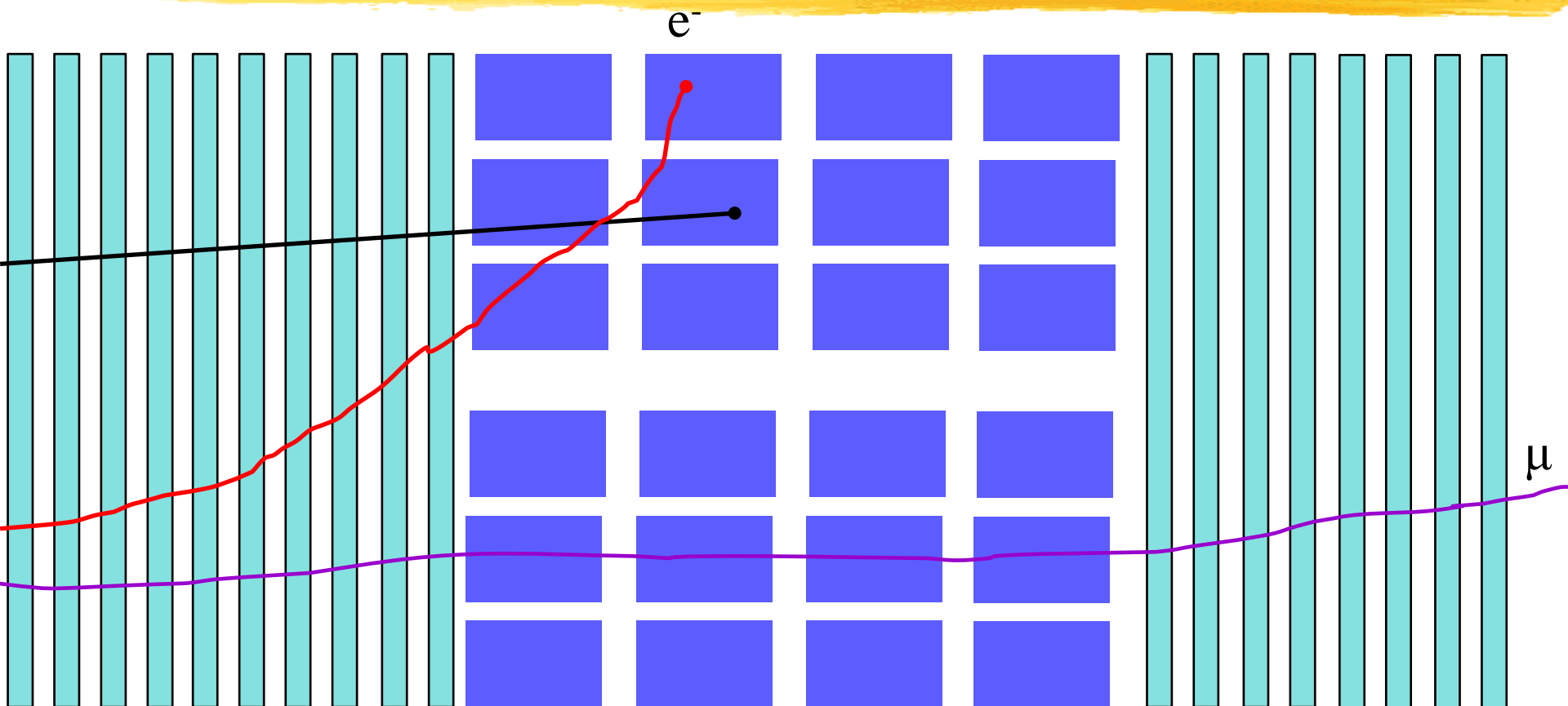


Τι είναι η ανακατασκευή?

- z Οι μετρήσεις είναι σαν ένας γρίφος
 - γ Τι τροχιές τις προκάλεσαν?
- z Κάθε μέτρηση θέσης βοηθάει
 - γ Υπάρχουν όμως 100-άδες ως χιλιάδες μετρήσεις
- z Η ανακατασκευή πρέπει να βρεί τη λύση!
 - γ Ξέροντας καλά το μαγνητικό πεδίο
 - x Βρίσκουμε ποιές μετρήσεις ανήκουν σε ποιές τροχιές

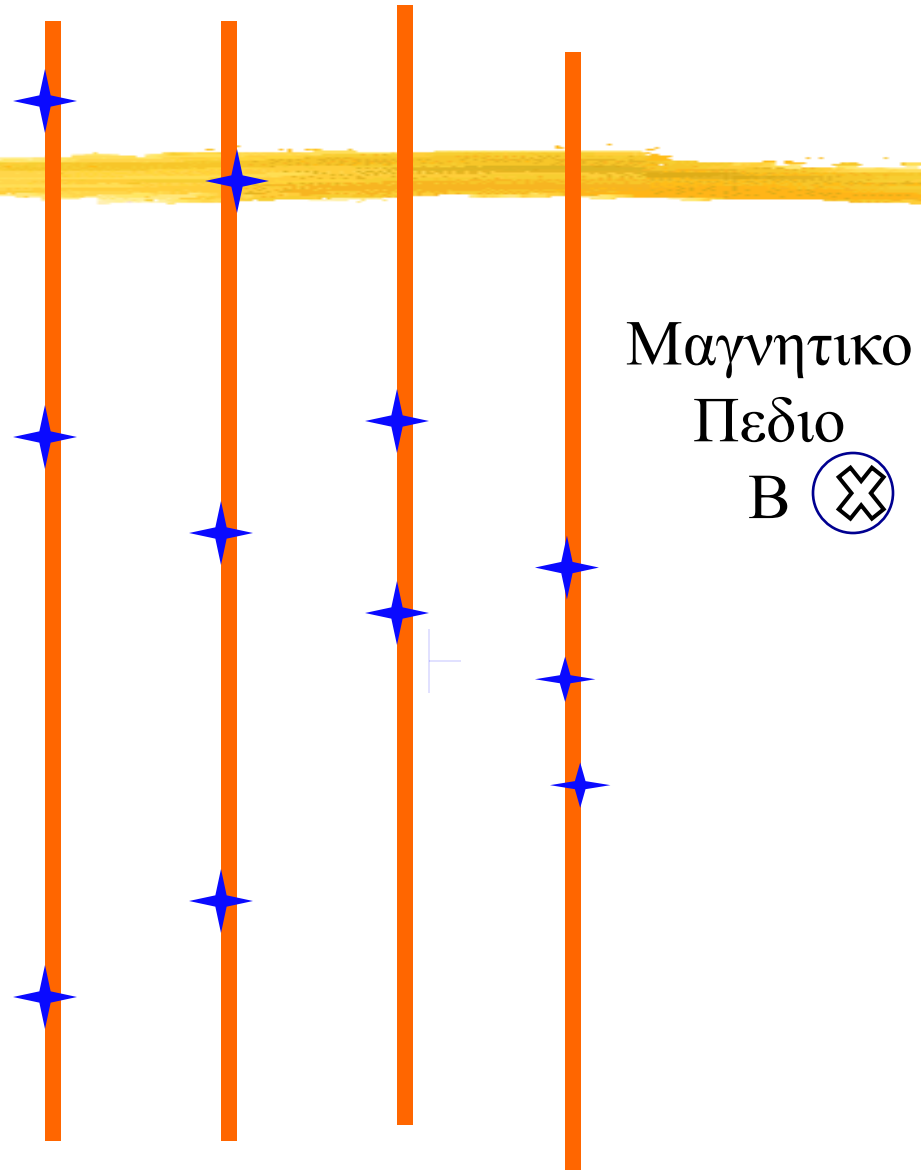


Απλουστευμένο πείραμα ΦΥΕ



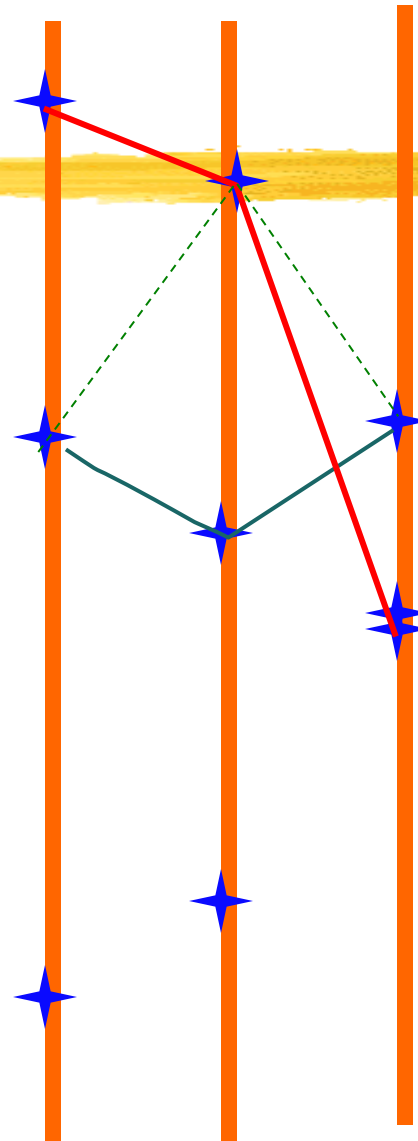
Ανακατασκευή στην πράξη

- Αρχίζει με τις θέσεις διάβασης 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
-



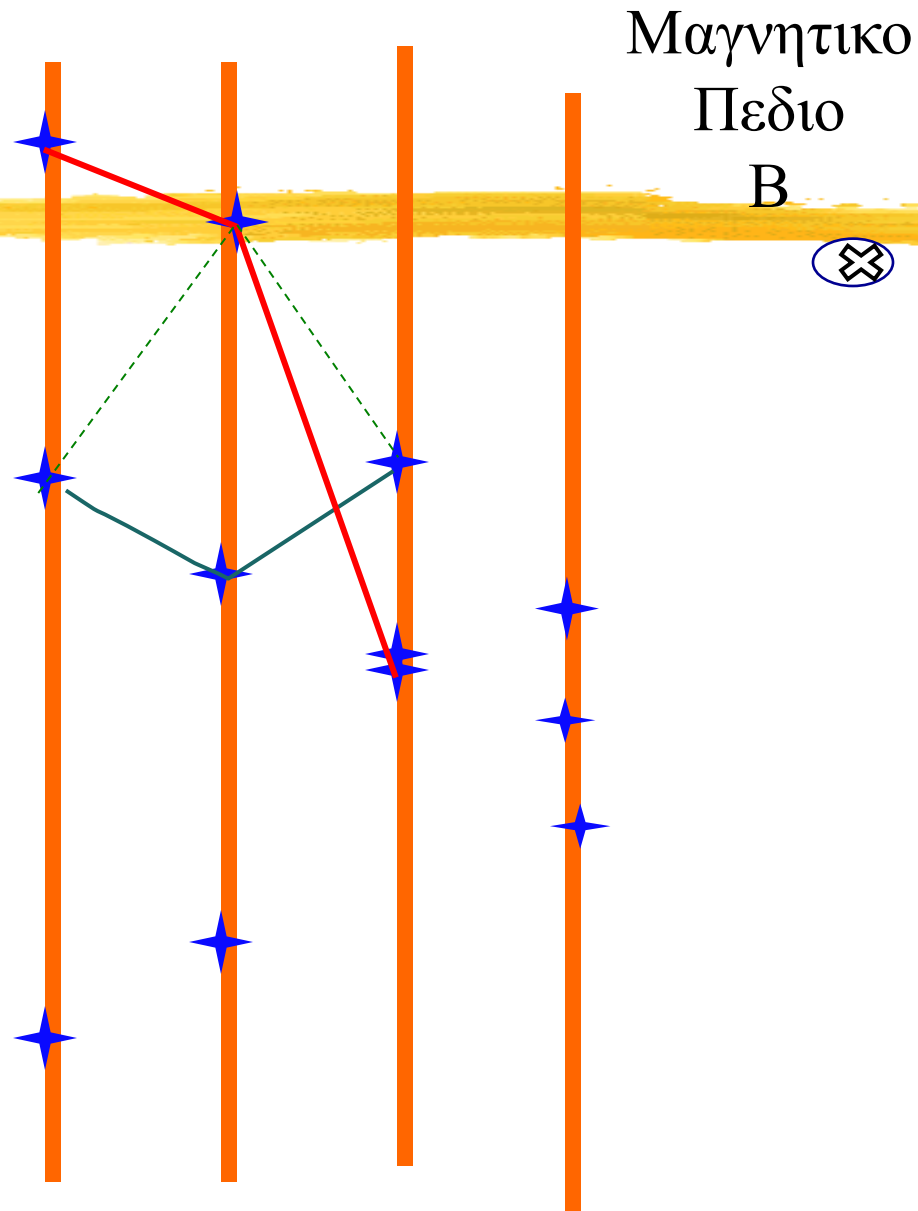
Μαγνητικό
Πεδίο
B



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

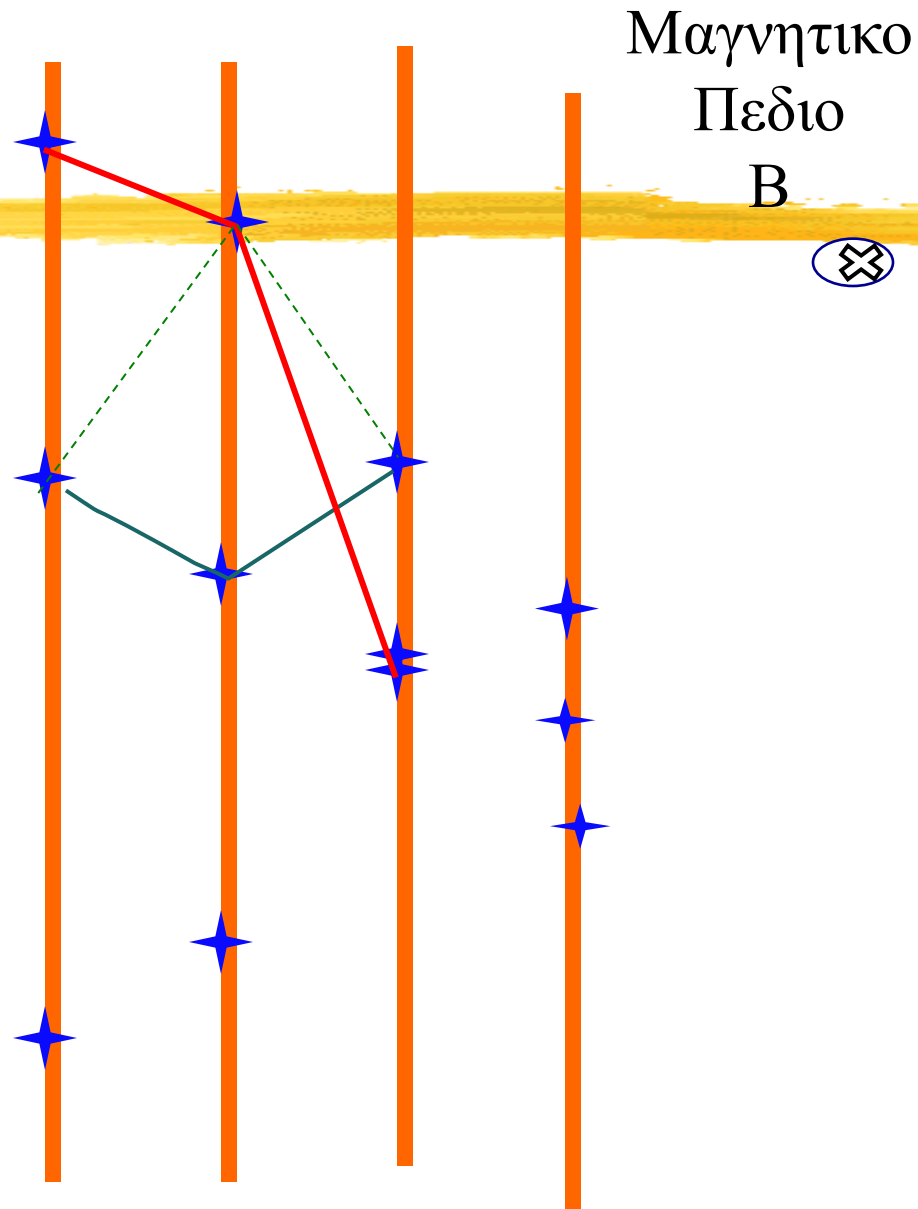
Ανακατασκευή στην πράξη

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



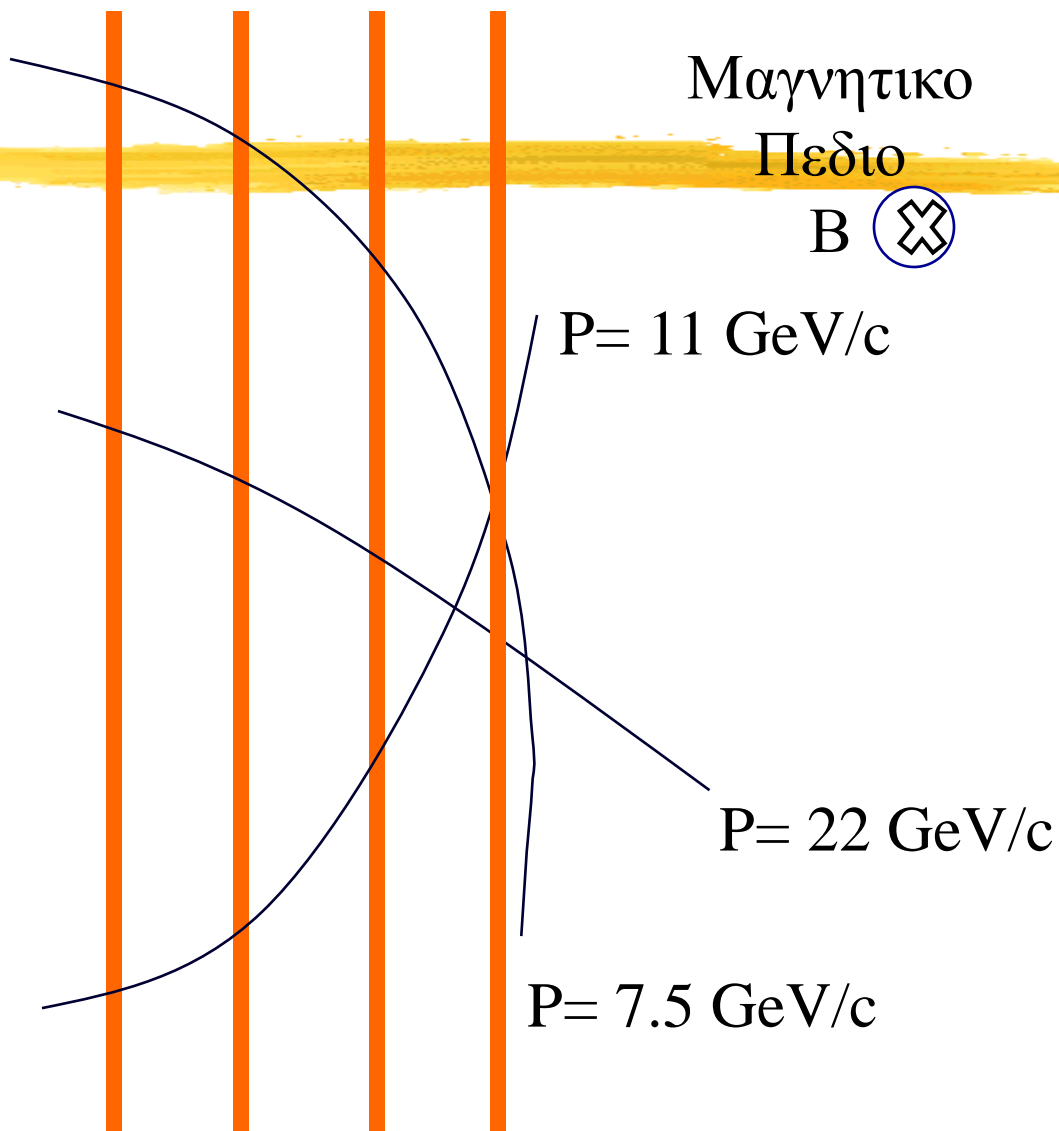
Ανακατασκευή στην πράξη

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



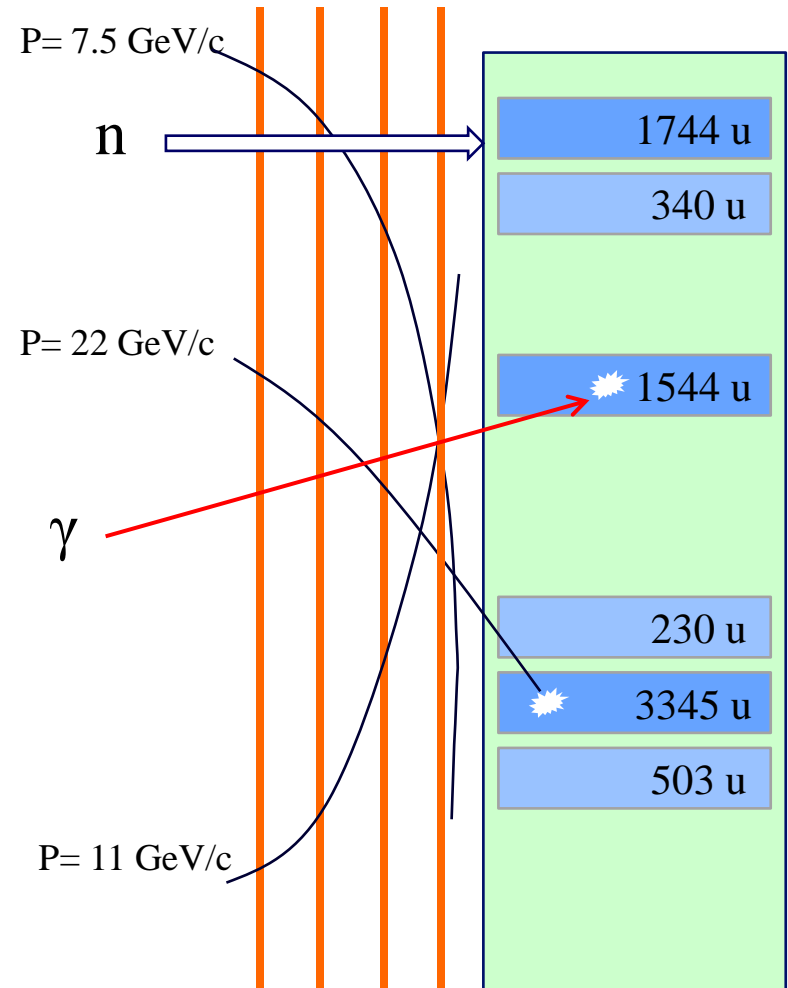
Ανακατασκευή: πρώτο αποτέλεσμα

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

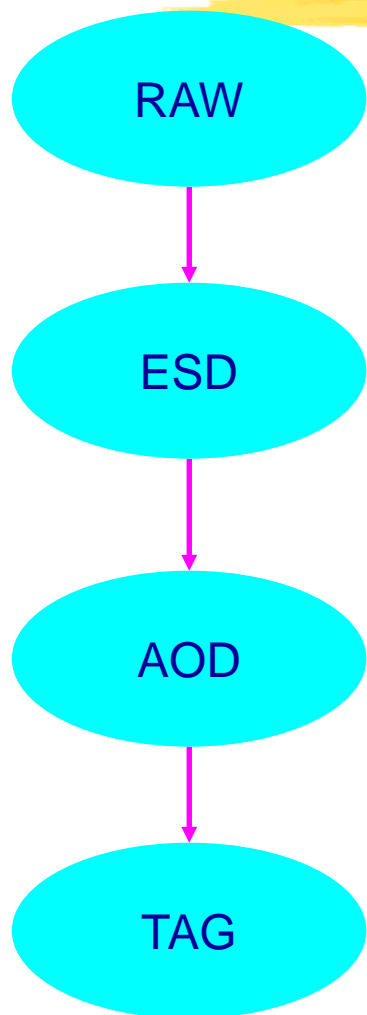


Από μερική προς ‘ολική’ ανακατασκευή

- z Μετά από αυτά τα πρώτα ίχνη
- z Προσθέτουμε ανιχνευτές
 - y Ευαίσθητους, σε περισσότερα σωματίδια, άλλες αλληλεπιδράσεις
- z Βγάζουμε πληροφορίες για τα άλλα σωματίδια (γ , π , K , ρ , n , ...)
- z Συγκρτόουμε «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

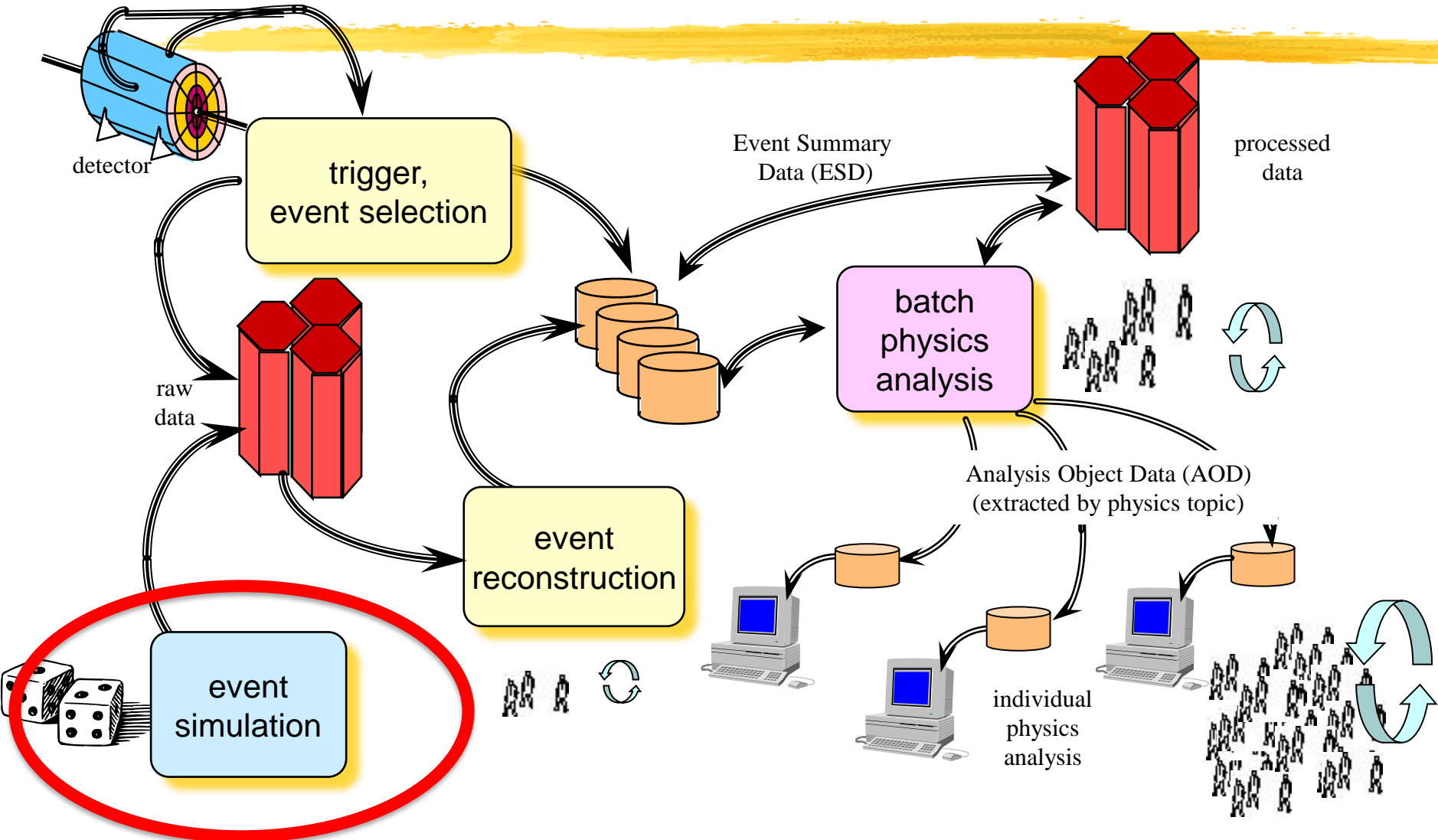
AOD

Πληροφορίες ταξινόμησης

Πληροφορίες σχετικές με τη γρήγορη επιλογή των κρουσεων

TAG

Στάδια Επεξεργασίας- (Προσομοίωση) Simulation



Προσομοίωση και Ανιχνευτές



Τι είναι η προσομοίωση
Γιατί υπάρχει
Πως γίνεται

Τι είναι η προσομοίωση?

Φυσικό σύστημα

Μοντέλο = εξισώσεις

Εξελίξη συστήματος

Βγαζω αποτελέσματα



ΓΙΑΤΙ ΚΑΝΟΥΜΕ ΠΡΟΣΟΜΟΙΩΣΗ?

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

Τι είναι προσομοίση;

z Φτιαχνουμε μοντελα

γ Του ανιχνευτη

x Γεωμετρια

x Υλικά

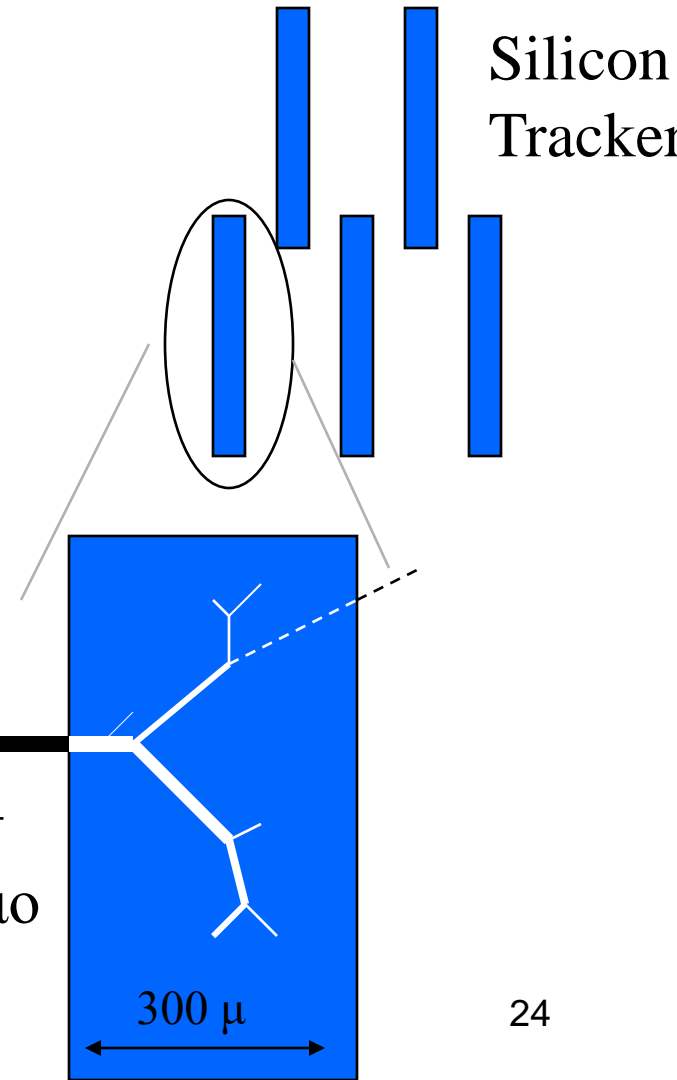
γ Των αλληλεπιδρασεων

x Καθε γνωστου τυπου

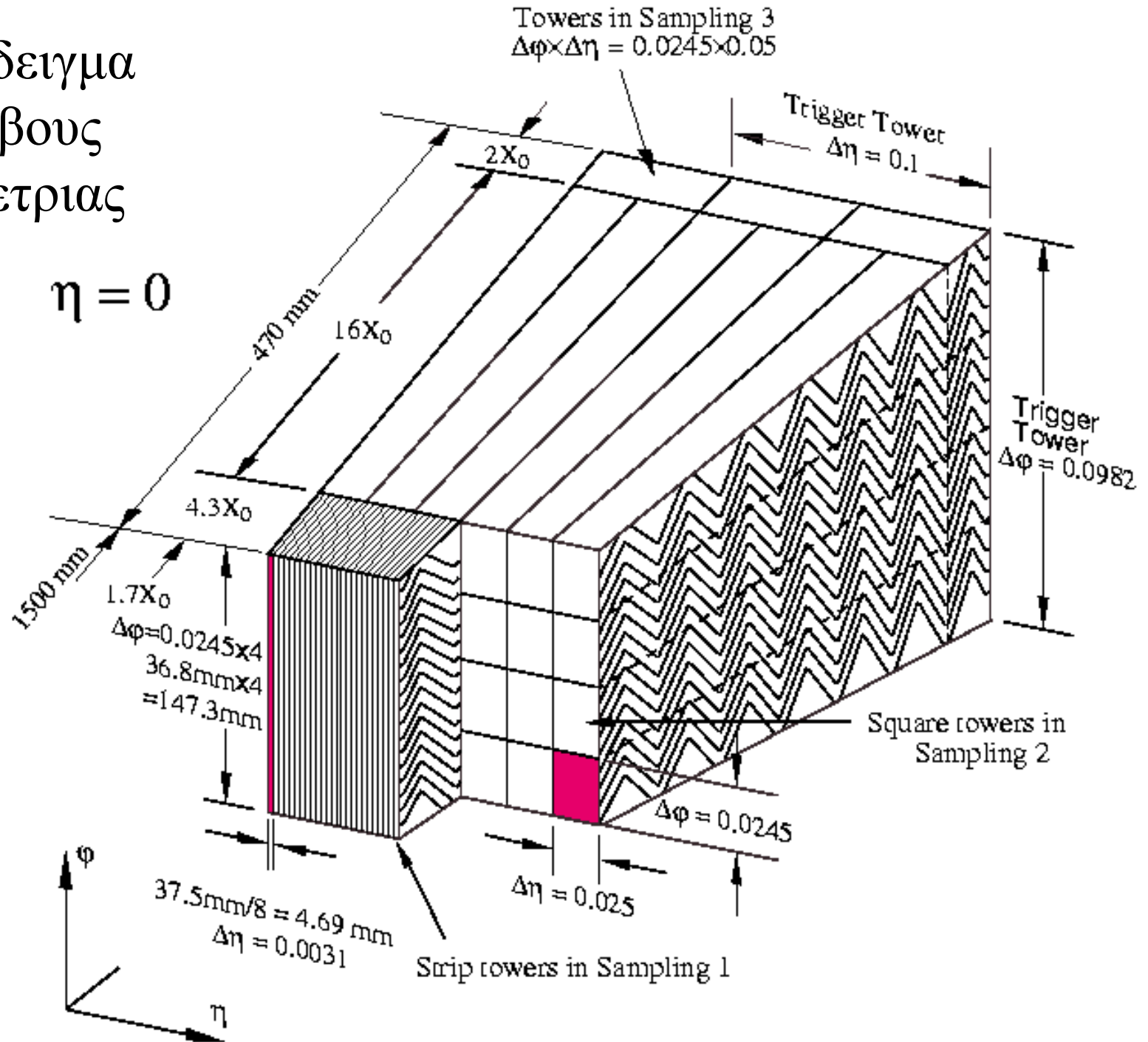
- Ηλεκτρομαγνητικου
- Υψηλου πυρηνικου

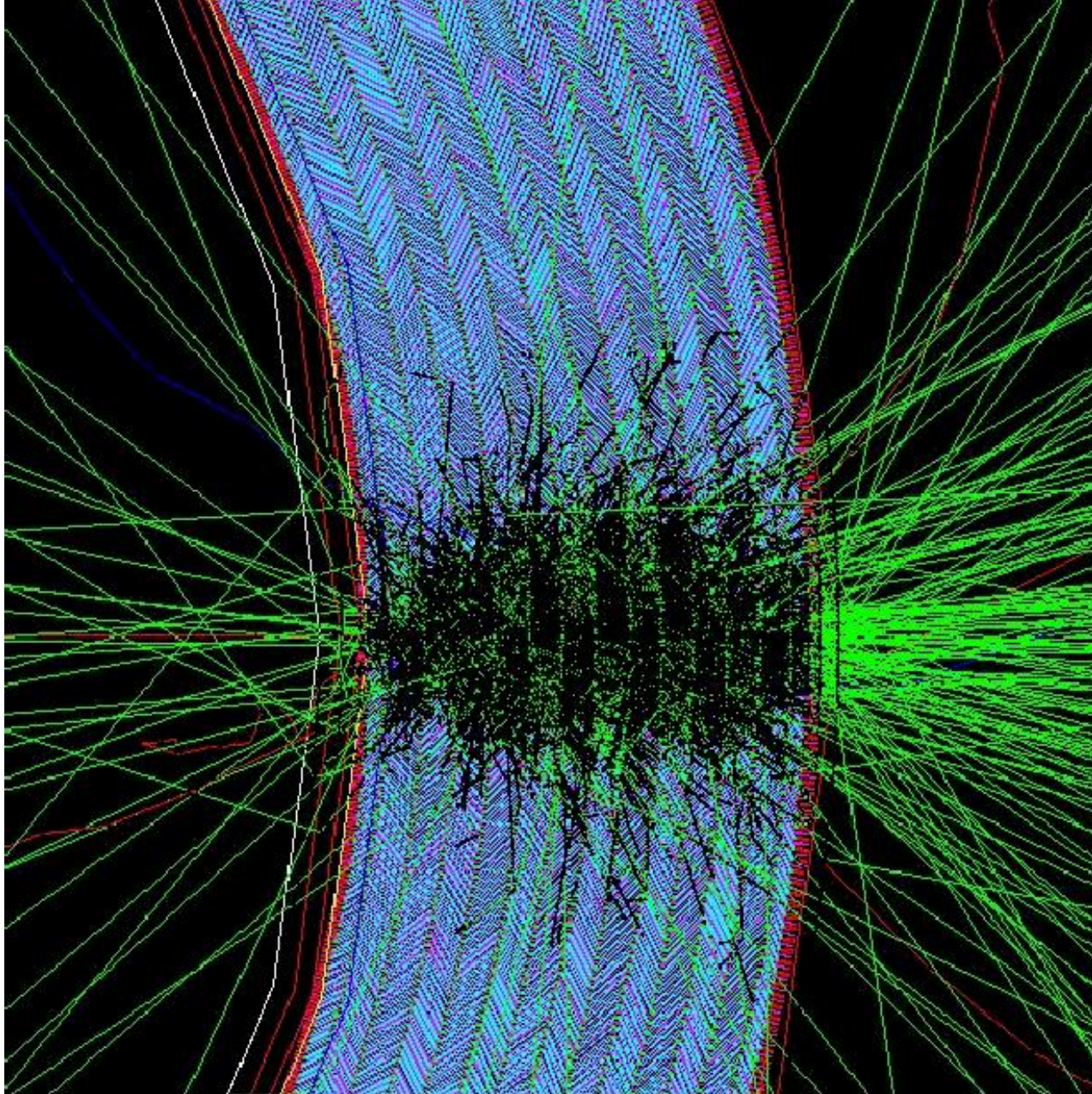
$$\sigma_{\text{συνολο}} = \sum \sigma_{\text{φαινομενου}}$$

2.5 MeV e⁻
ηλεκτρονιο

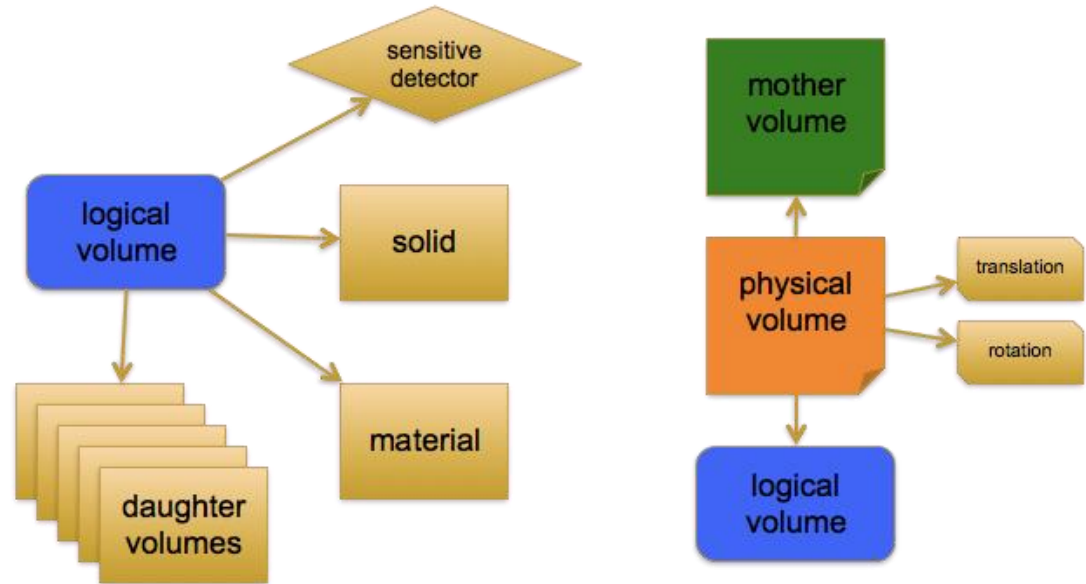
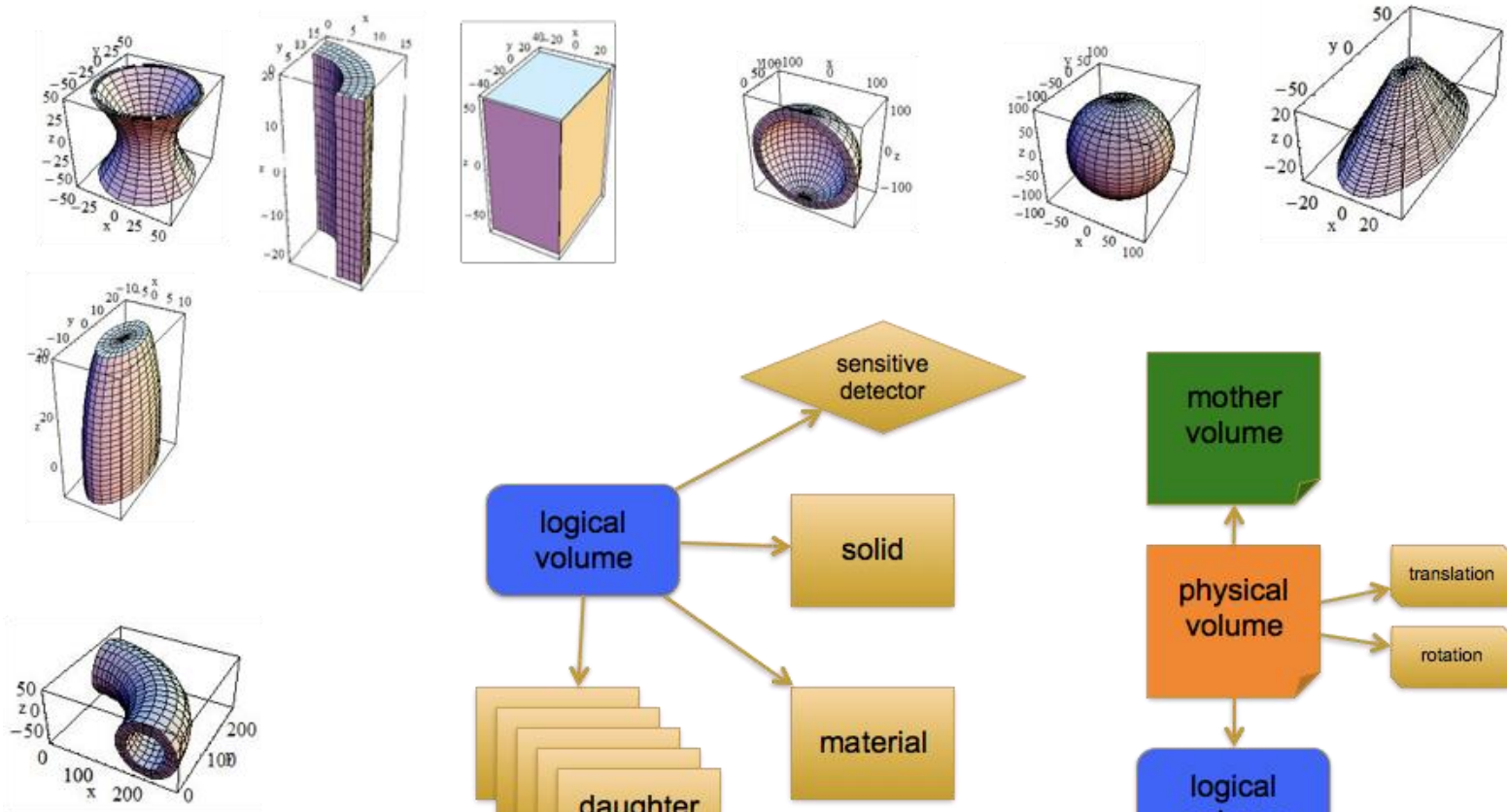


Παραδειγμα ακριβους γεωμετριας





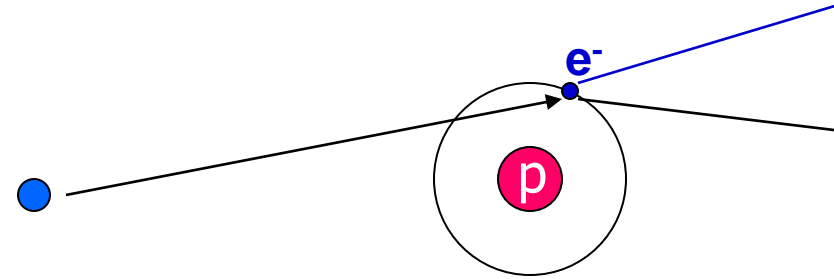
Geometry construction



Βασικές Αλληλεπιδρασεις

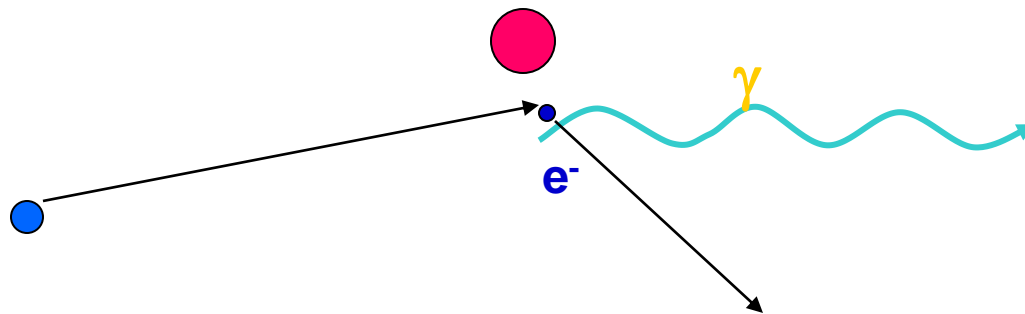
- z Οι διαφορες αλληλεπίδρασης του σωματιδίου με το υλικό (τμήμα του ανιχνευτή η άλλο)
- γ παραγωγή δευτερευοντος σωματιδιου

x Ιονισμός



x Ακτινοβολία πέδησης

x (Bremsstrahlung)



Ενεργος διατομη (Cross-section)

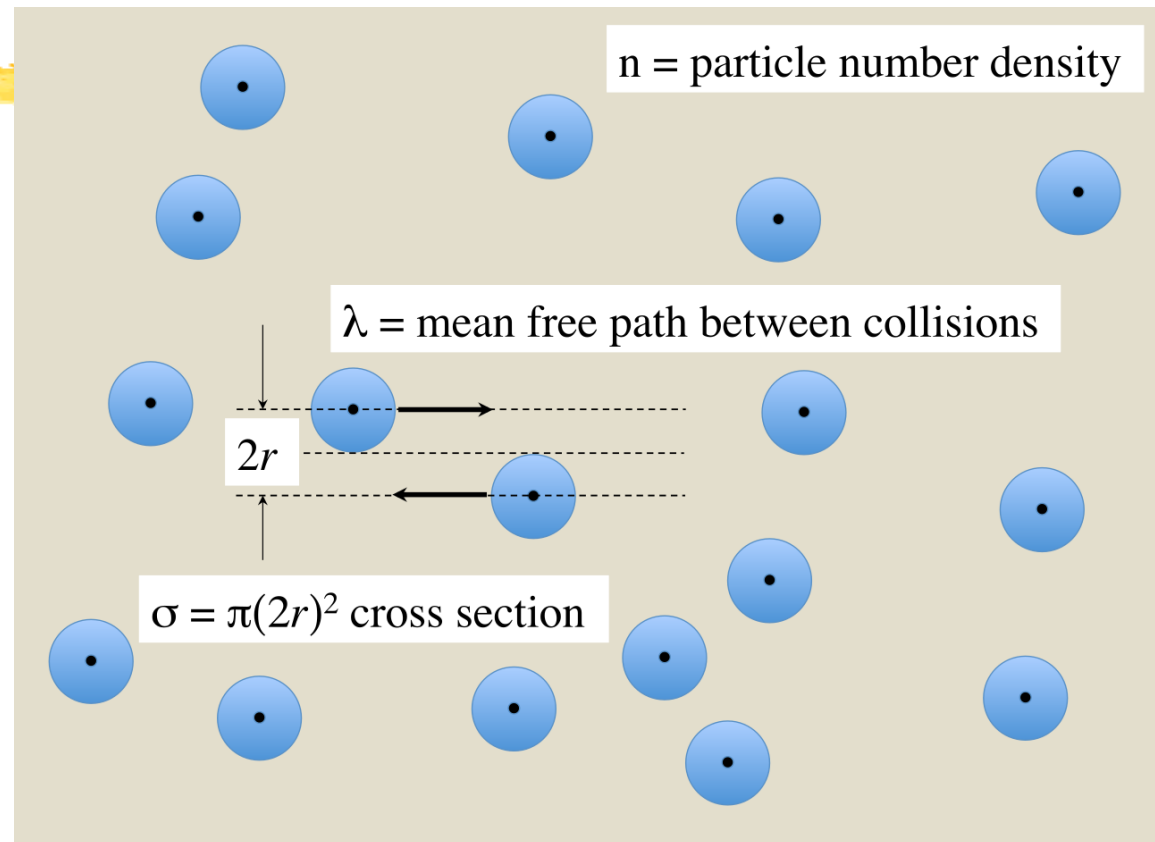
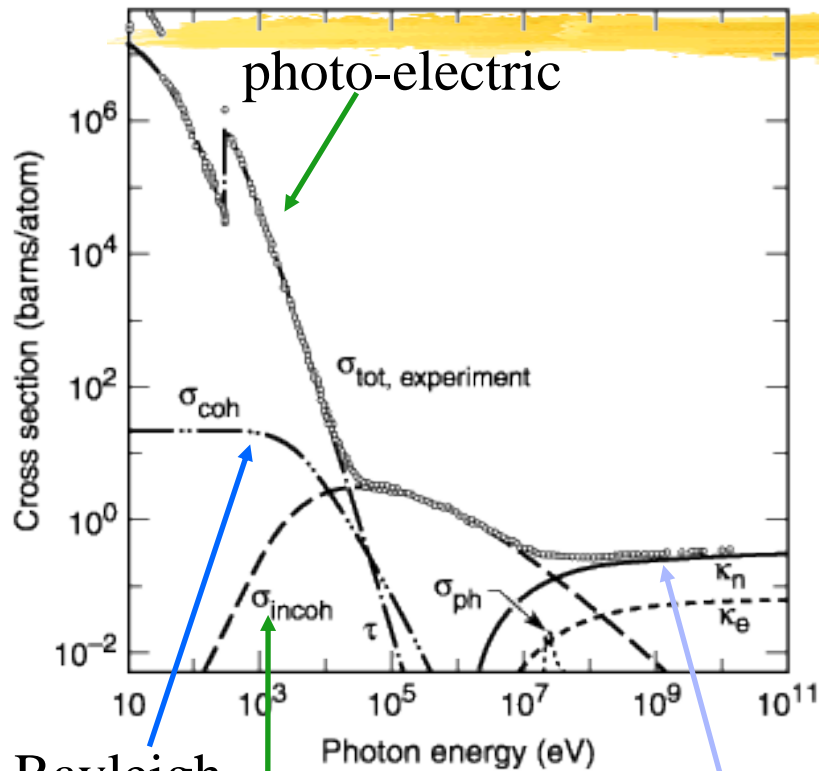


Figure 1. In a gas of particles of individual diameter $2r$, the cross section σ , for collisions is related to the particle number density n , and mean free path between collisions λ .

Ενεργος διατομη (Cross-section)

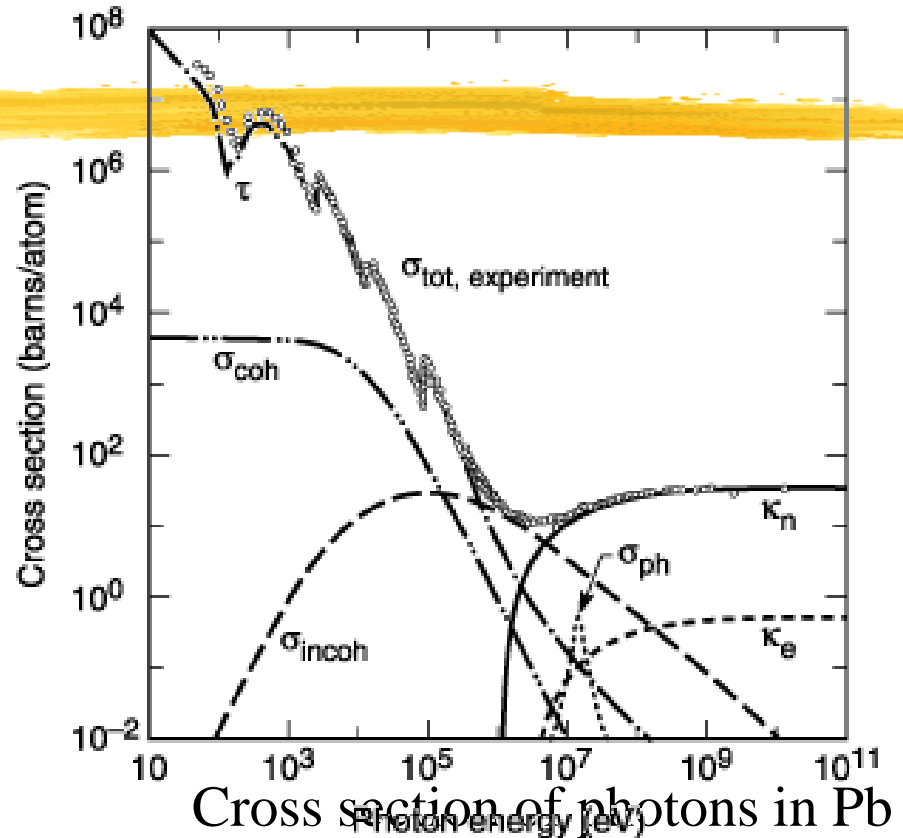


Rayleigh

Compton

gamma conversion

<https://pdg.lbl.gov/>



Cross section of photons in Pb
(Lead)

https://xdb.lbl.gov/Section3/Image_Sec3/Sec3135.gif

Σκεδάσεις ηλεκτρονίων

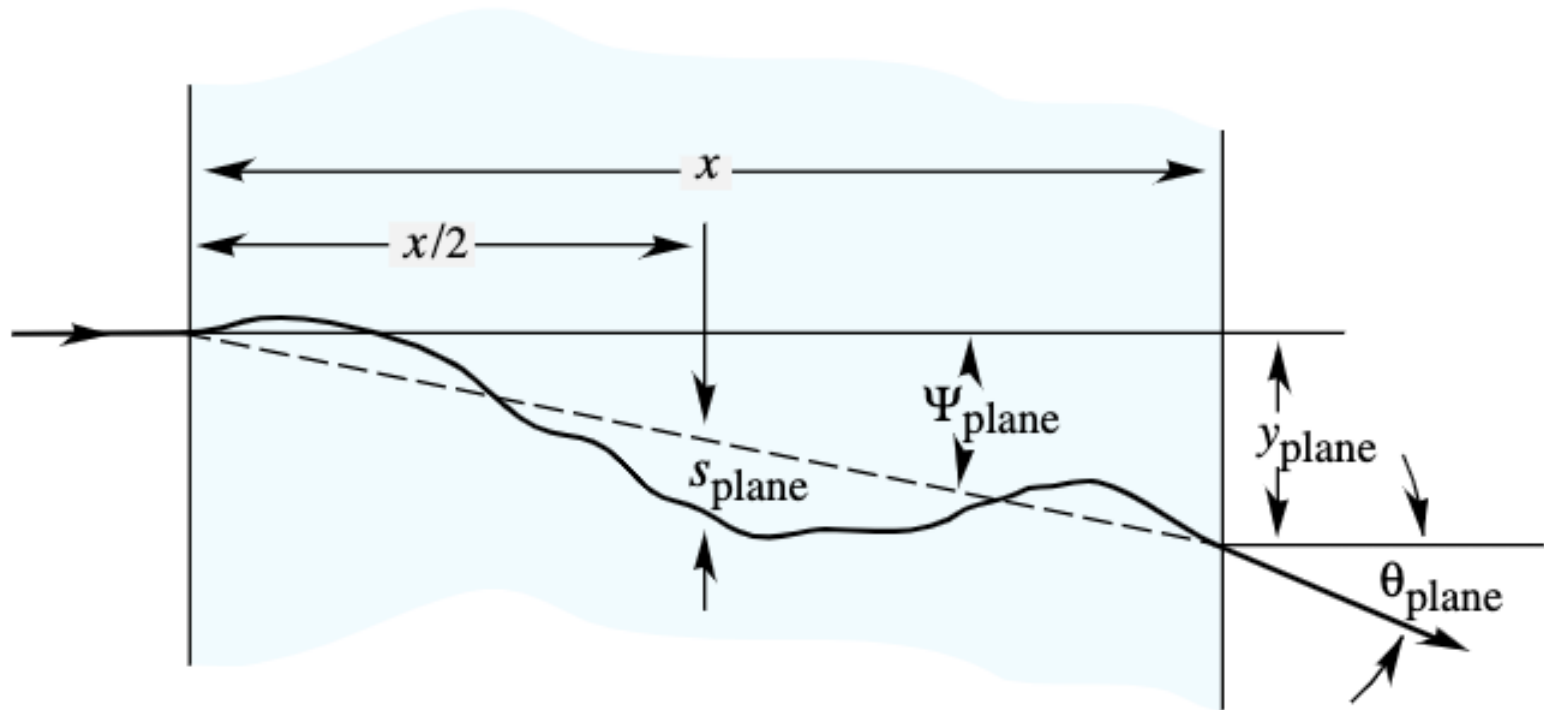
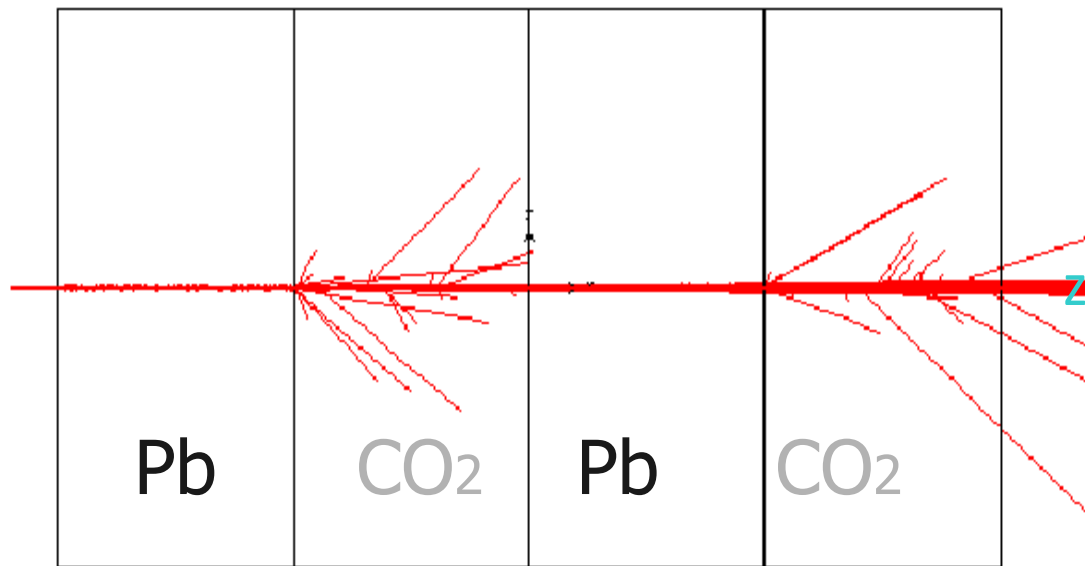


Figure 34.10: Quantities used to describe multiple Coulomb scattering. The particle is incident in the plane of the figure.

From “Passage of particle through matter” - <https://pdg.lbl.gov/>

Ένα απλο παραδειγμα



GEANT 3

z Στο μολυβδο παραγονται πολλα δευτερευοντα σωματιδια

- γ Τα περισσοτερα μενου κοντα,
- γ Μερικα ξεφευγουν.

Το διοξειδιο του ανθρακα, σαν αεριο, εχει μικρη πυκνοτητα

- γ Οσα σωματιδια φτανουν η παραγωνται, πανε μακρια
- γ Παραγονται λιγοτερα

Γιατι προσομοίωση ?



- z Για να σχεδιάσουμε τους ανιχνευτες
- z Για να ετοιμάσουμε τις μεθόδους ανακατασκευής
- z Για να καταλάβουμε τα αποτελέσματα του ανιχνευτή

Άλλες χρήσεις προσομοίωσης

z Ιατρική

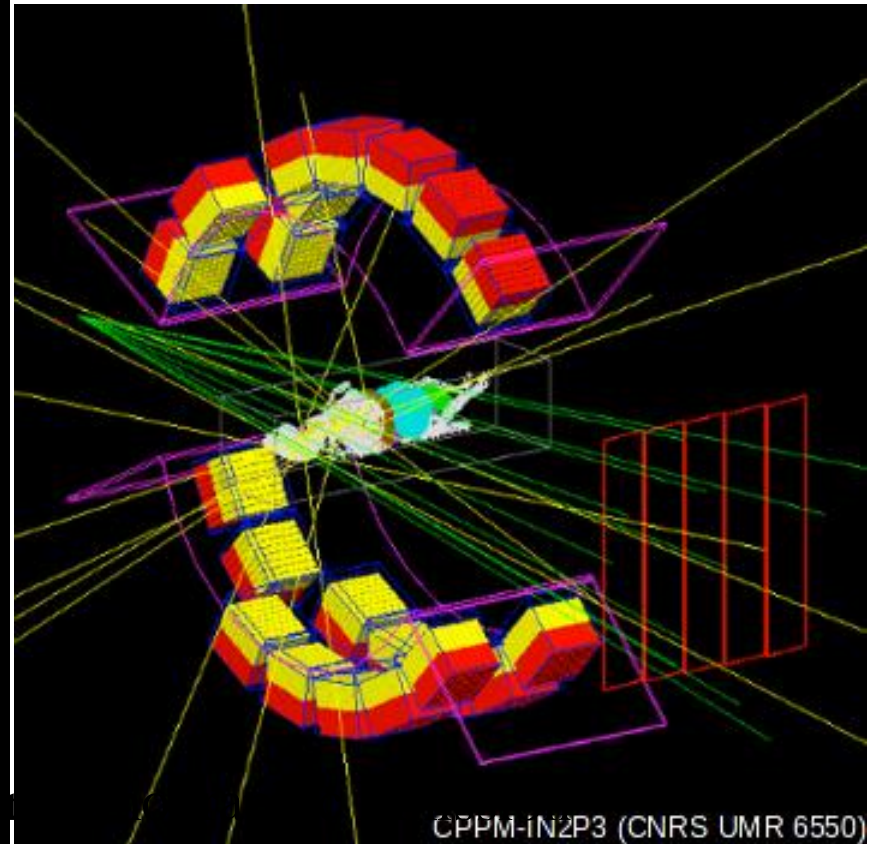
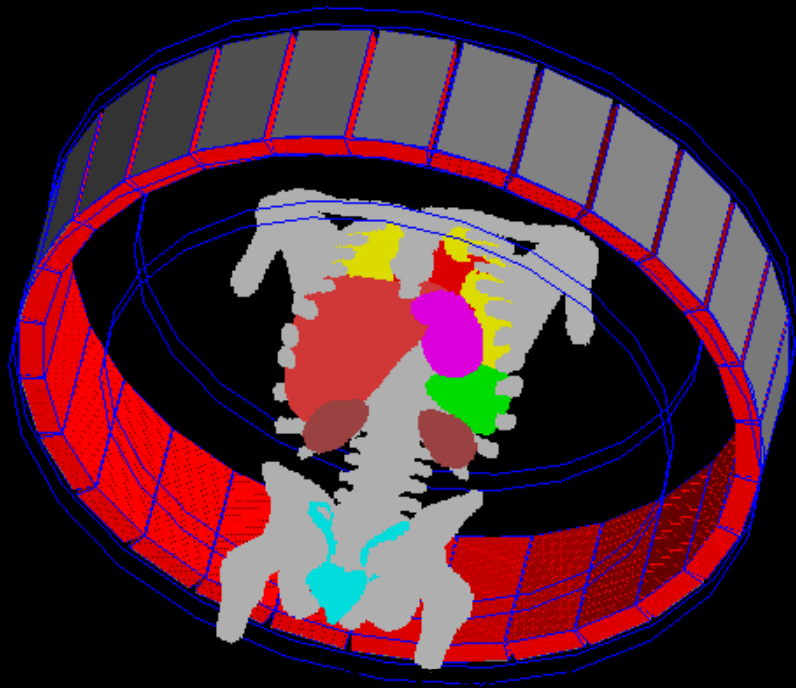
- γ Για ιατρική διάγνωση

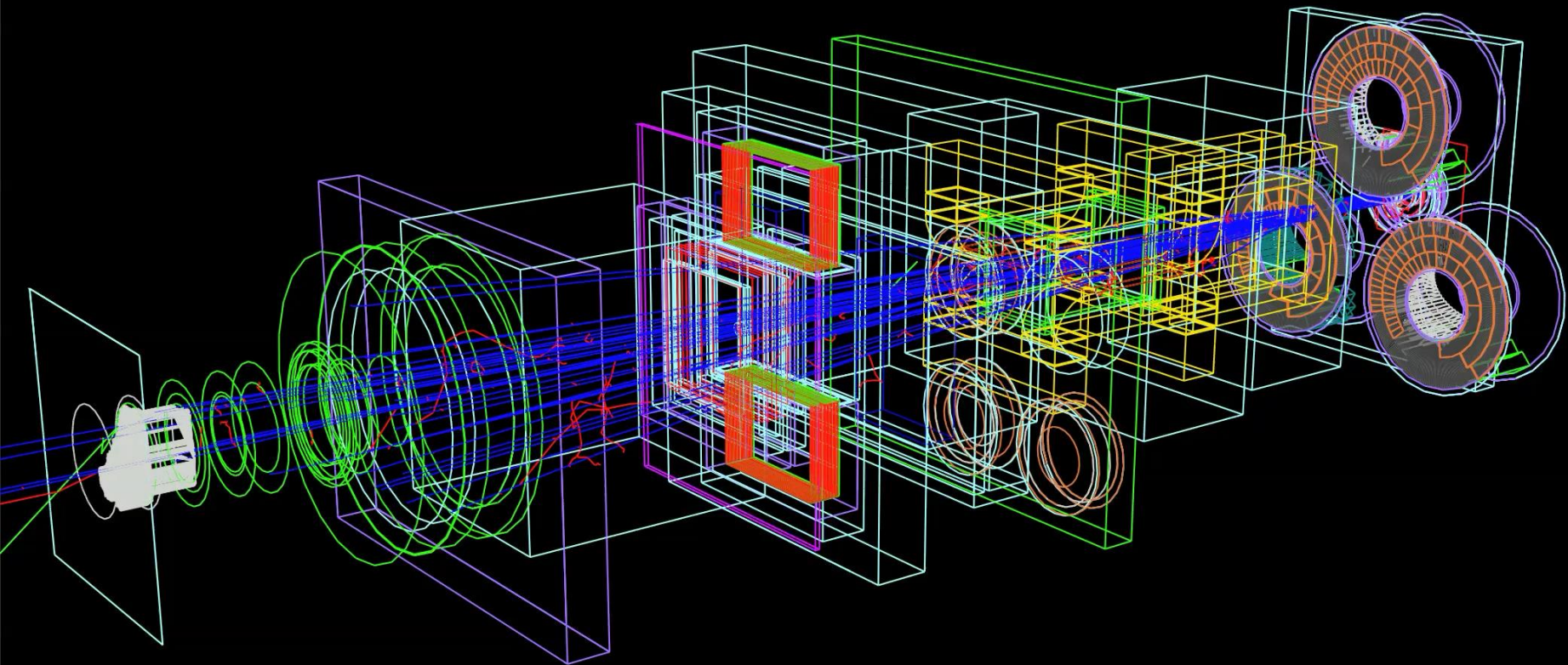
- γ Ραδιο-ιατρική – θεραπεία με πρωτόνια

z Πλανήτη

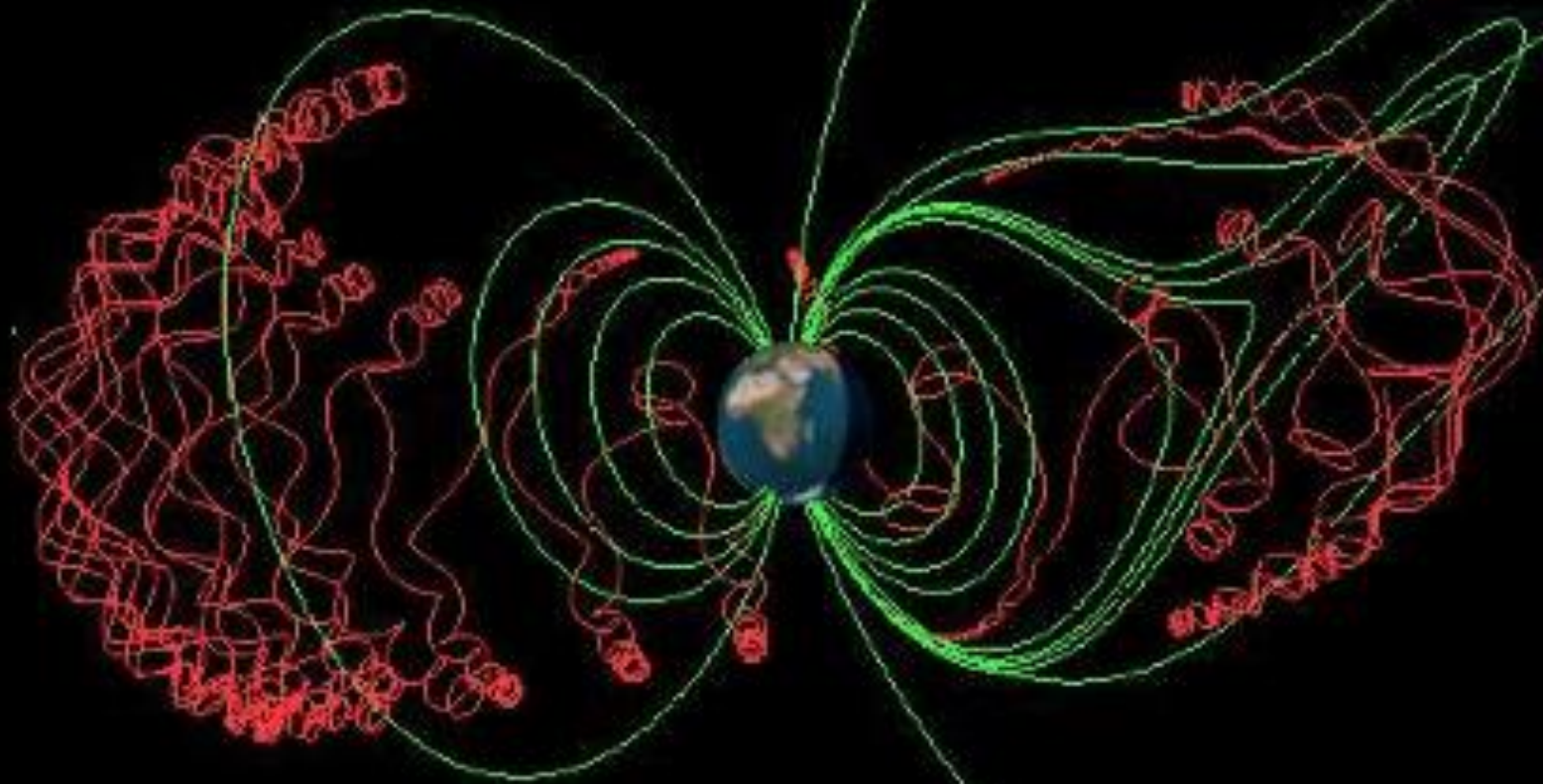
- γ Αλληλεπίδραση ηλεκτρονίων/πρωτονίων με ατμόσφαιρα και μαγνητικό πεδίο της Γης

Positron Emission Tomography (PET)

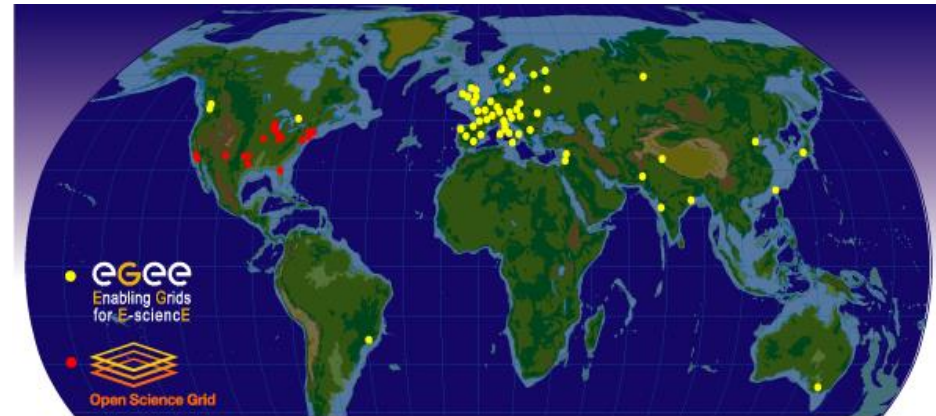




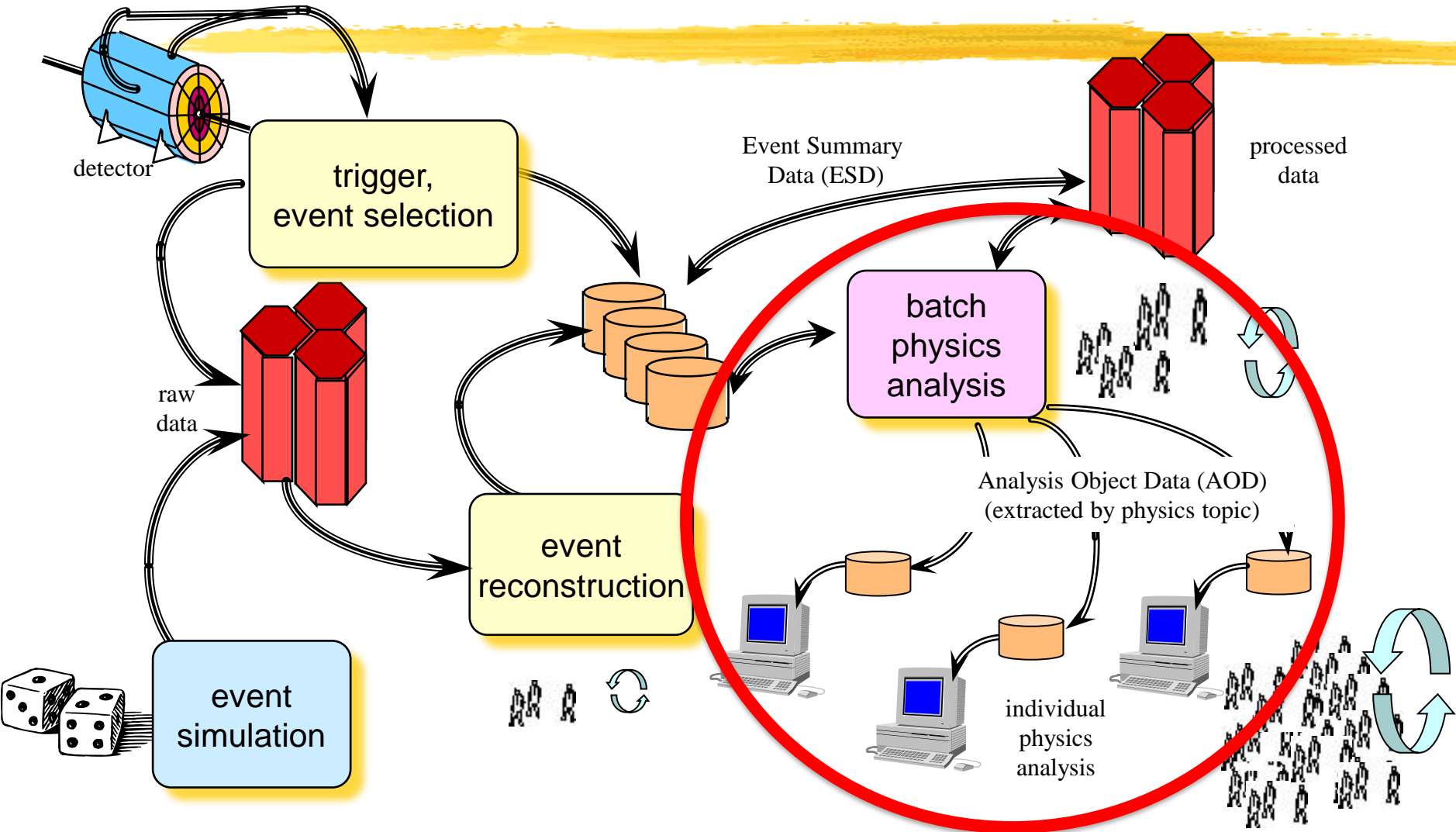
Trapped particles in Earth's Magnetic field



Ανάλυση δεδομένων



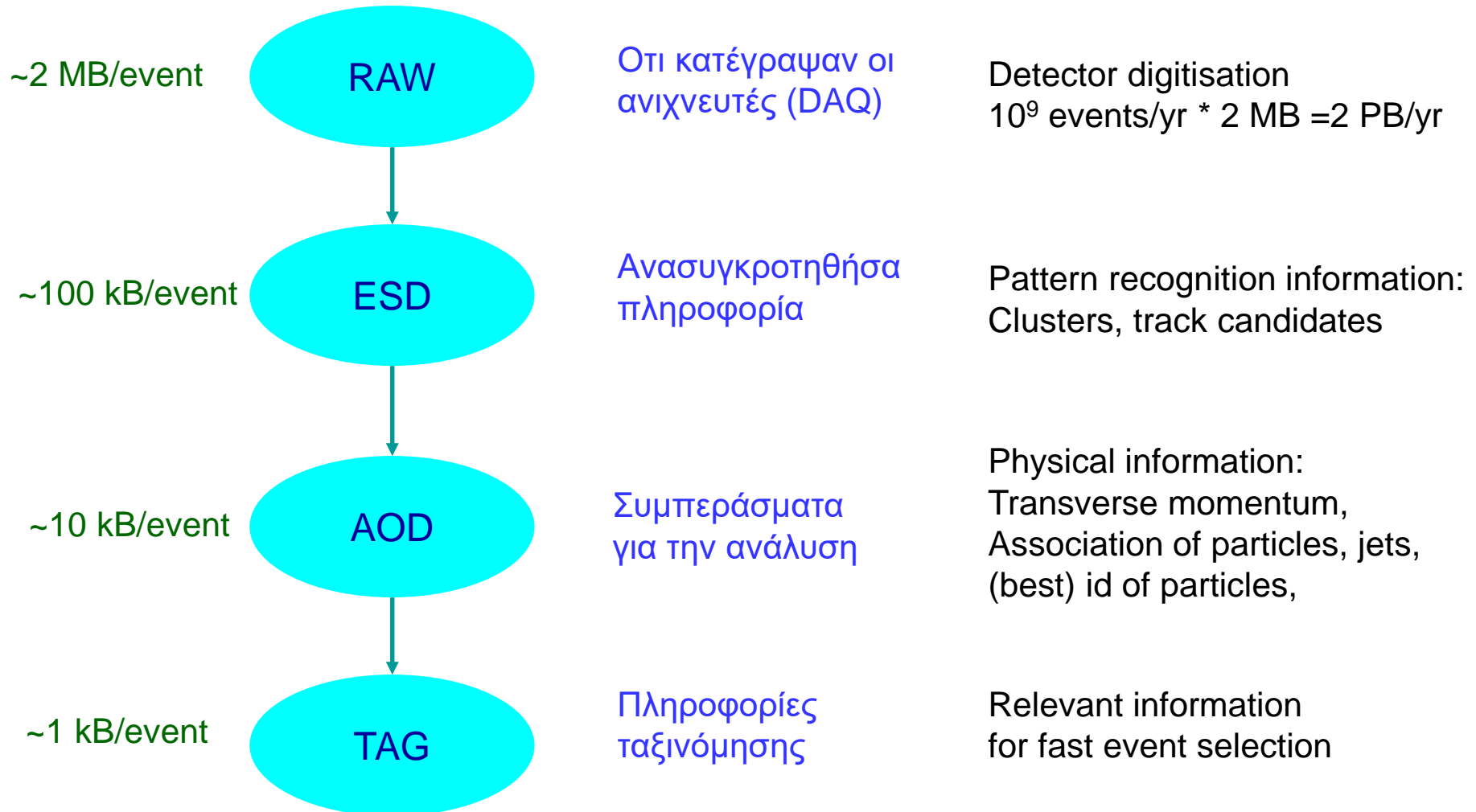
Στάδια Επεξεργασίας Ανάλυση Δεδομένων - Analysis



Ανάλυση:

- ❑ Έπειτα από την ανακατασκευή
- ❑ Χρησιμοποιεί τα 'δεδομένα' (Data) προϊόντα της ανακατασκευής
 - Ιεραρχία 'δεδομένων' από Ολικά (RAW), ως συνοπτικά/συμπερασματικά (AOD)
- ❑ Οι ομάδες ενός πειράματος χρησιμοποιούν τα δεδομένα
 - Στο CERN (δύσκολο) η που ? ... στο Πλέγμα (GRID)
- ❑ Hyratia: μικρό κομάτι ανάλυσης πειράματος στο σχολείο
 - [Εισαγωγή](#), Κοσμός/[Portal](#) Εφαρμογή - Δ.Φασουλιώτης (Εθν.Καπ. Παν. Αθήνας), Α. Αλεξόπουλος(CERN)
 - <http://hyratia.iasa.gr/en/index.html>
 - <http://indico.cern.ch/conferenceDisplay.py?confId=257353#2013-07-08>

Ιεραρχία Δεδομένων (Data)





Distribution of Computing Services

Summary of Con

All experiments - 2008

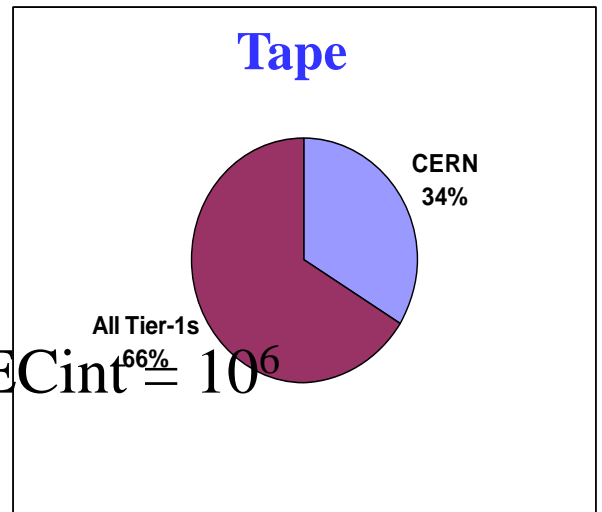
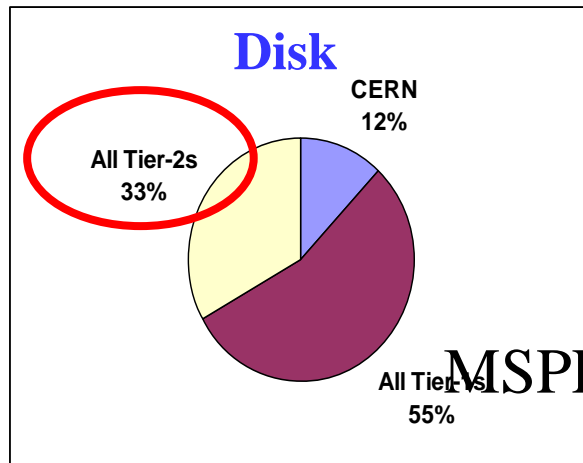
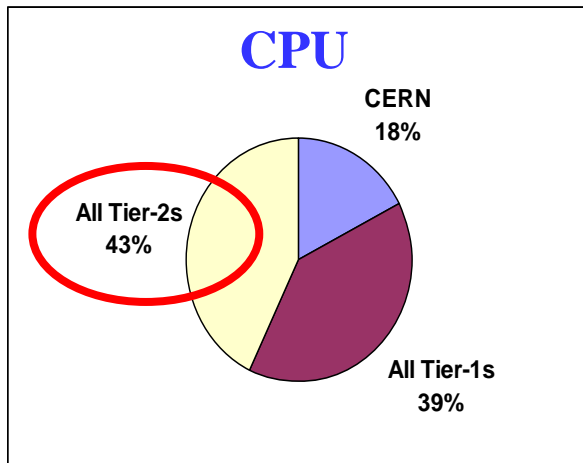
From LCG TDR - June 2008

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
δισκων είναι διαδεδομενα

σε 120 υπολογιστικά κέντρα

MSPECint = 106



Λύση : το Πλέγμα (Grid)

- Χρησιμοποιούμε το Πλέγμα να ενώσουμε τους υπολογιστικούς πόρους των ινστιτούτων ανά τον κόσμο

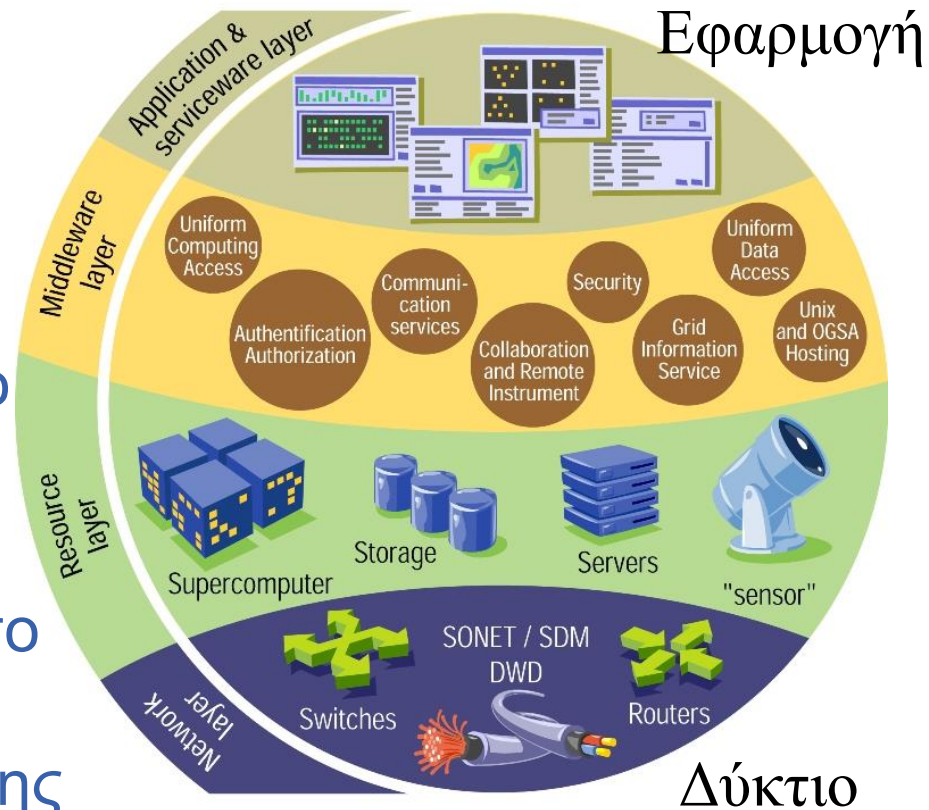
Το World Wide Web παρέχει απλή πρόσβαση σε πληροφορίες που είναι αποθηκευμένες σε πολλά εκατομμύρια διαφορετικές γεωγραφικές τοποθεσίες

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

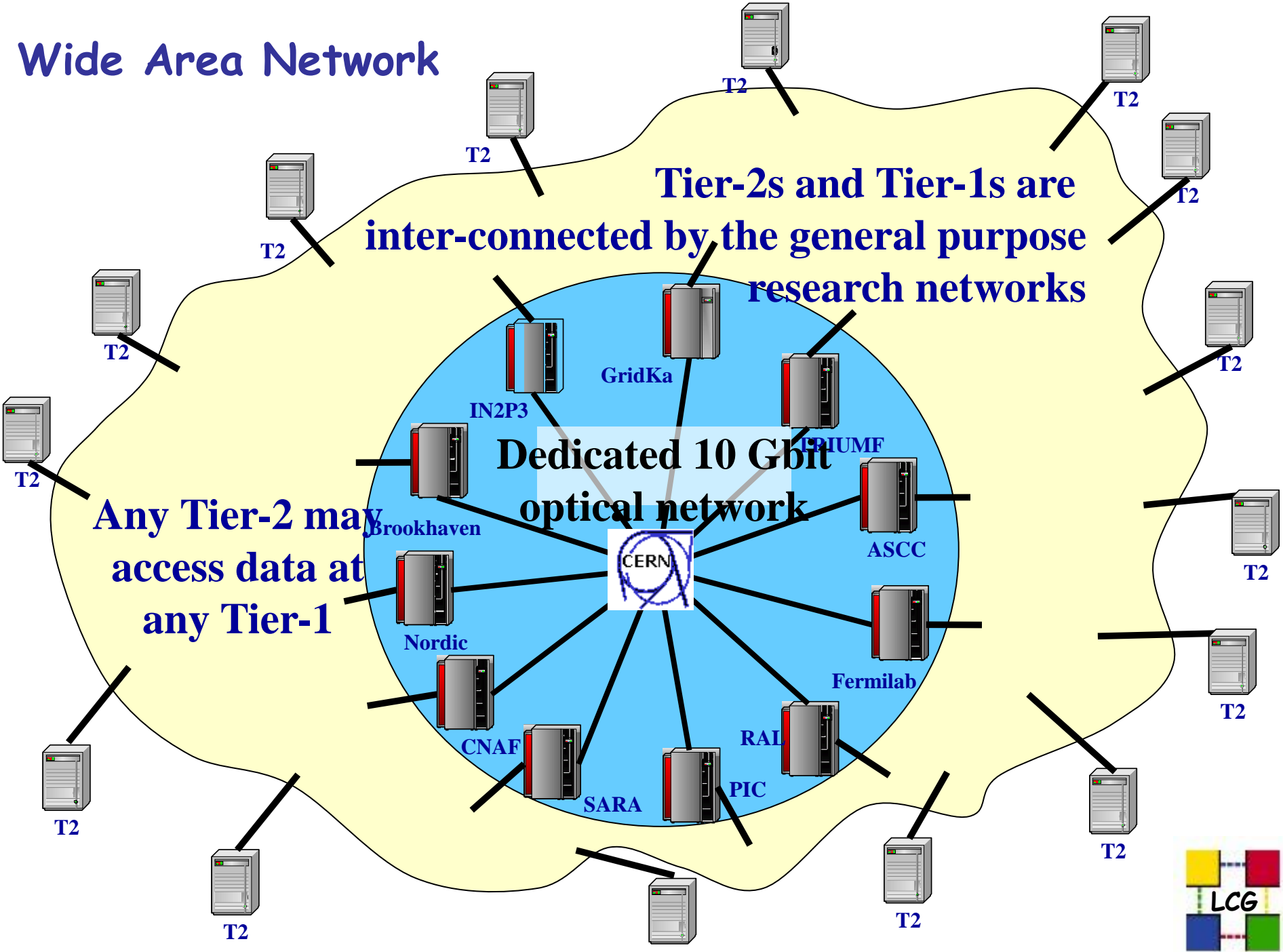


Πως δουλεύει το Πλεγμα (Grid)

- Βασίζεται σε ειδικευμένο λογισμικό, το λεγόμενο **μεσο-λογισμικό** (middleware).
- Η ιδέα του Middleware είναι να βρει αυτόματα τα **στοιχεία** που χρειάζεται ο/η ερευνητής, και το **υπολογιστικό δυναμικό** να τα επεξεργαστεί.
- Middleware εξισοροπεί το φόρτο σε διάφορους πόρους και εγκαταστάσεις. Χειρίζεται επίσης **ασφάλεια**, λογιστική, επιτήρηση και πολλά άλλα.



Wide Area Network

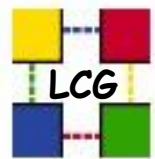


Tier-2s and Tier-1s are inter-connected by the general purpose research networks

Dedicated 10 Gbit optical network



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



- Δεκάδες εφαρμογές σε διάφορους τομείς
 - Φυσική Υψηλών Ενεργειών (Pilot domain)
 - 4 πειράματα LHC, DESY, Fermilab
 - Βιοϊατρική (Pilot domain)
 - Βιοπληροφορική (Bioinformatics)
 - Ιατρική απεικόνιση (Medical imaging)
 - Γεωεπιστήμες
 - Γεω-επισκόπηση
 - Φυσική Στερεάς Γης (Solid Earth Physics)
 - Υδρολογία, Κλίμα
 - Υπολογιστική Χημεία
 - Τήξη (Fusion)
 - Αστρονομία
 - Κοσμικό υπόβαθρο μικροκυμάτων
 - ακτίνων-γ
 - Γεωφυσική
 - Βιομηχανικές εφαρμογές



Λειτουργικά συστήματα

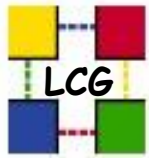
- z Linux σε σχεδόν όλους τους υπολογιστές
 - γ CERN Data center: 450,000 πυρήνες και 10,000 servers
 - γ Χαμηλότερο κόστος ανά υπολογιστική δύναμη
- z Στά γραφεία / laptops
 - γ MacOS για προγραμματιστές – παρόμοιο Unix (σαν το Linux) + λογισμικά σαν το 'Office'
 - γ PC με Windows
 - γ PC/laptops με Linux

Γλώσσες προγραμματισμού

- z C++ για τα κύρια τμήματα
 - γ Τα πιο πολύπλοκα, όσα κάνουν πολύ δουλειά
- z Python – για ελαφριές 'δουλιές' η ειδικές
 - γ κώδικα που ενώνει, καθοδηγεί, επιλέγει
 - γ ανάλυση δεδομένων, βαθειά μηχανική μάθηση
- z Java – αυτοματισμούς, JavaScript - ιστό
- z CUDA – για προγραμματισμό GPU
- z Julia - πειραματικές χρήσεις

Οι φυσικοί προγραμματίζουν

- z Ολη η ανάλυση μετρήσεων γίνεται με υπολογιστές
- z Οι φυσικοί ΥΕ χρησιμοποιούν ειδικευμένα προγράμματα
 - y Μερικοί γράφουν μεγάλες κομμάτια η πολλές υπορουτίνες (methods) σε C++, η άλλες γλώσσες πχ Python
 - y Οι περισσότεροι γράφουν μικρές Python υπορουτίνες, για τις δικές τους ανάγκες
- z Όλοι χρησιμοποιούν λογισμικά 'εργαλεία'
 - y για να δουν τις περιλήψεις των μετρήσεων

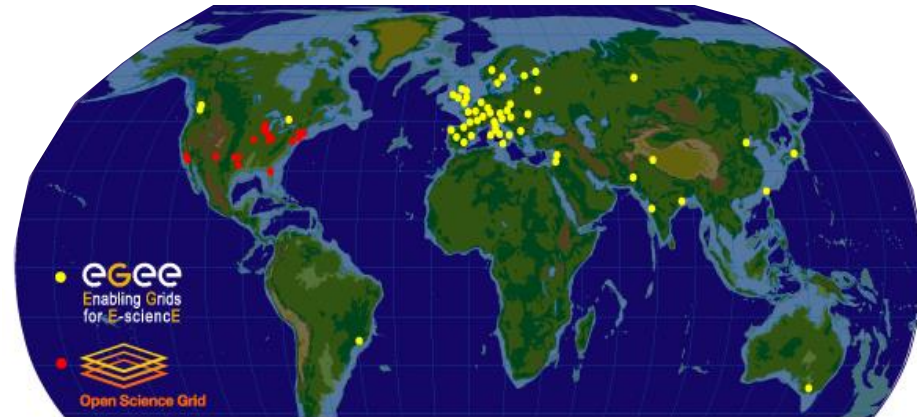


Backup



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



From:
Cosmus
Project,

AIRES
Project,

Sergio
Sciutto

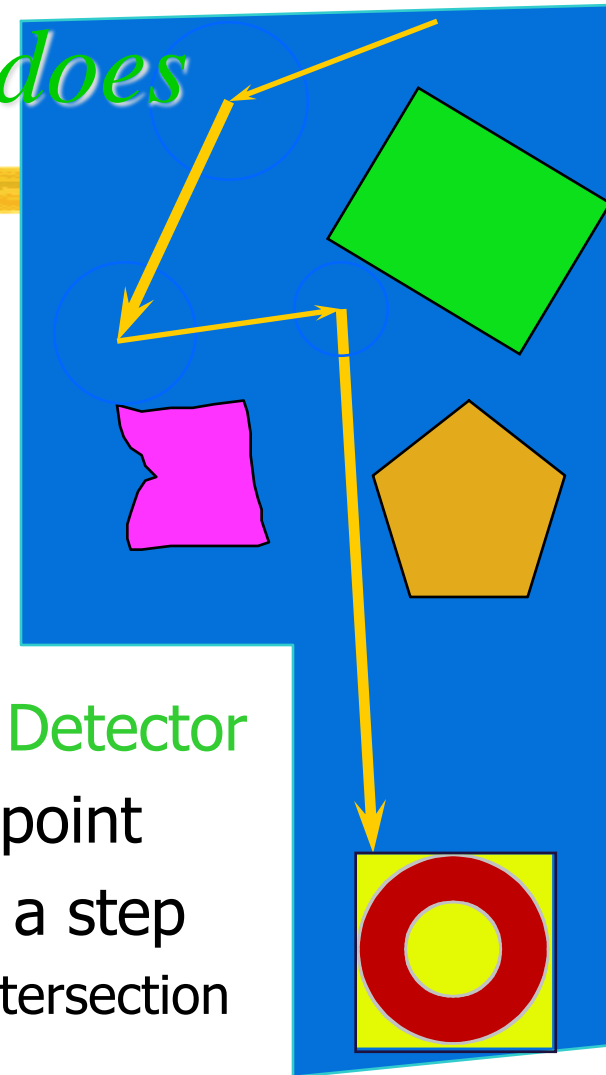
Geant4 geometry: what it does

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

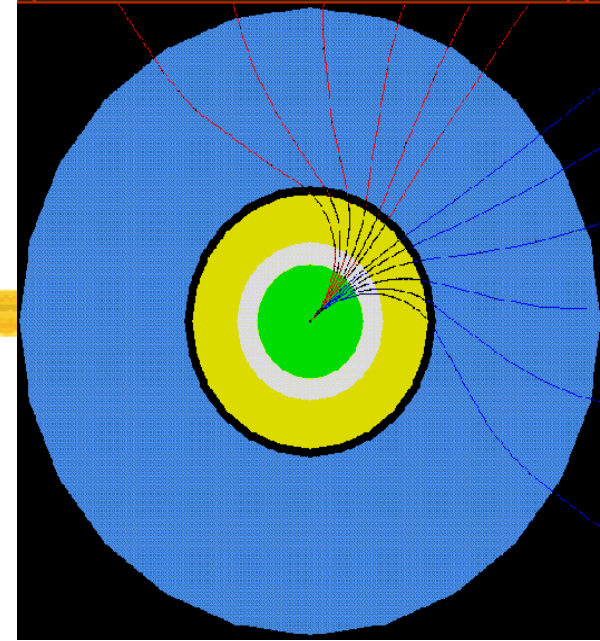
Navigates in Detector

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



Propagating in a field

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



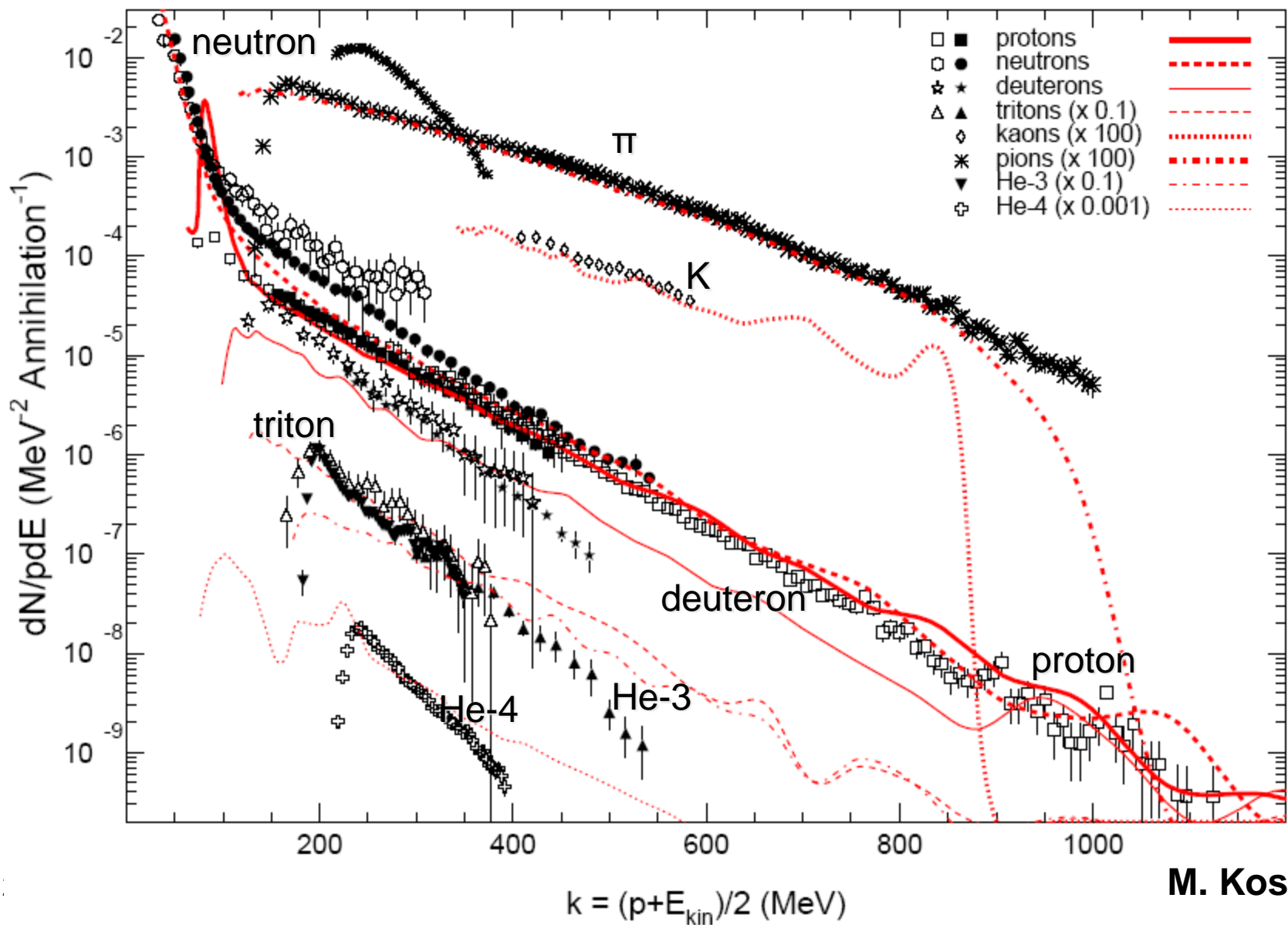
- z It is possible to tailor
 - y the accuracy of the splitting of the curve into linear segments,
 - y the accuracy in intersecting each volume boundaries.
- z These can be set now to different values for a single volume or for a hierarchy.

Electromagnetic physics

- z Gammas:
 - y Gamma-conversion, Compton scattering, Photo-electric effect
- z Leptons(e , μ), 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}U nucleus

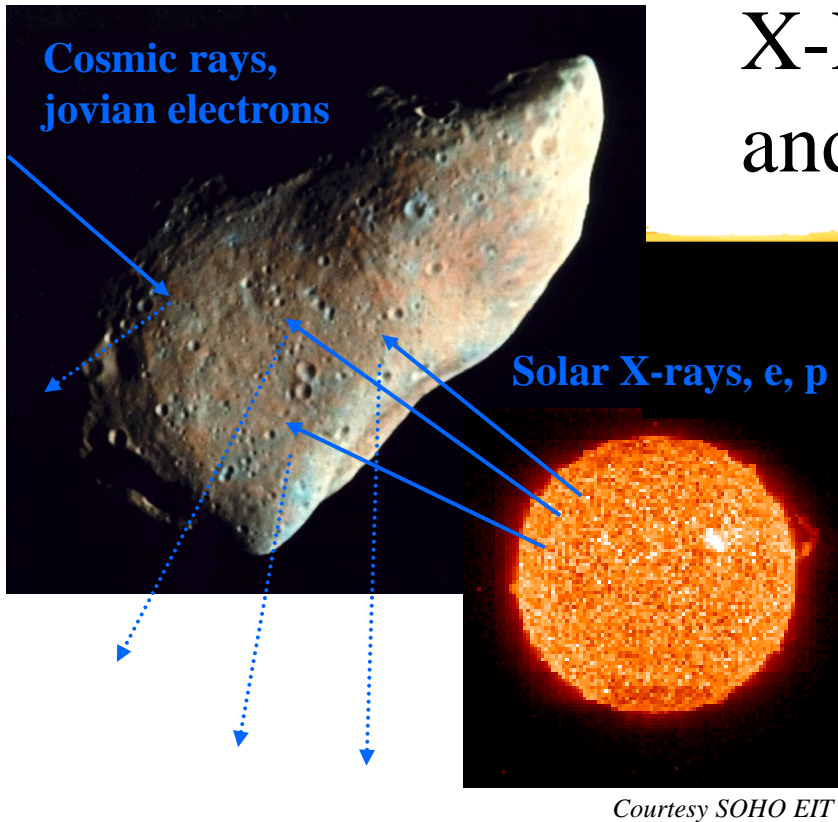


Simulation ‘packages’

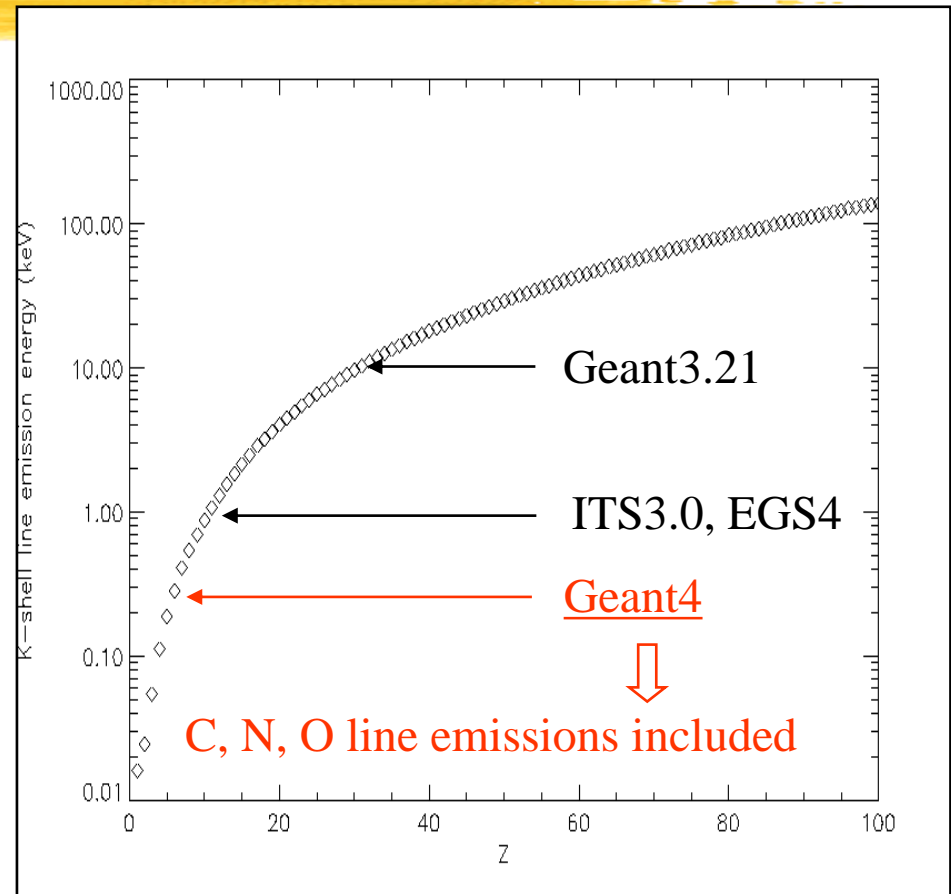


- z Provides the means to simulate
 - y the **physical processes** and
 - y **detector response** of an experiment.
- z As was realised by many in the past,
 - y **most of the parts** needed can be **common** between experiments (eg physics, geometry blocks) .
- z So it makes eminent sense to create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

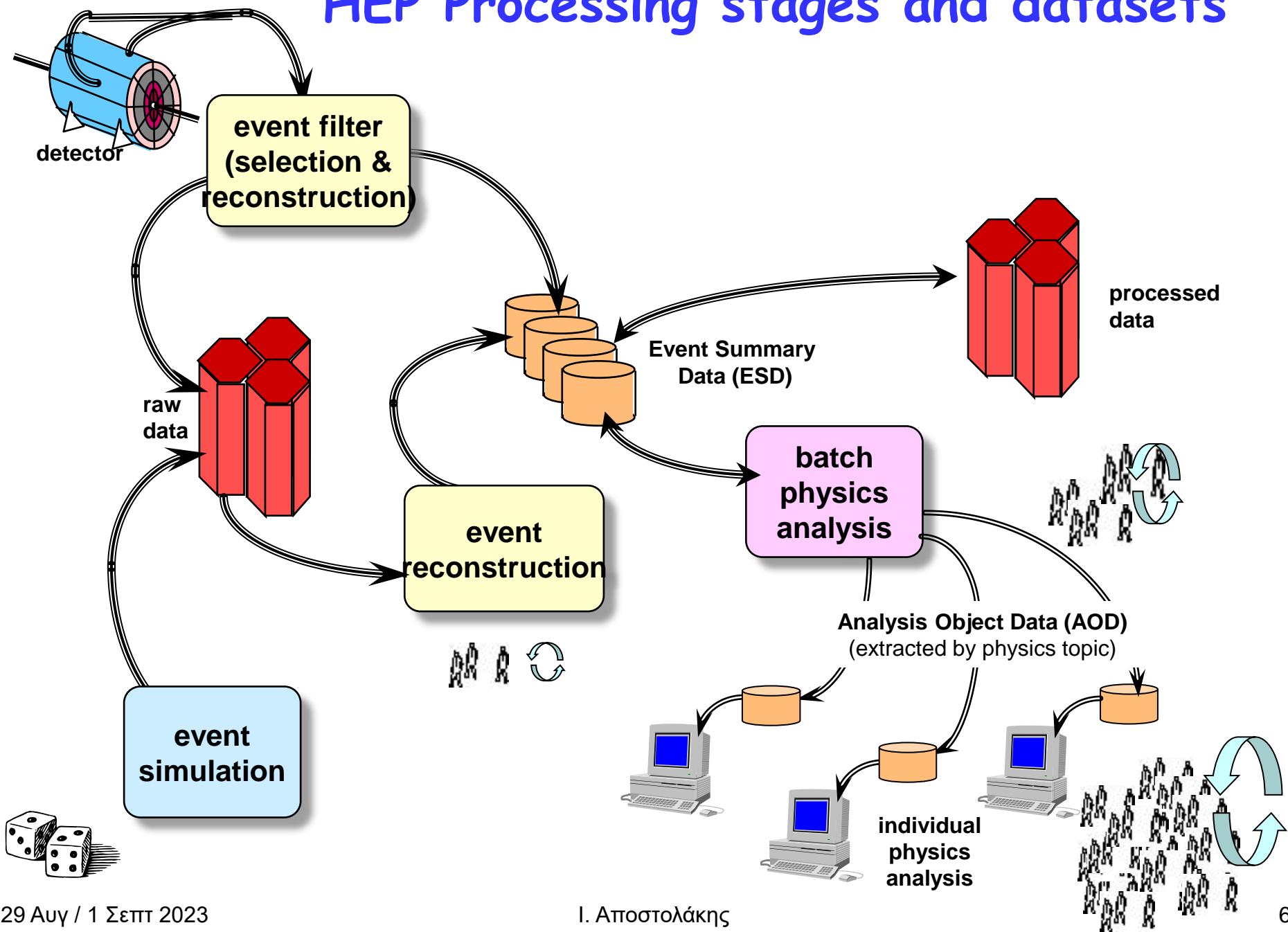
X-Ray Surveys of Asteroids and Moons



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

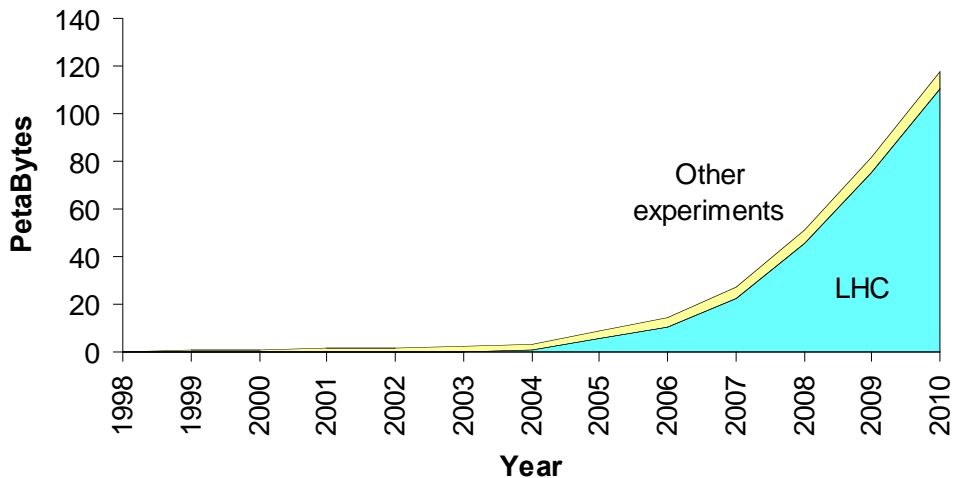


HEP Processing stages and datasets

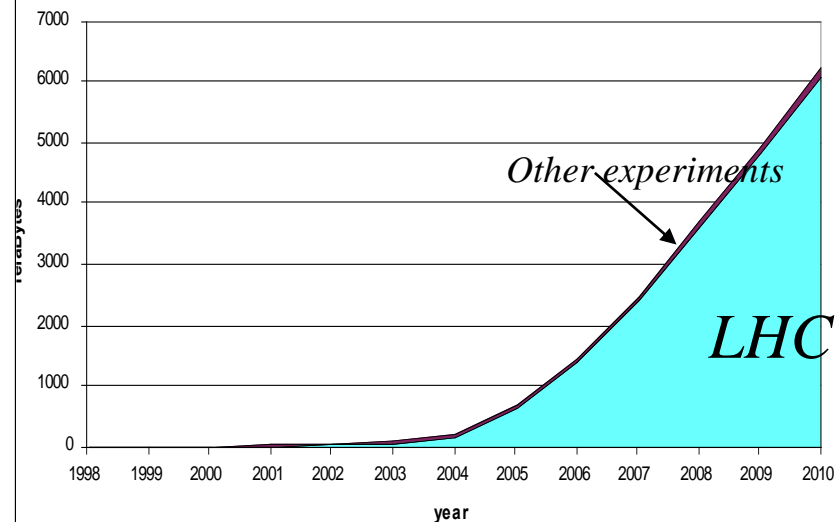


CERN Centre Capacity Requirements for all

Estimated Mass Storage at CERN



Estimated DISK Capacity at CERN



processing

K 20000

3 700

8 2000

10 1000

25,000

34,000

disk

PE

5.0

6.7

tape media

PE

36

48

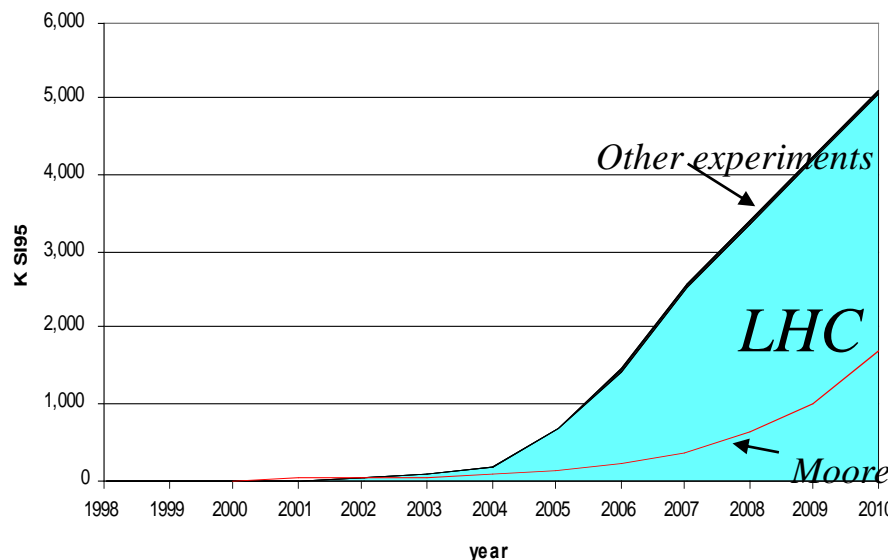
tape I/O

GI

39

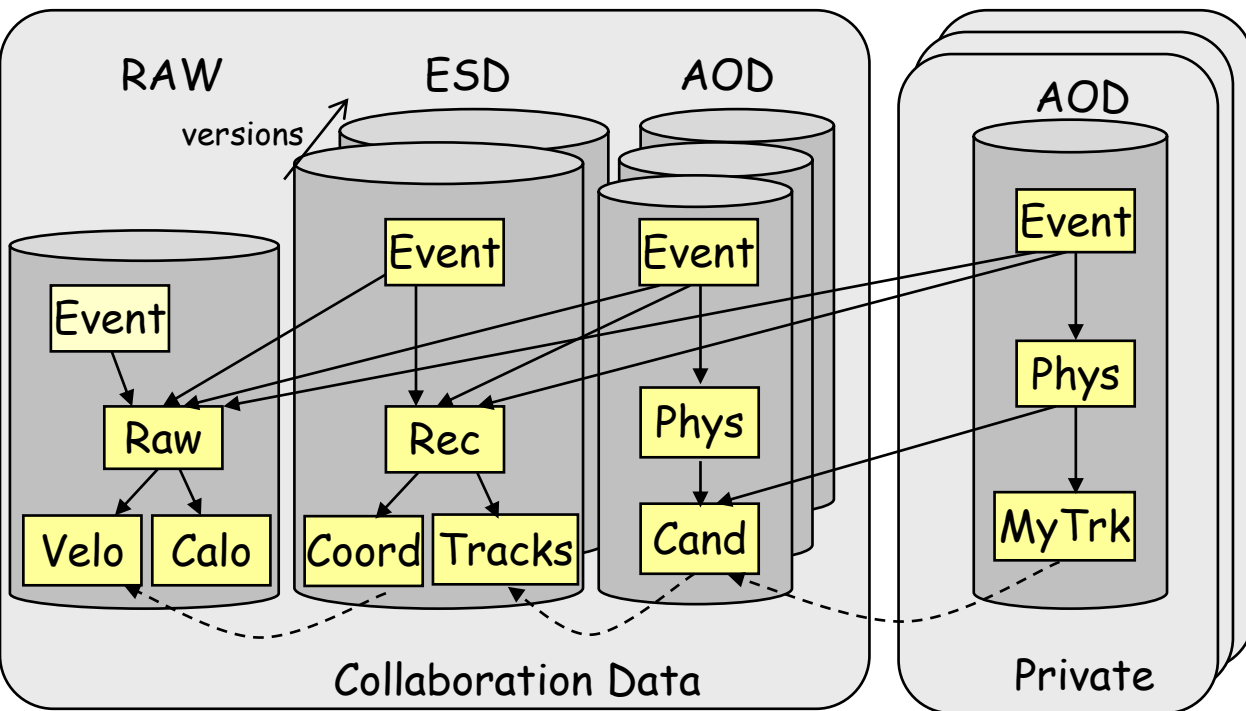
39

Estimated CPU Capacity at CERN



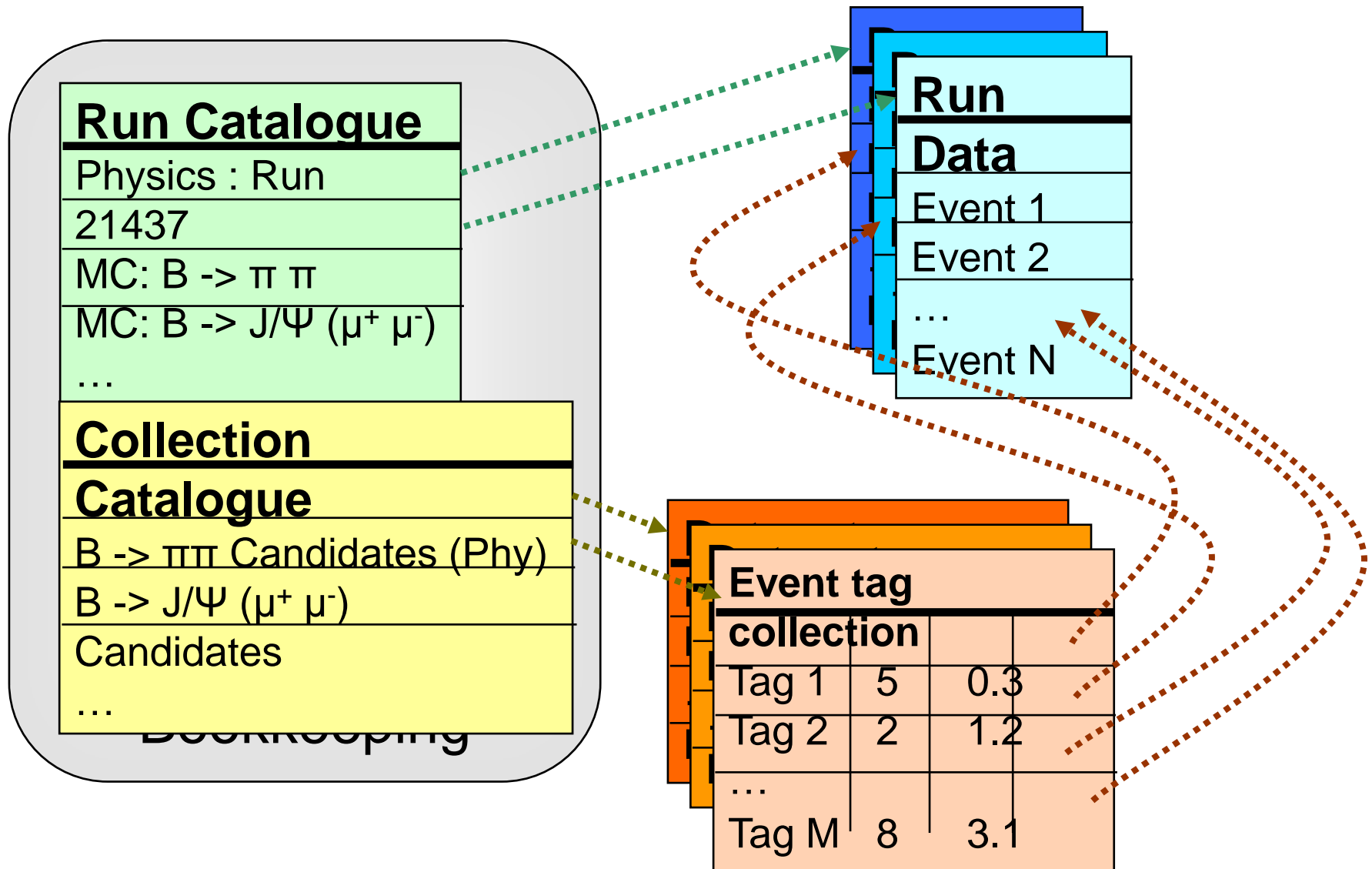
Moore's law

Event Data



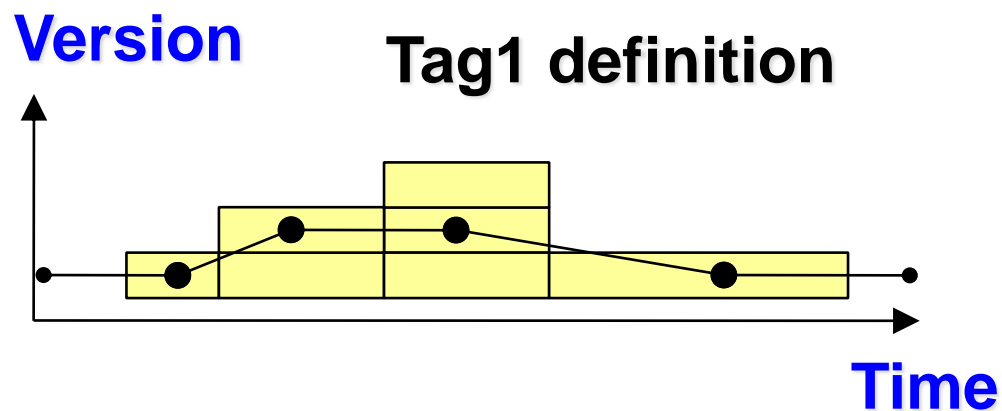
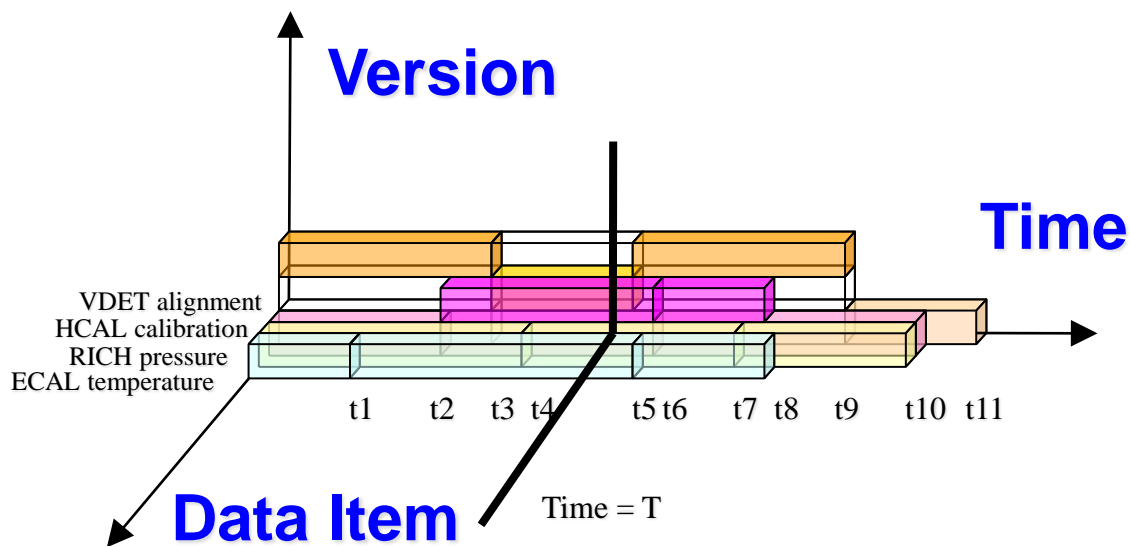
- ❑ Complex data models
 - ~500 structure types
- ❑ References to describe relationships between event objects
 - unidirectional
- ❑ Need to support transparent navigation
- ❑ Need ultimate resolution on selected events
 - need to run specialised algorithms
 - work interactively
- ❑ Not affordable if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

HEP Metadata - Event Collections

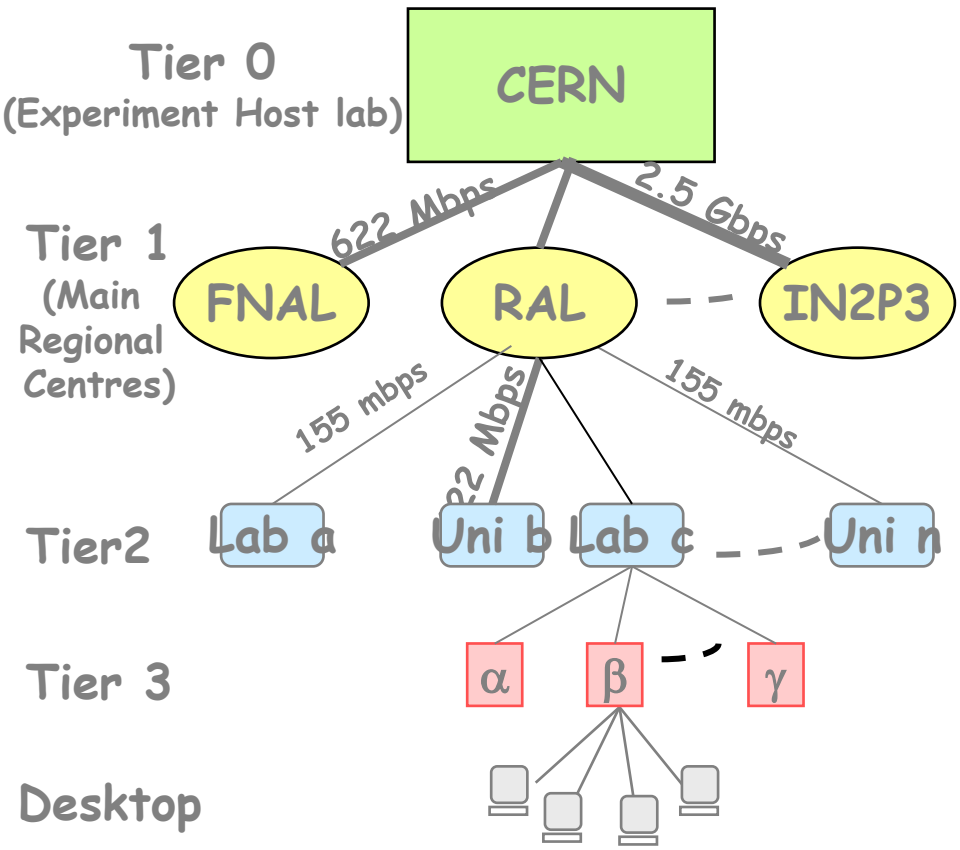


Detector Conditions Data

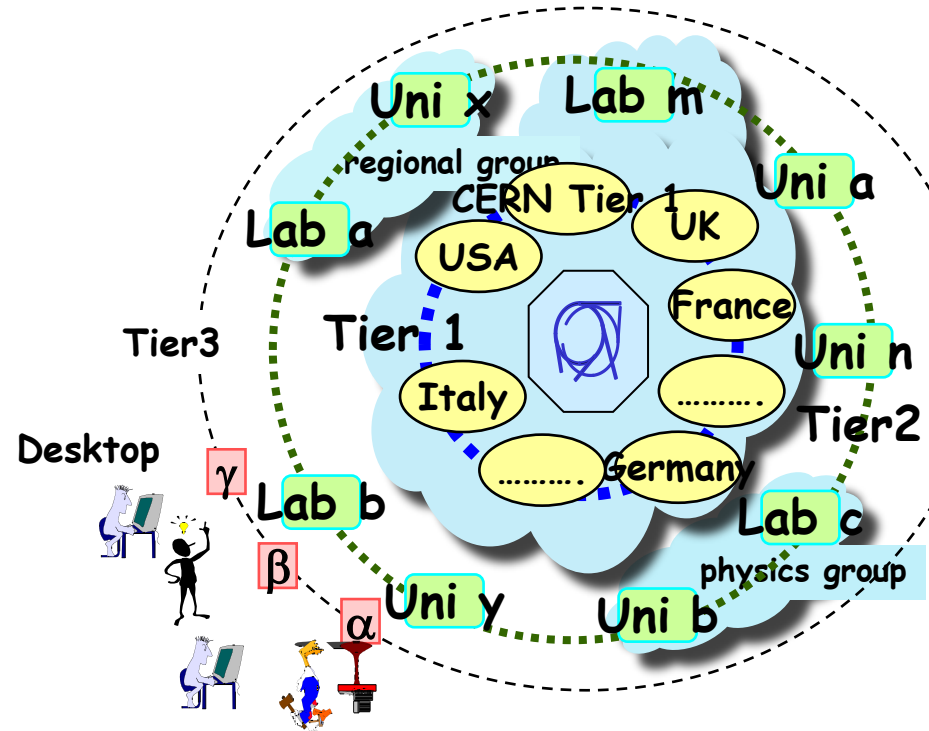
- ❑ Reflects changes in state of the detector with time
- ❑ Event Data cannot be reconstructed or analyzed without it
- ❑ Versioning
- ❑ Tagging
- ❑ Ability to extract slices of data required to run with job
- ❑ Long life-time



A Multi-Tier Computing Model



Manager View

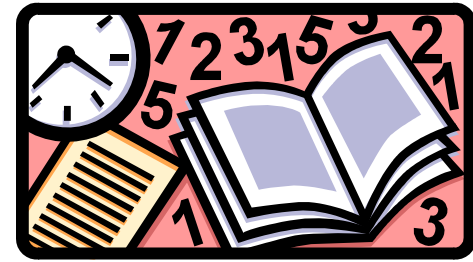


User View

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

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’).

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)

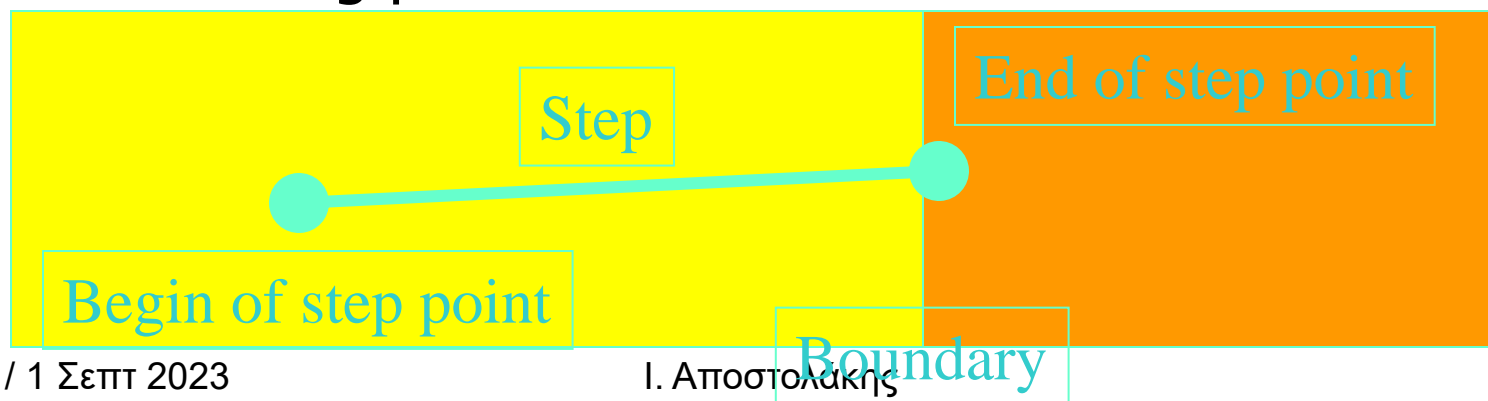
- z During a step (continued)
 - y An (optional) user-written 'action' is called,
 - x Which can be used eg to create histograms or tallies.
 - y If the current volume contains a sensitive detector, that is addressed, allowing it eg
 - x to record the energy deposited,
 - x to record the exact positionin 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.



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:
 - y LHC
 - y heavy ions, CP violation, cosmic rays
 - y medical and space science applications
- z **World-wide collaboration**