Data Processing for Particle Physics

- Focus on LHC

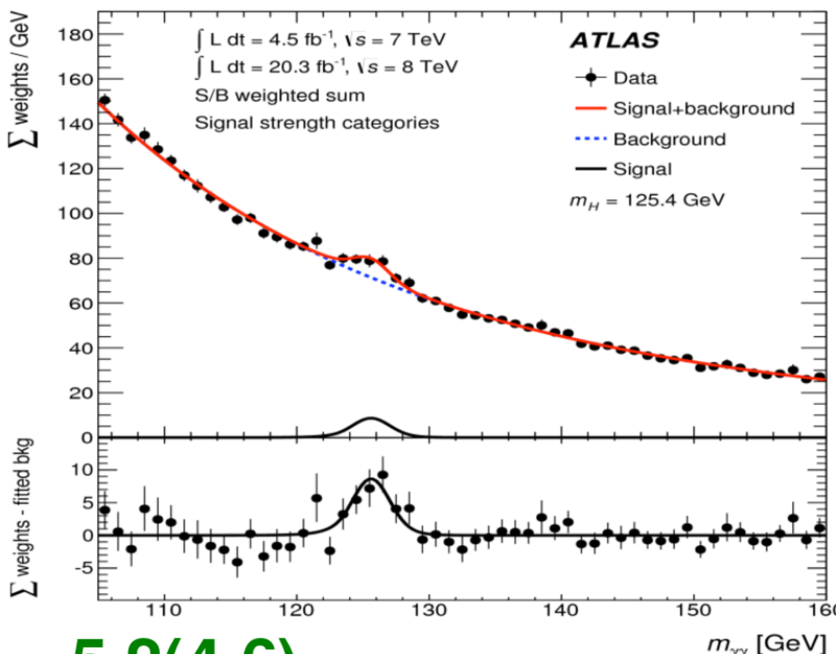
Jike Wang

声明 : lots of materials borrowed from M. Elsing,
B. Dahmes, etc.

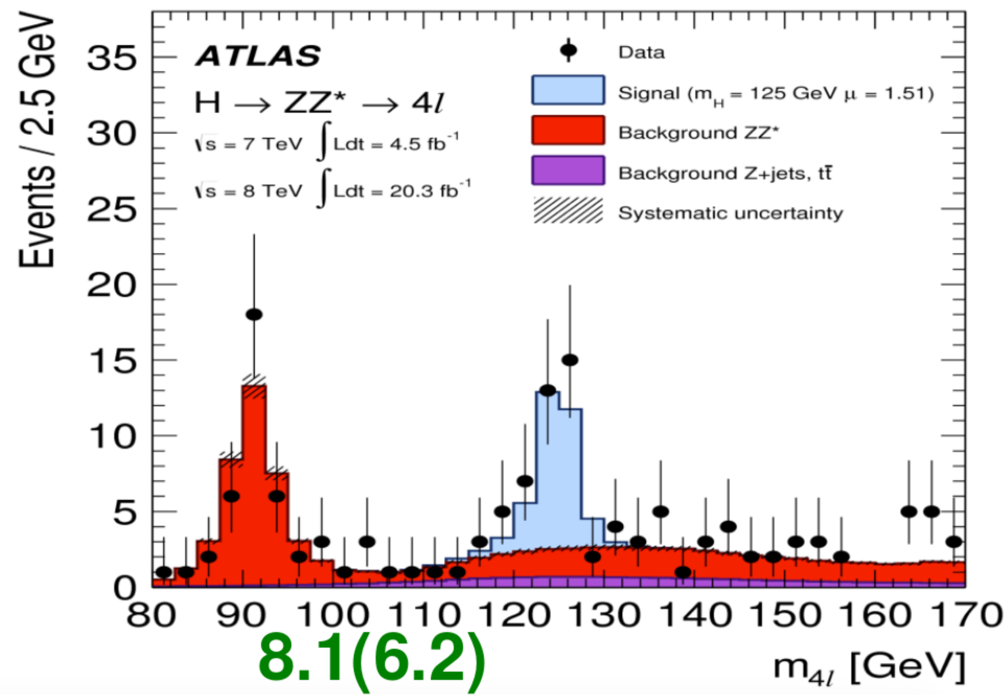
18/July/2018, iSTEP2018, Wuhan, China

Higgs Plots

- This is a beautiful mass distribution. But where are the data come from and how ?
- In detector we only have electronic info

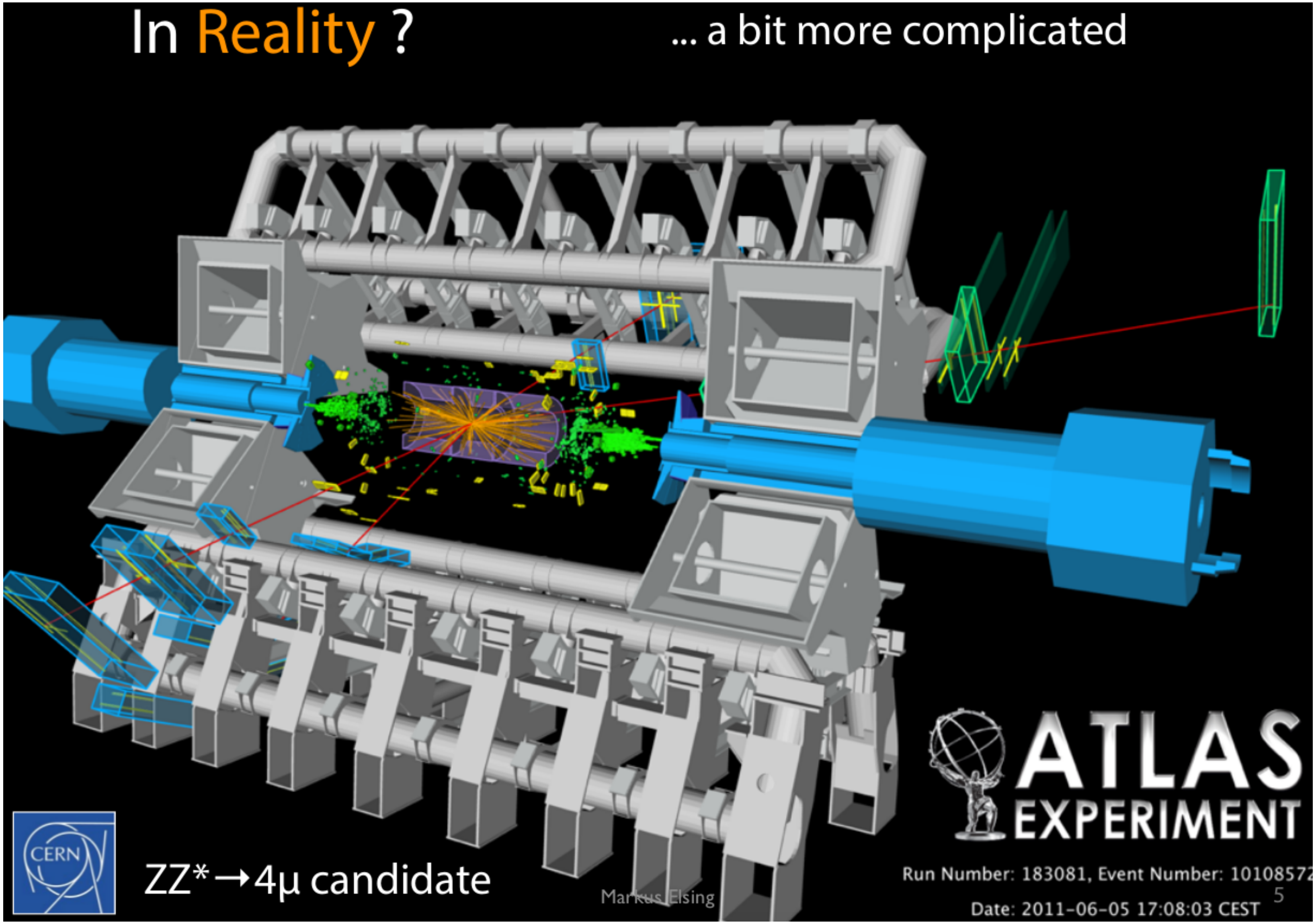


5.2(4.6)



8.1(6.2)

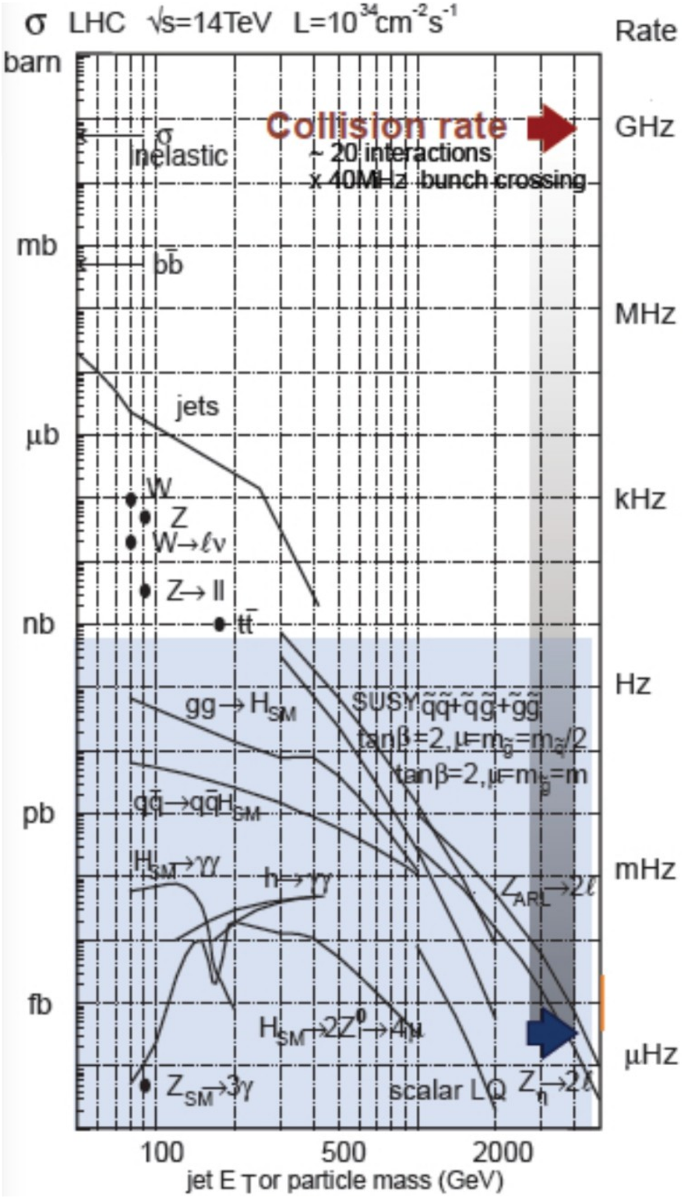
Higgs→ZZ→4l


In Reality ?
... a bit more complicated
ATLAS
EXPERIMENT
CERN
ZZ*→4μ candidate
Markus Elsing
Run Number: 183081, Event Number: 10108572
Date: 2011-06-05 17:08:03 CEST
5

Challenging

- Most of the interesting physics processes have very low production rates
- Most are rubbish

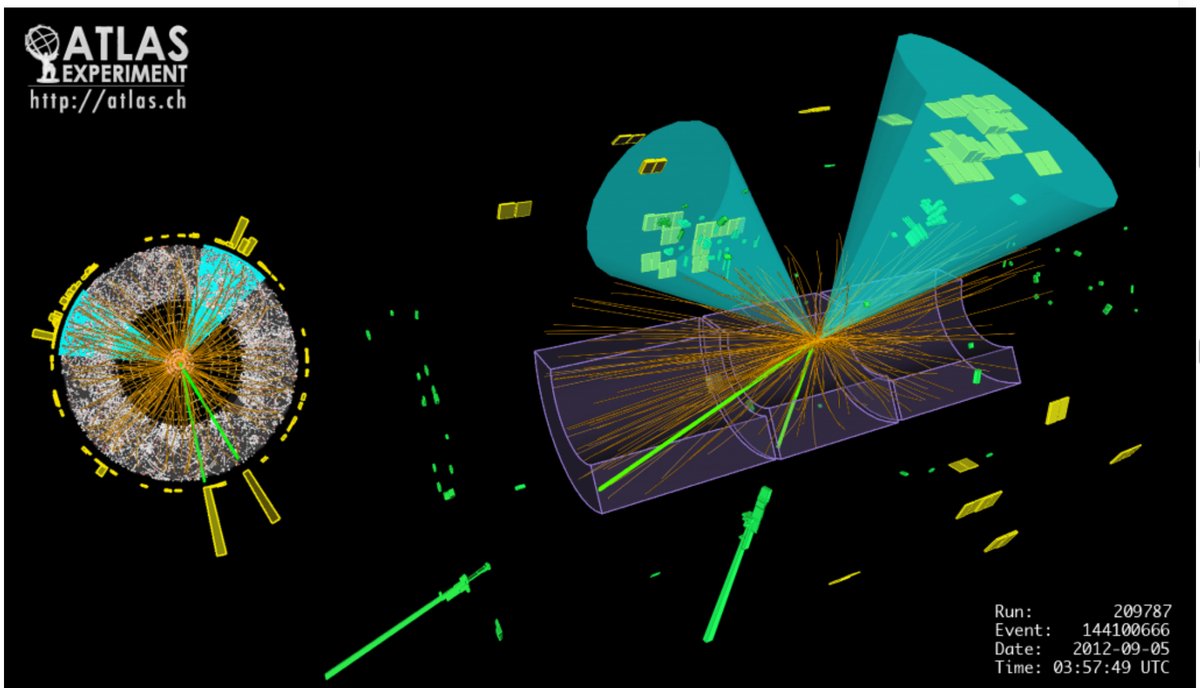
Process	Production Rate $10^{34}\ cm^{-2}s^{-1}$
inelastic	~1 GHz
bbbar	5 MHz
$W \rightarrow l\nu$	150 Hz
$Z \rightarrow l\nu$	15 Hz
ttbar	10 Hz
Z’	0.5 Hz
H(125) SM	0.4 Hz

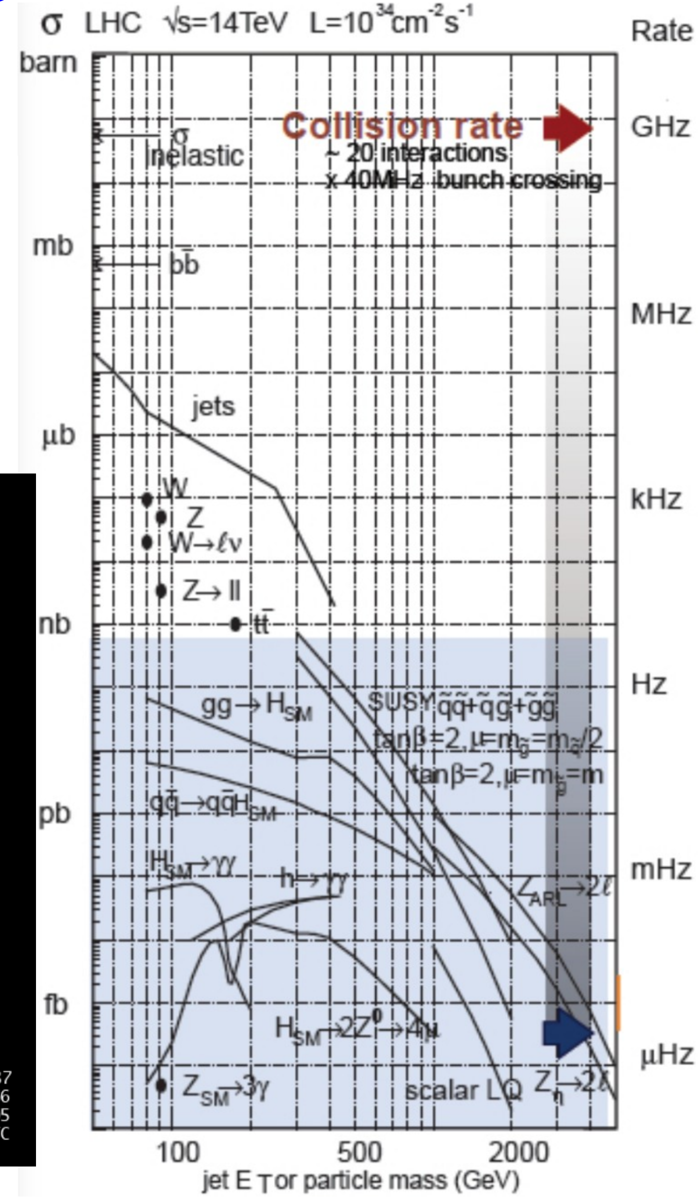


Challenging

- Roughly one higgs produced for every **10,000,000,000** pp interactions

A Zh→μμbb candidate





Challenging

- 1 in 10,000,000,000:
  - Like looking for a single drop of water from the Jet d'Eau over 30 minutes

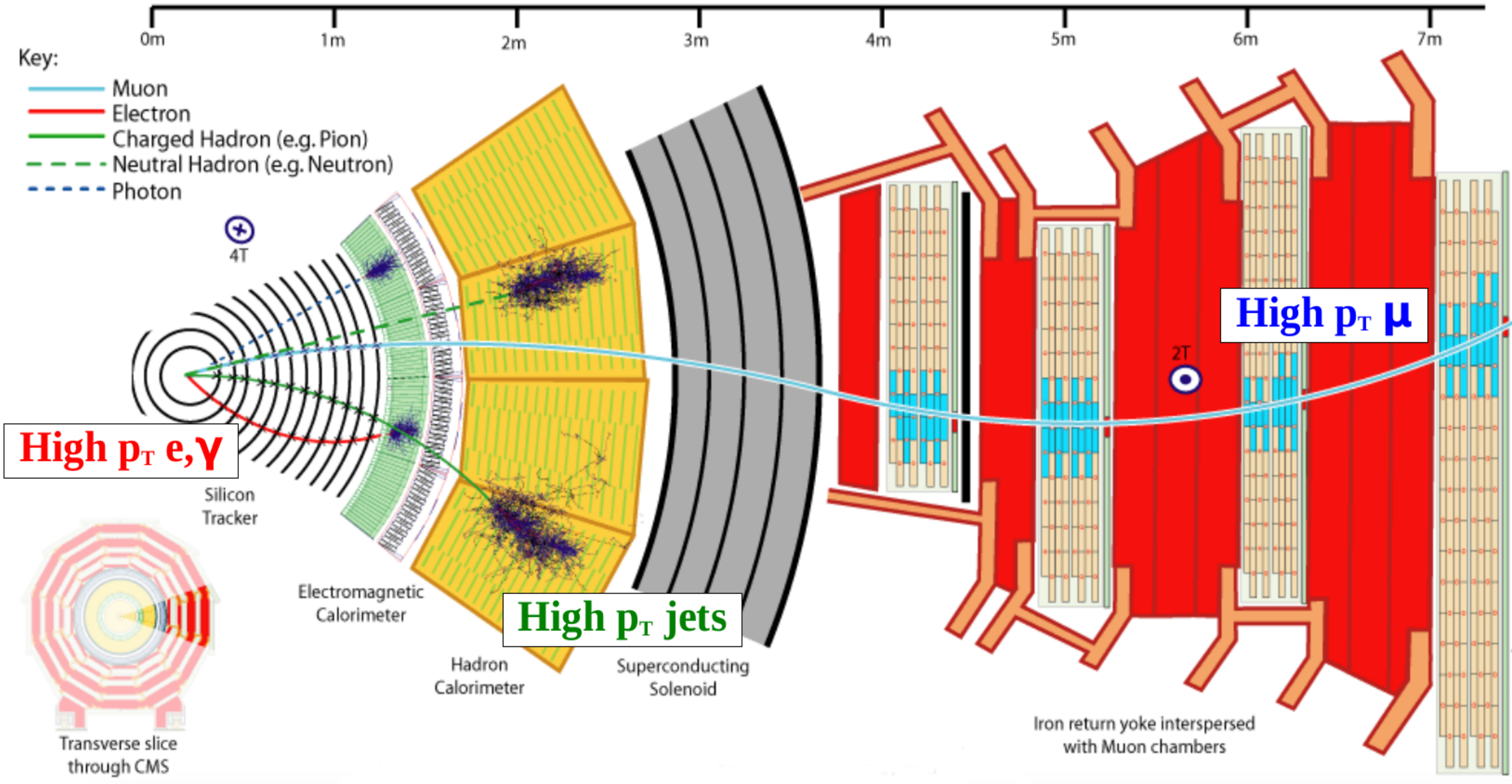
Trigger ?!

• Trouble: We must analyze and **reject most** LHC collisions prior to storage

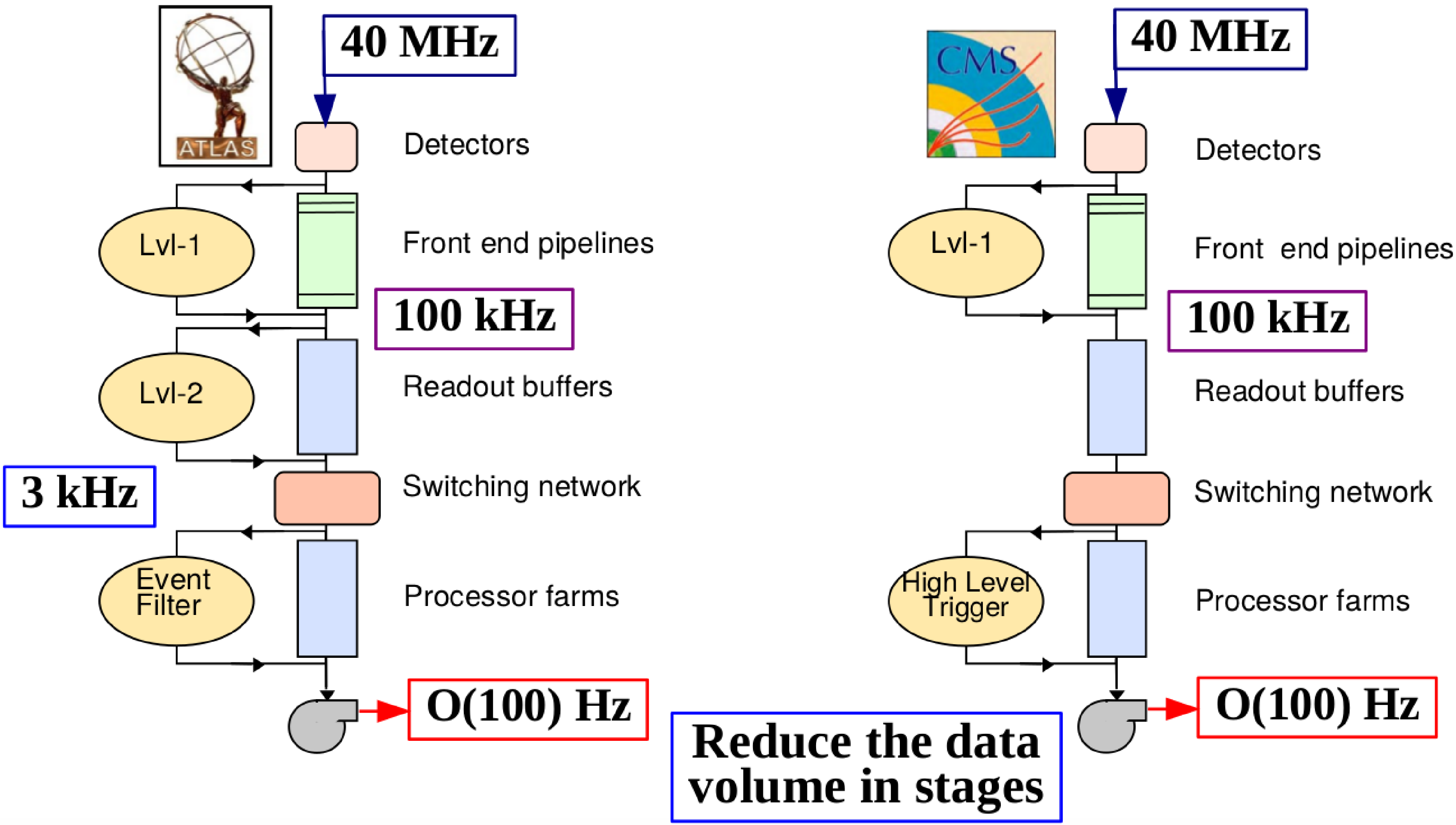
• **Solution: Trigger**

• **Should be:**
- Fast processing
- High rejection factor
- High efficiency for interesting physics
  → If most of signals are killed, why we do all the thing ??
- Flexible
- Affordable

Trigger Objects



Trigger Setup



Trigger Setup

- **Level 1: Custom hardware and firmware:**
  - Reduces the rate from 40 MHz to 100 kHz
  - Advantage: speed

- **Level 2: Computing farm (software):**
  - Further reduces the rate to a few kHz
  - Reconstruct a region surrounding the L1 trigger object
  - Advantage: Further rejection, still relatively fast

- **Level 3: Computing farm (software):**
  - Store events passing final selection for offline analysis
  - Advantage: The best reconstruction

High Level Trigger

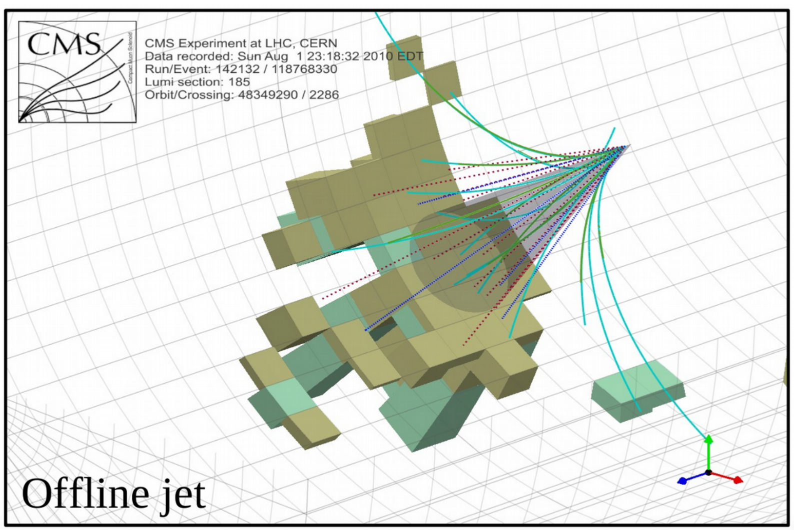
L1 Trigger

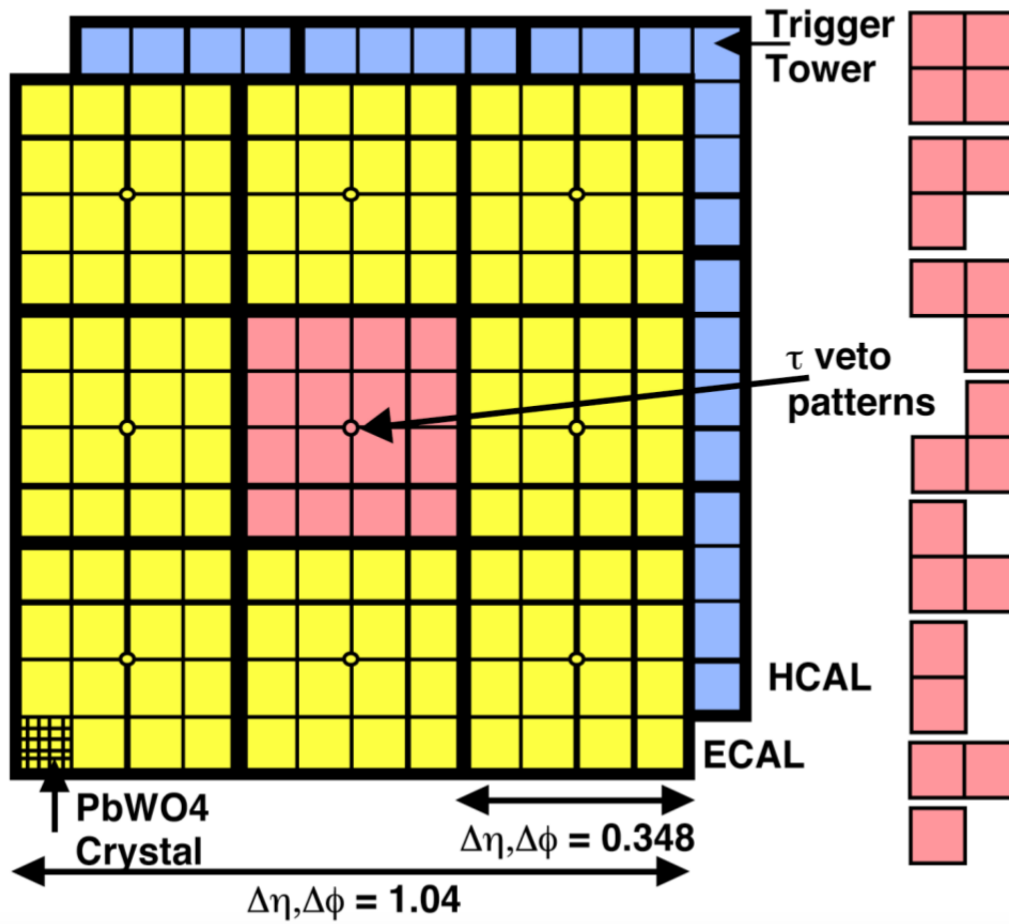
- **Custom electronics designed to make very fast decisions:**
  - Application-Specified Integrated Circuits (ASICs)
  - Field Programmable Gate Arrays (FPGAs)

- **Must be able to cope with input rate of 40 MHz:**
  - Otherwise trigger wasting time (and money), as new events keep arriving
  - Event buffering is expensive, too

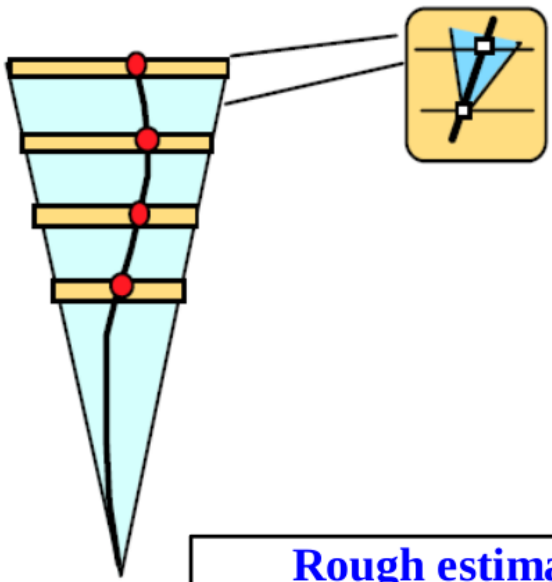
- **L1 Trigger: Pipeline**
  - Process many events at once
  - Parallel processing of different inputs as much as possible

L1 Calorimeter Trigger

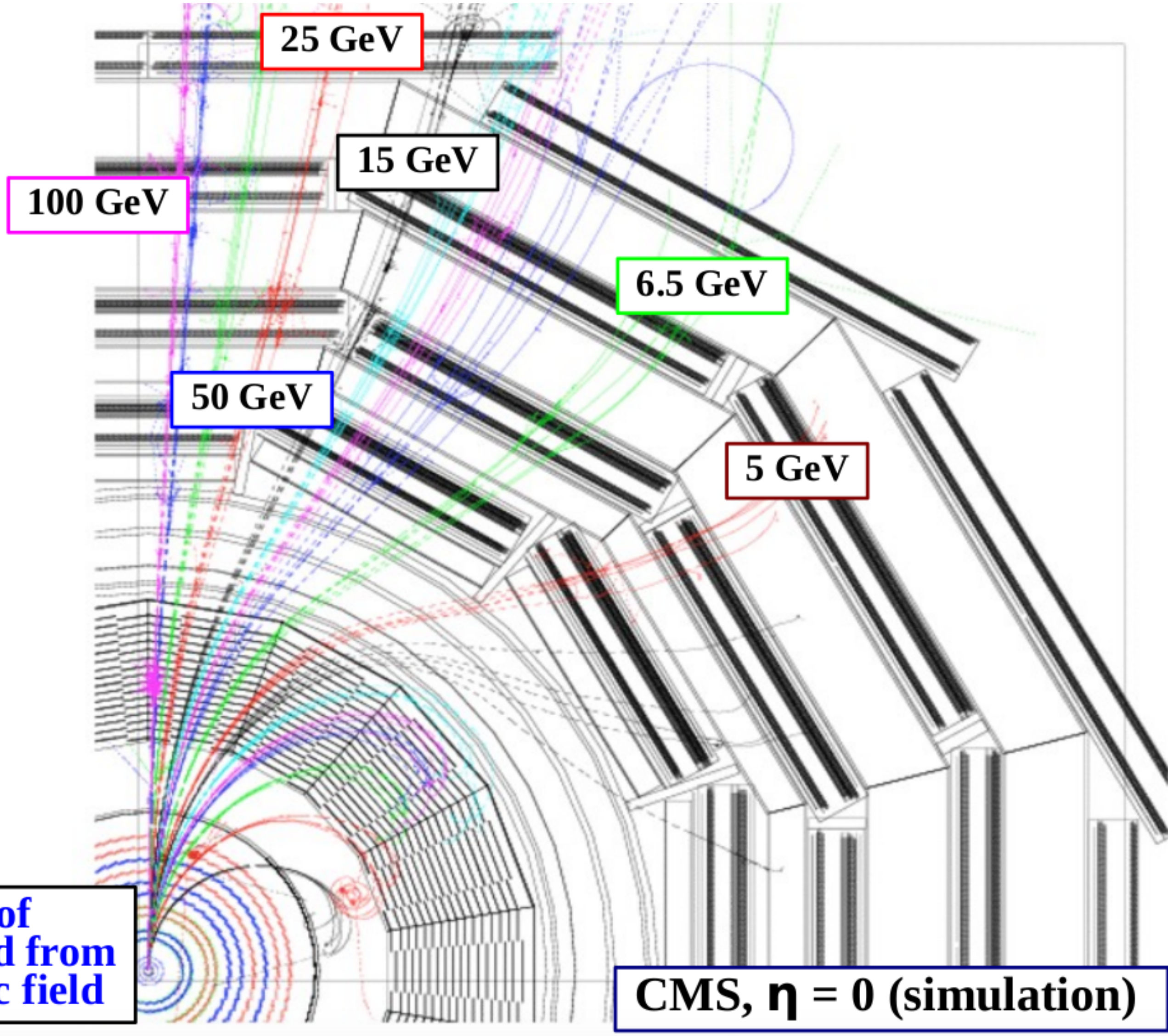




L1 Muon Trigger

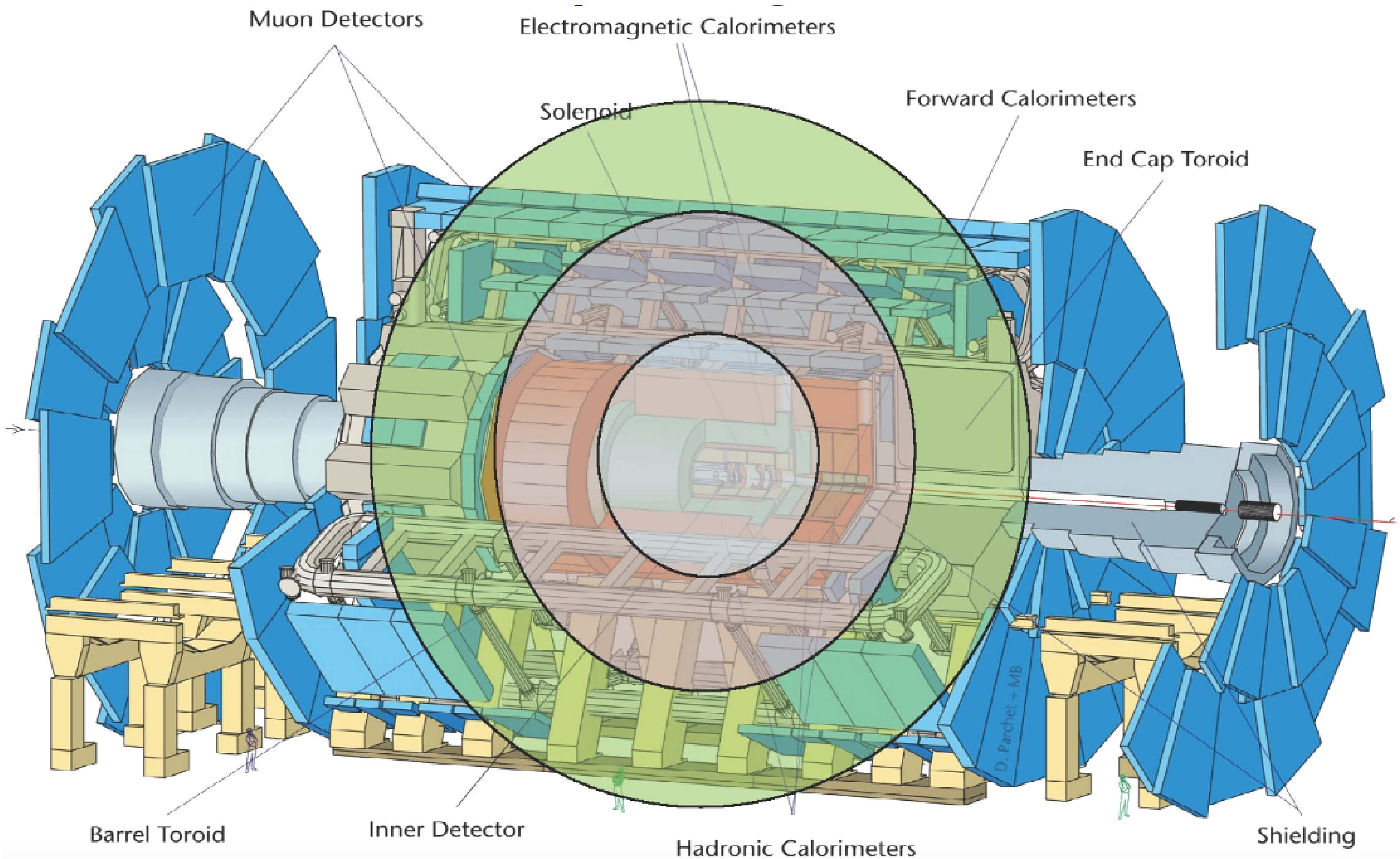


Rough estimate of
muon pT determined from
bending in magnetic field



Global

- **We still need a global decision**
  - We have the information, does the event pass?
  - Decision needs to be made quickly



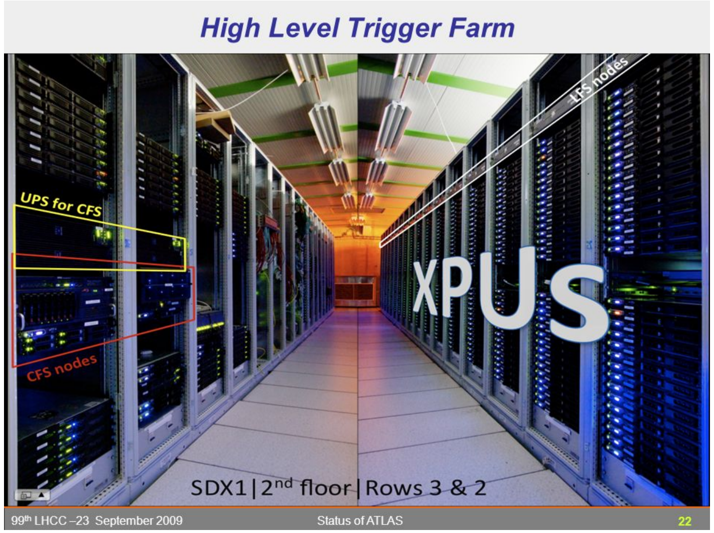
High Level Trigger

- From L1 we expect a large rate (up to 100 kHz) of events that "might be interesting"

- These events are not kept yet (rate too high for storage), but sent to the HLT for additional filtering
  - Massive commercial computer farm
  - ATLAS: L2 and L3 handled by separate computing farms
    Roughly 17k CPUs that can be freely assigned to either
    CMS: Single computing farm (roughly 13k CPUs)

- Parallel processing, each CPU processes individual event

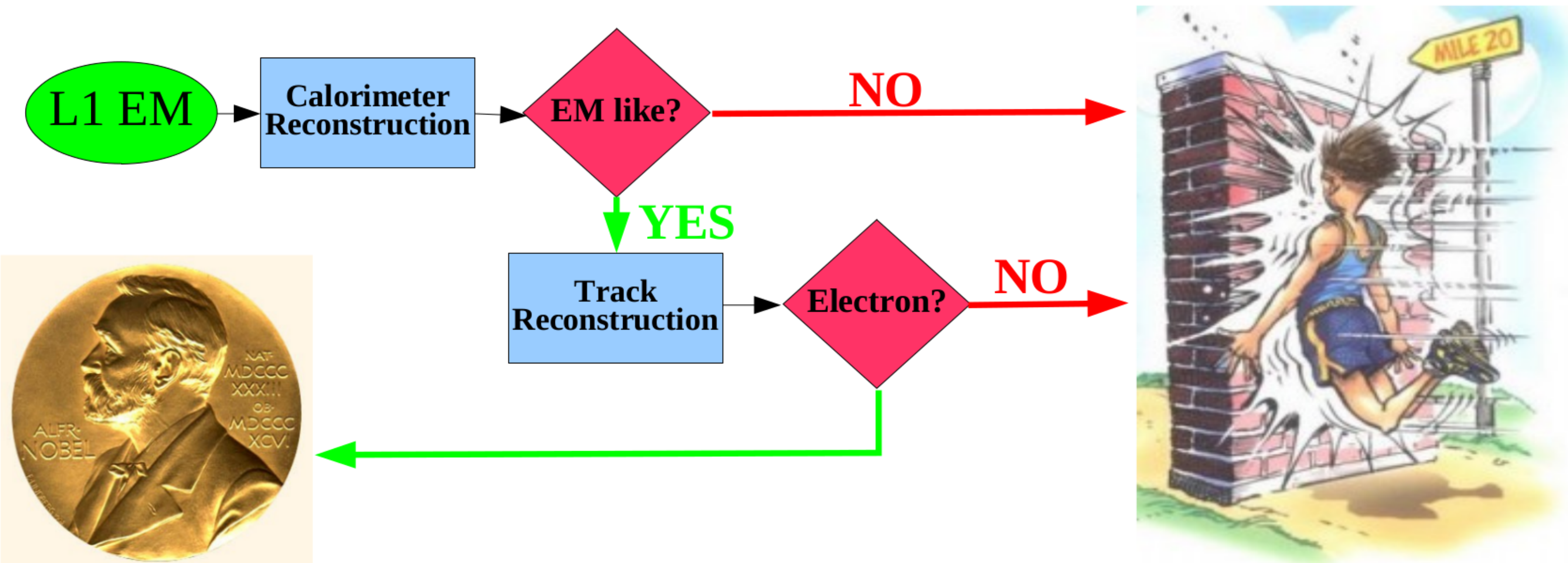
- Resources are still limited
  - Offline: Full reconstruction takes seconds (minutes)
  - Online latency: milliseconds (input rate dependent)



Should Be Fast

- HLT is composed of hundreds of trigger algorithms
  - Software design, so no strict limit on the number of algorithms
  - Each designed with a specific physics signature in mind

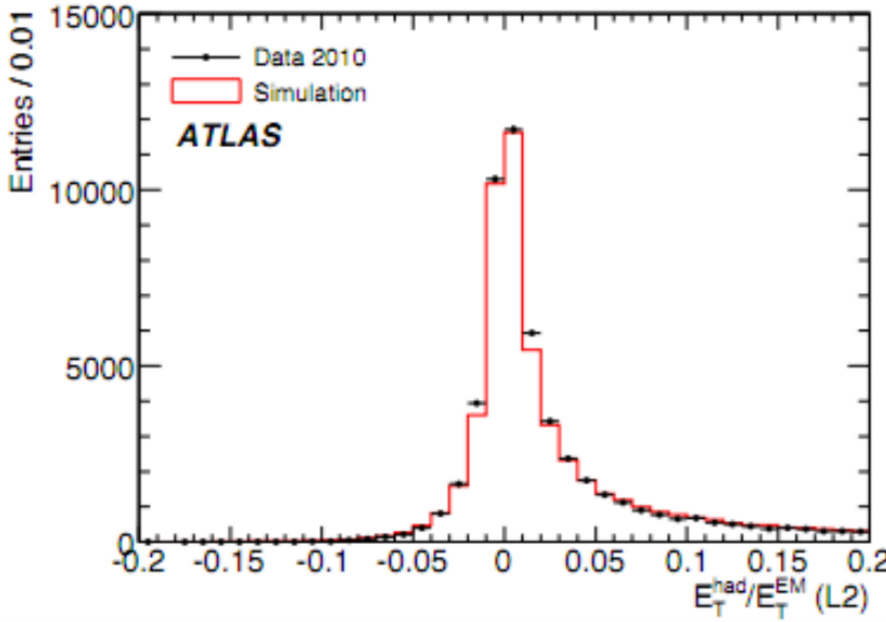
- Algorithm speed enhanced by various checkpoints
  - Opportunity to reject early and save processing time



HLT Electrons/Photons

- **Start from L1 e/γ seed with sufficient ET Reconstruct the cluster in EM Calorimeter**

  - Is there enough energy to continue?
  - Does the cluster shape look like that of an electron/photon?
  - Make sure the cluster is not a hadron (check Hadronic Calorimeter)
  - Is the candidate isolated in the calorimeters?

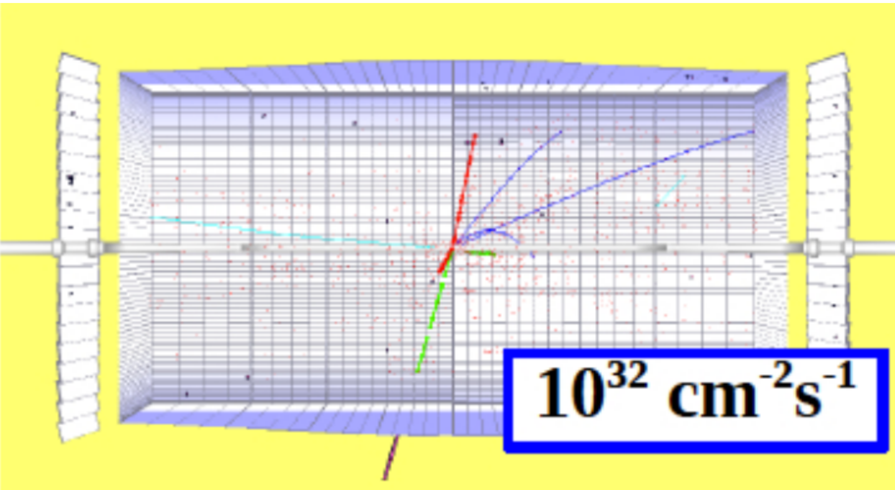


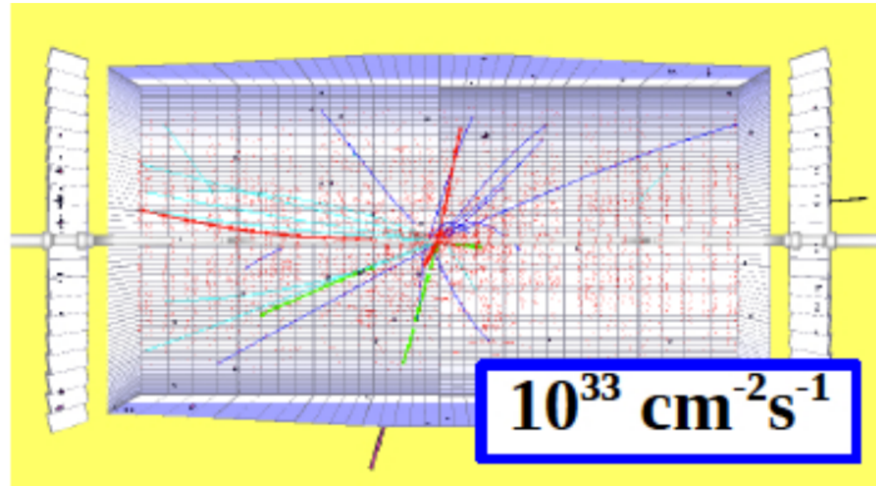
- **Electrons:**
  - Is there a track matched to the cluster?
  - Is the electron isolated in the tracker?

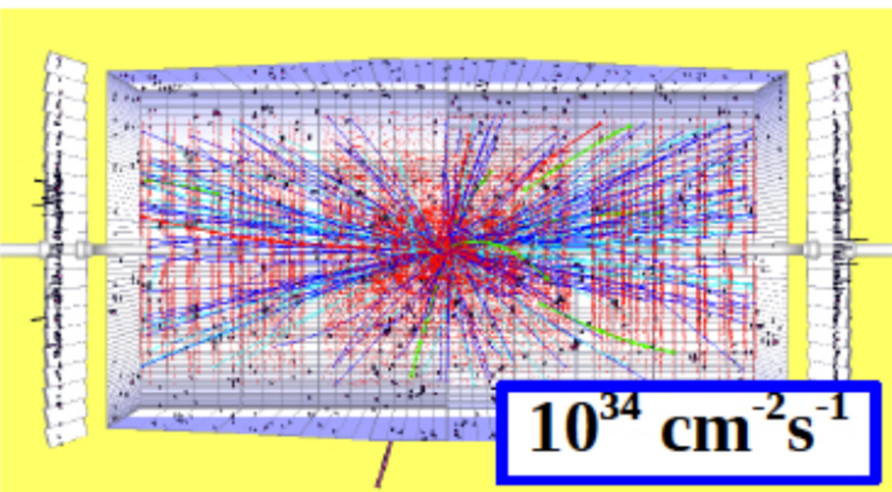
- **Photons**:
  - Check for tracks pointing to the cluster

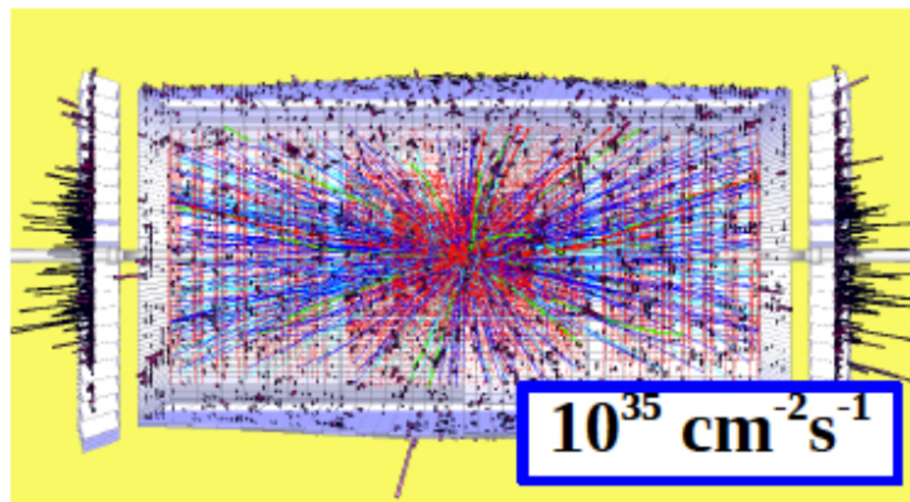
Pileup

- Simulation of 300 GeV H → ZZ → eeμμ



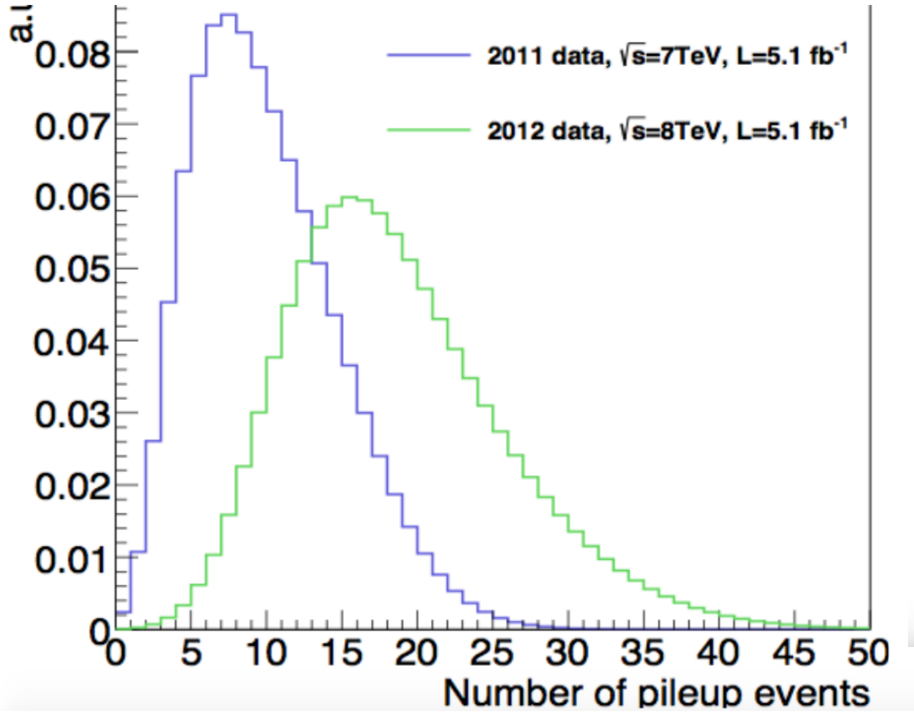


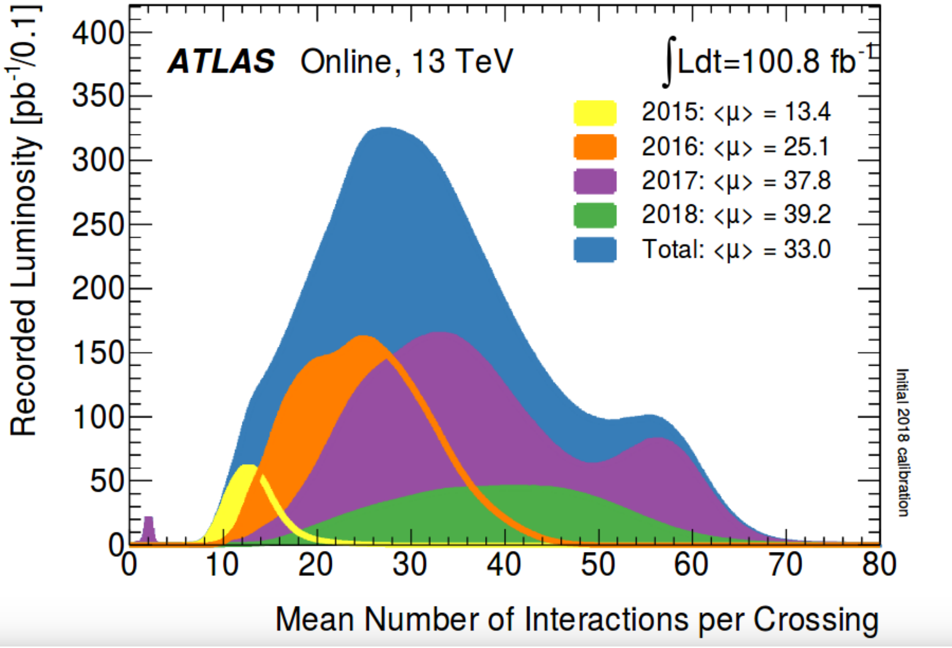




Pileup

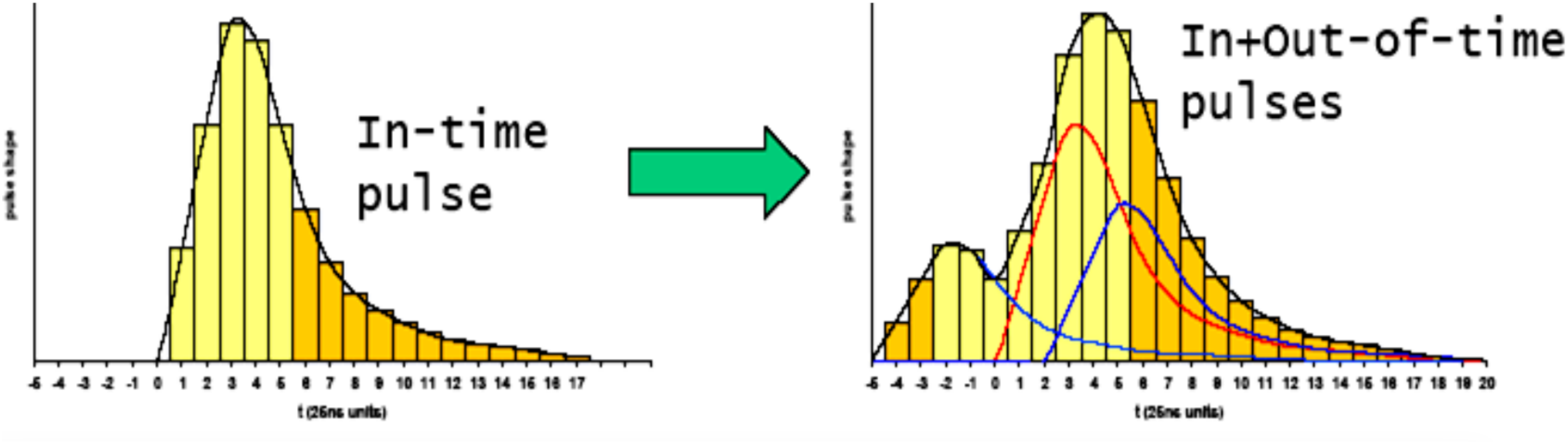
- LHC Design: 20 collisions per crossing
- Today:





L1 at high pileup

- L1 Trigger must cope with high collision rate
  - Tighten trigger requirements to reject extra background
  - Trade-off: Possible loss of signal efficiency
- Multiple collisions per crossing impacts the L1 trigger
- All this was “known” already, as part of the LHC detector design
  - HL-LHC: New challenges



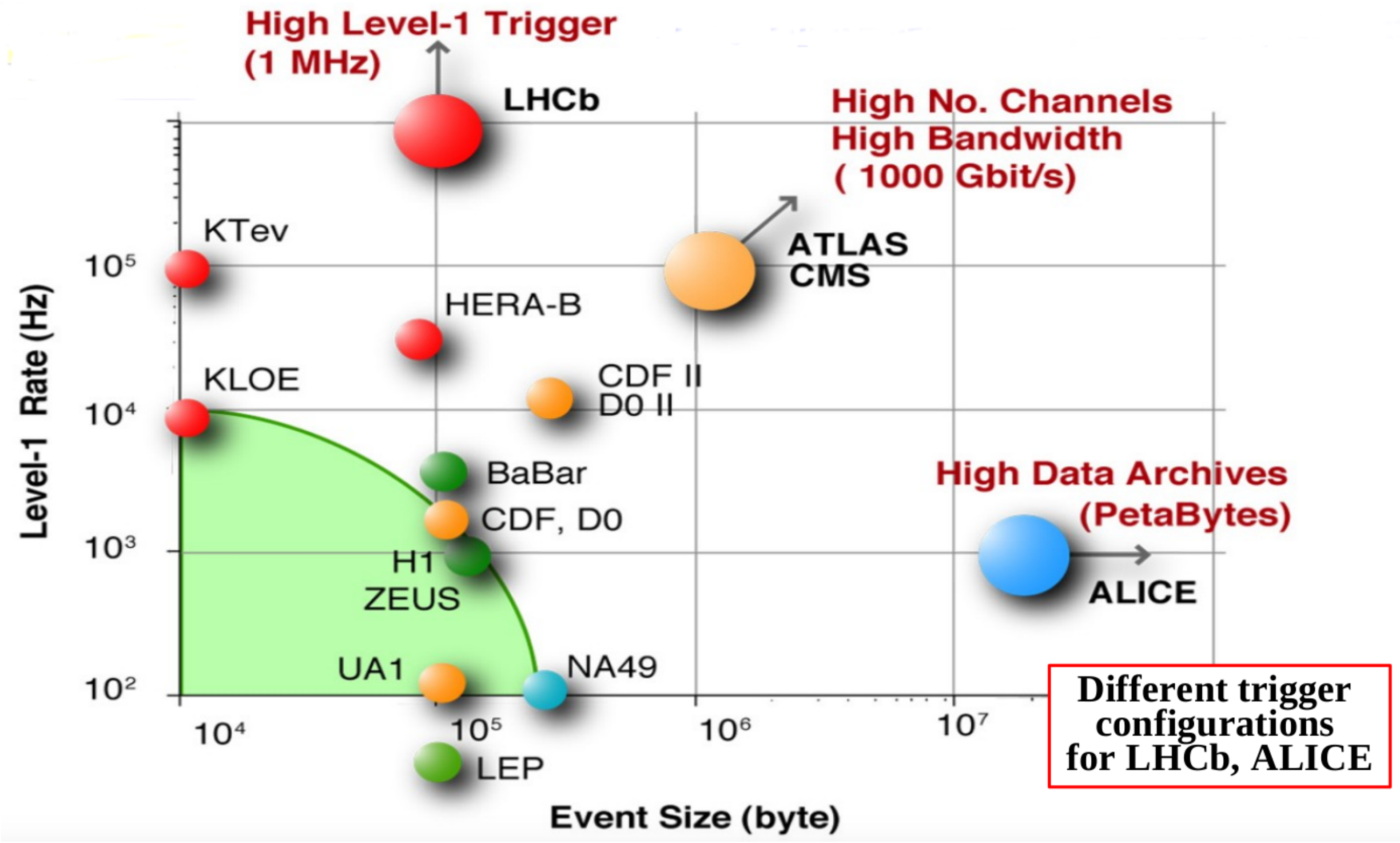
Evolution of Trigger

- **The trigger is by design very flexible:**
  - Should always be able to respond to the present physics demand
  - And demands can change quickly!

(e.g. 2010→ 2012
increase of 1,0000)

Trigger at Experiments

- **LHC experiments have much higher trigger requirements than previous experiments**



Different trigger configurations for LHCb, ALICE

Trigger Interface with Analysis

- **Physicists start with an analysis idea:**
  - Determine what you want to look for (i.e. where you want to go)
  - Then figure out how to select the data

- There is little point in trying to do an analysis if every “interesting” event fails the trigger

- Want to build a trigger that has loose requirements that you tighten up offline

- Design a trigger to meet analysis goals, but … (next page)

Competing

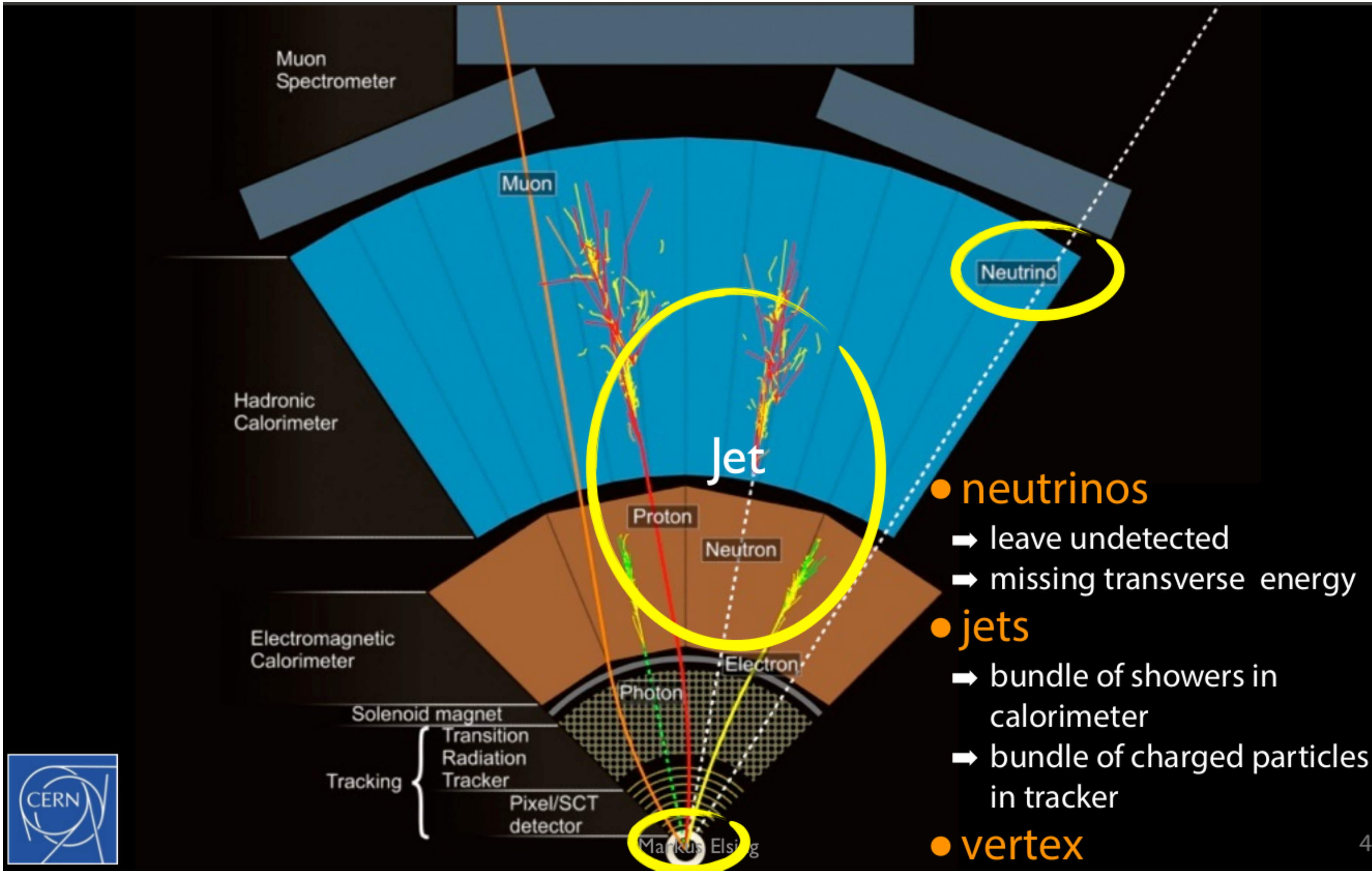
- **There are hundreds to thousands of physicists on a LHC collaboration**
  - All are competing for the same resources
  - Only O(100) Hz of collision data available
    → At L = 10^34 , this is roughly the rate of $W \to l\nu$ production!

- **How do you make sure your (very important) data is kept for later analysis?**
  - Need to meet physics needs with limited bandwidth

- **Cutting at the trigger level throws away data forever**
  - Potential bias to events that you analyze
  - Loss of interesting data

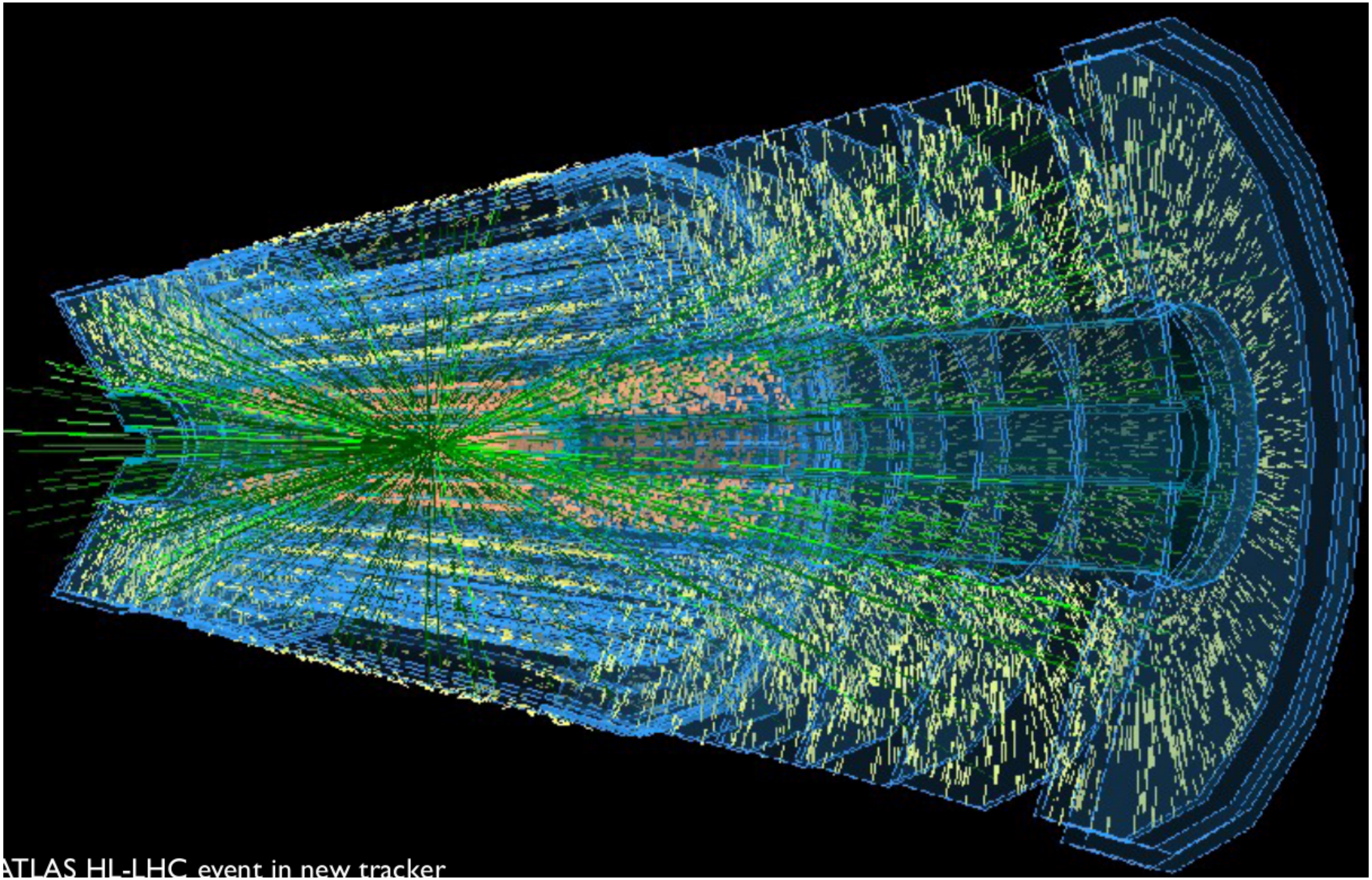
“The Trigger does not determine which Physics Model is right, only which Physics Model is left”

Move to Reconstruction



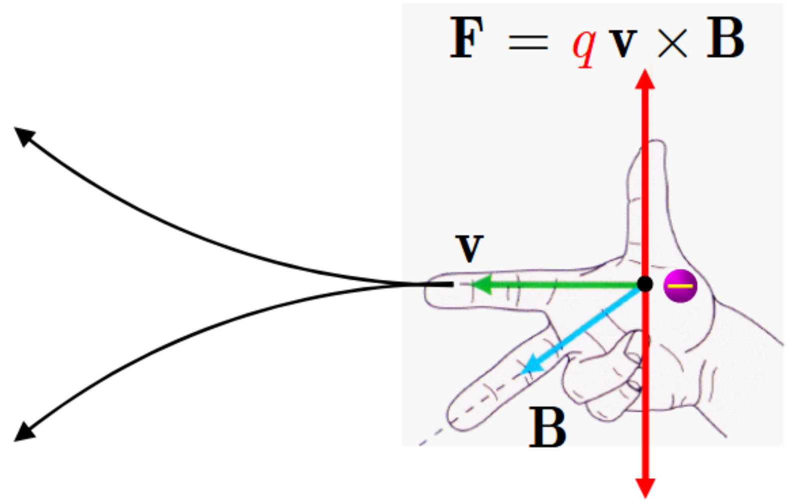
- neutrinos
  - ➡ leave undetected
  - ➡ missing transverse energy
- jets
  - ➡ bundle of showers in calorimeter
  - ➡ bundle of charged particles in tracker
- vertex

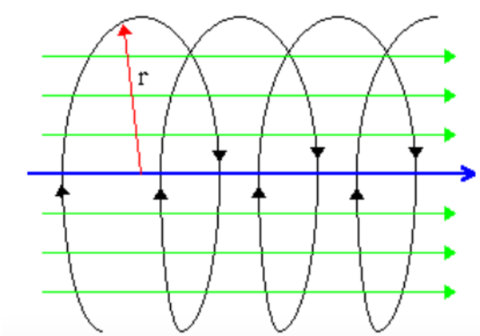
Tracking



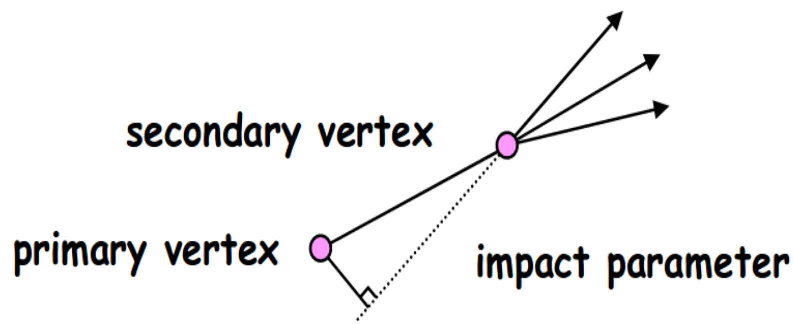
Tracking

- Tracking is concerned with the reconstruction of charged particles trajectory (tracks)
- in experimental particle physics the aim is to measure (not a full list):





$$p = 0.3 \cdot B \cdot R$$

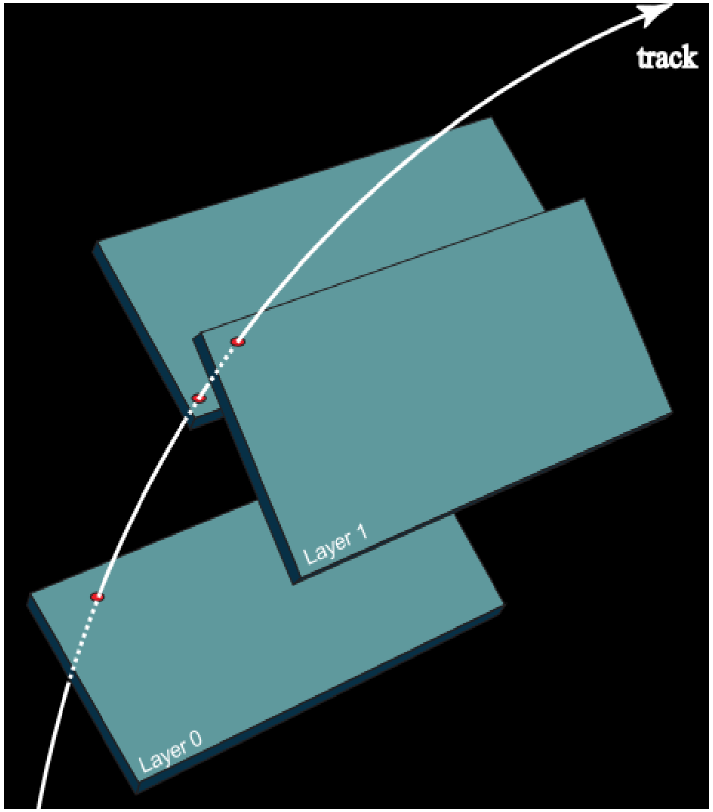


Track Fitting

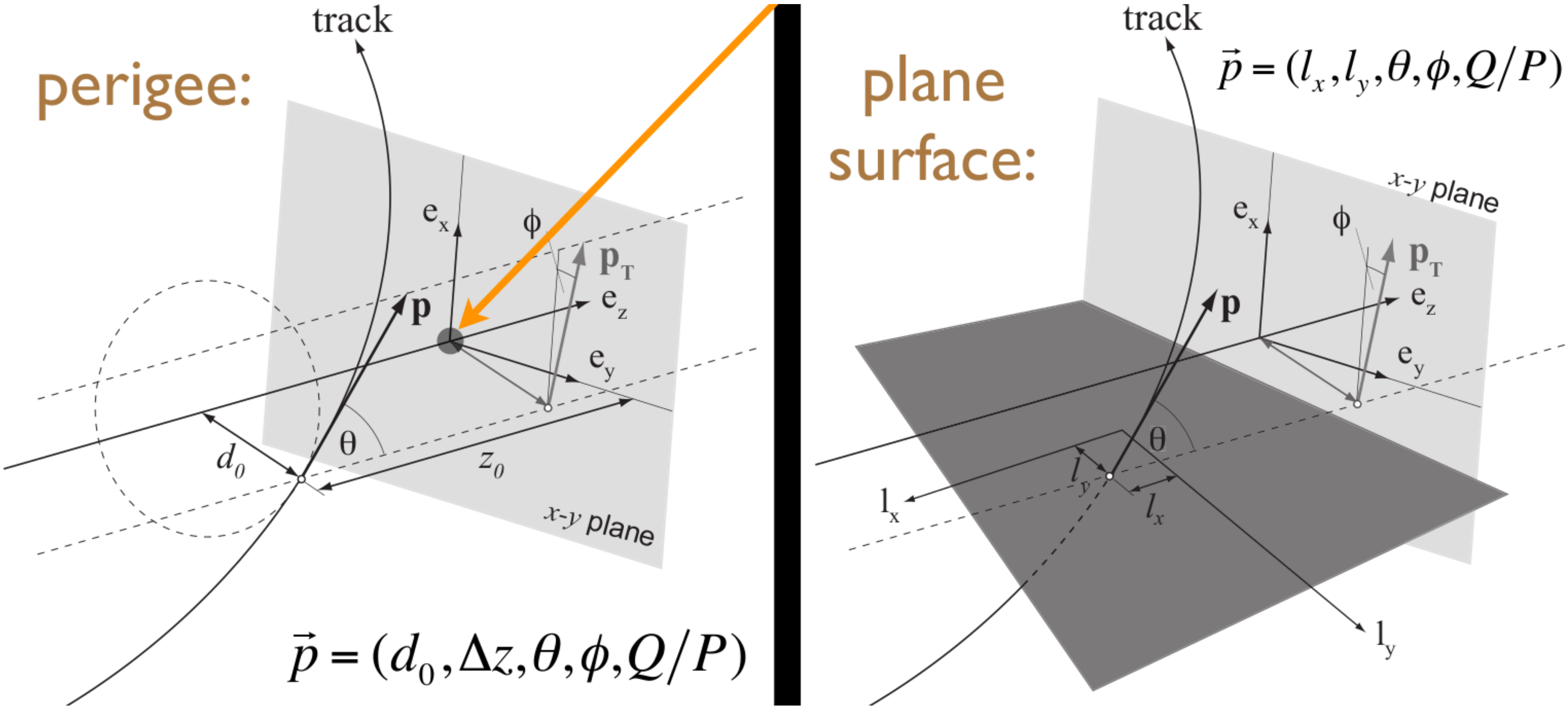
A Trajectory of a Charged Particle

- **in a solenoid B-field a charged particle trajectory is describing a helix**
  - a circle in the plane perpendicular to the field (Rφ)
  - a path (not a line) at constant polar angle(θ) in the R-Z plane

→ a trajectory in space is defined by 5 parameters
→ the local position (l1 , l2) on a plane, a cylinder, on the surface or reference system
→ the direction in θ and φ plus the Curvature q/pT



The Perigee Parametrization

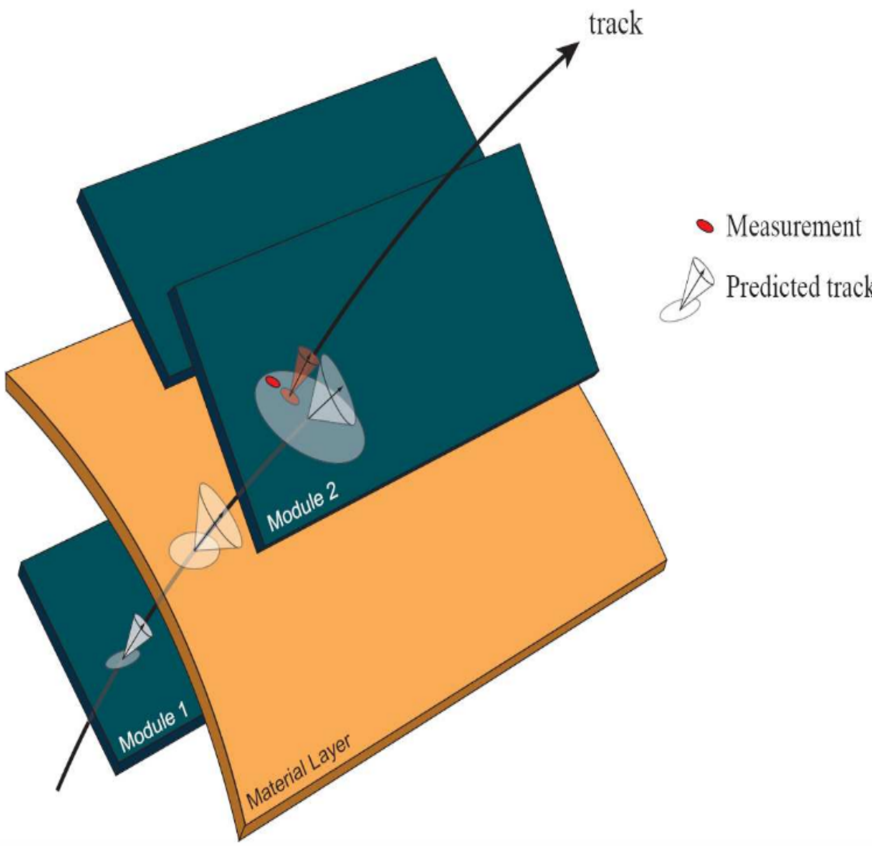


To express the track parameters near the production vertex or on plane surface

Following the Particle Trajectory

- **basic problems** to be solved in order to follow a track through a detector:

➡ next detector module that it intersects?
➡ what are its parameters on this surface? what is the uncertainty of those parameters ?
➡ for how much material do I have to correct for ?



- **requires:**

➡ a detector geometry track surfaces for active detectors passive material layers
➡ a method to discover which is the next surface (navigation)
➡ a propagator to calculate the new parameters and its errors often referred to as “track model”

Tracking

- Almost all High Energy experiments done at accelerators have a magnetic spectrometer to measure the momentum of charged particles

The equation of motion for a particle with charge q in magnetic field B:

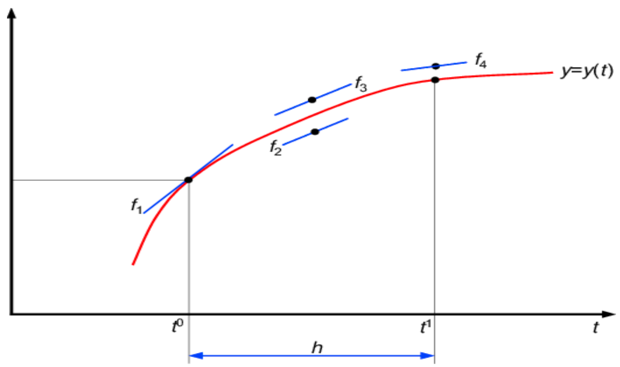
$$\frac{d\vec{p}}{dt} = q\vec{v} \times \vec{B}$$

Can be written as set of differential equations for motion along z with x(z)

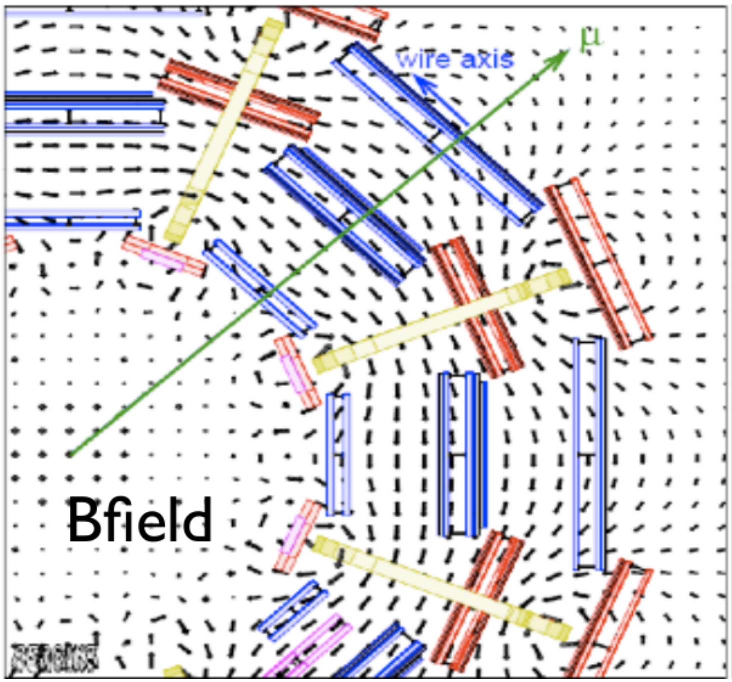
$$\frac{d^2x}{dz^2} = \frac{q}{p}R\left[\frac{dx}{dz}\frac{dy}{dz}B_x - \left(1 + \left(\frac{dx}{dz}\right)^2\right)B_y + \frac{dy}{dz}B_z\right]$$

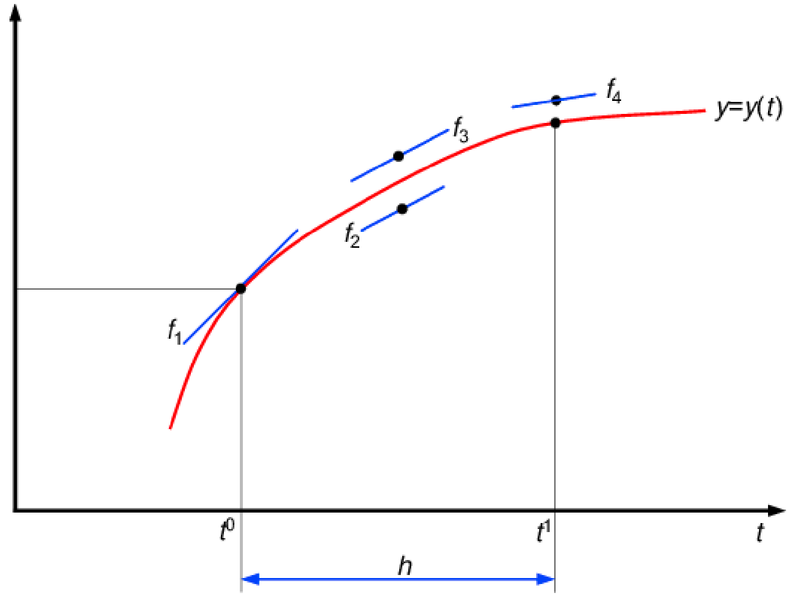
$$\frac{d^2y}{dz^2} = \frac{q}{p}R\left[\left(1 + \left(\frac{dy}{dz}\right)^2\right)B_x - \frac{dx}{dz}\frac{dy}{dz}B_y - \frac{dx}{dz}B_z\right]$$

- No analytical solution for inhomogeneous B-field, requires numerical integration
- numerical integration done using Runge-Kutta technique



Track Propagation in realistic B-Field

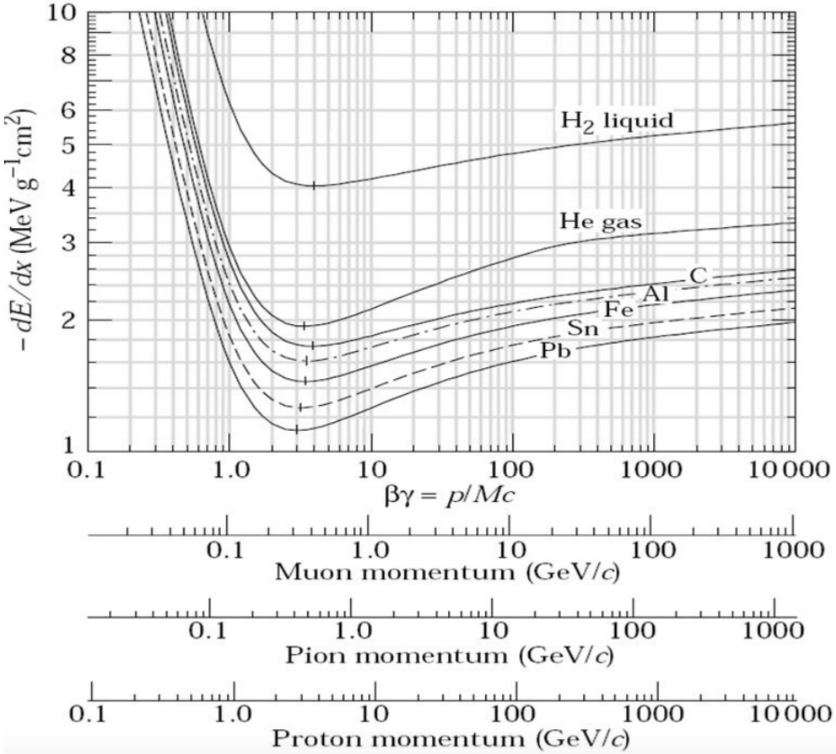


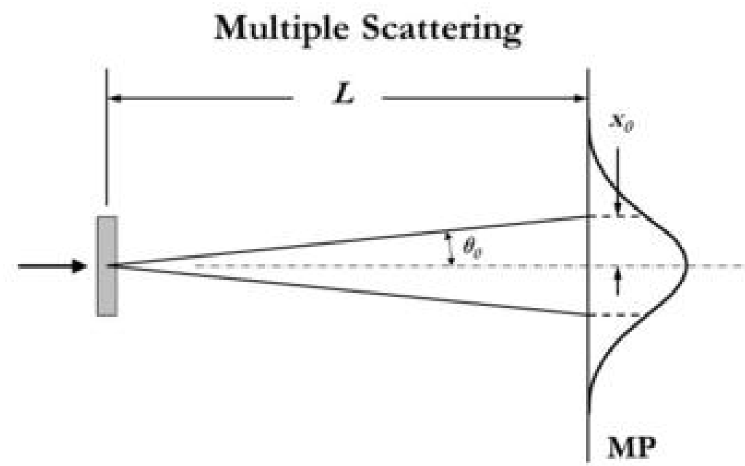


$$
\begin{aligned}
k_1 &= h f(x_n, y_n) \\
k_2 &= h f(x_n + h/2, y_n + k_1/2) \\
k_3 &= h f(x_n + h/2, y_n + k_2/2) \\
k_4 &= h f(x_n + h, y_n + k_3) \\
y_{n+1} &= y_n + \frac{k_1}{6} + \frac{k_2}{3} + \frac{k_3}{3} + \frac{k_4}{6}
\end{aligned}
$$

Well, not only B-field

- Energy loss
  - impact on the momentum
- Multi-scattering
  - increases uncertainty on direction of track





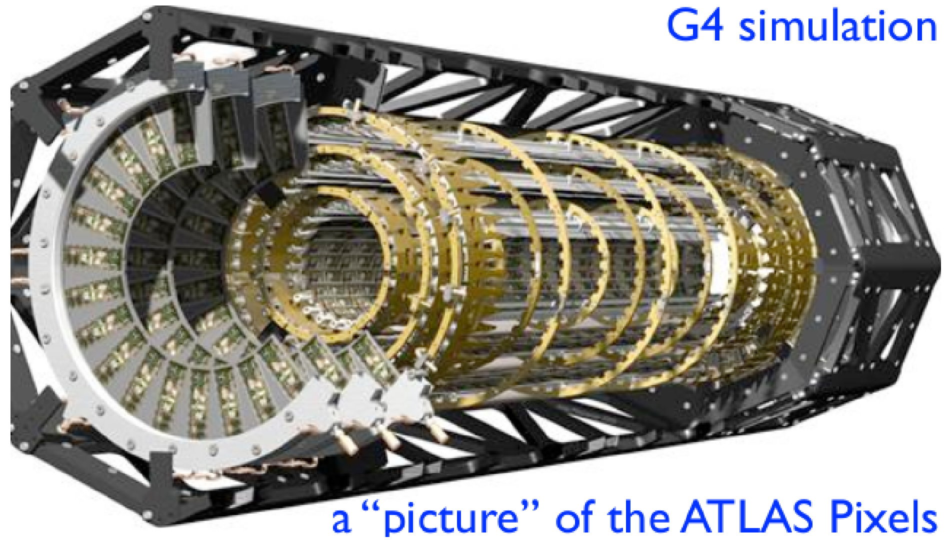
When protons pass through a slab of material they suffer millions of collisions with atomic nuclei. The statistical outcome is a *multiple scattering angle* whose distribution is approximately Gaussian. For protons, this angle is always small so the projected displacement in any measuring plane **MP** is also Gaussian. The width parameter of the angular distribution is θ_0. The corresponding displacement, x_0, can easily be measured by scanning a dosimeter across the MP. The task of multiple scattering theory is to predict θ_0 given the scattering material and thickness, and the incident proton energy.

Detector Geometry

- interactions in detector material limiting tracking performance:

➡ LHC detectors are complex require a very detailed description of their geometry
➡ experiments developed geometry models (translation into G4 Simulation) huge number of volumes

- physics requirement to reach LHC goals (e.g. W mass)
  - ➡ control material close to beam pipe at % level



a "picture" of the ATLAS Pixels

	model	placed volumes
ALICE	Root	4.3 M
ATLAS	GeoModel	4.8 M
CMS	DDD	2.7 M
LHCb	LHCb Det.Des.	18.5 M

Weighing Detectors during Construction

- Huge effort in experiments:
  - ➡ important to reach good description in simulation and reconstruction
  - ➡ each individual detector part was put on balance and compare with model
  - ➡ CMS and ATLAS measured weight of their tracker and all of its components
  - ➡ correct the geometry implementation in simulation and reconstruction

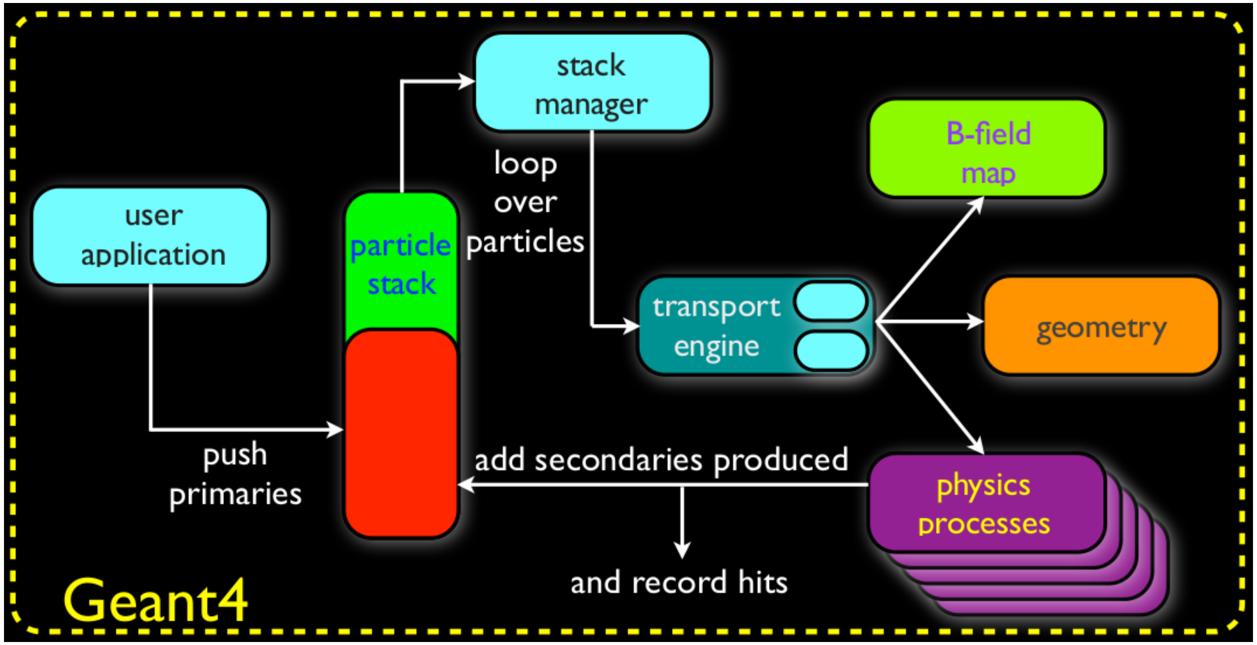
CMS	estimated from measurements	simulation
active Pixels	2598 g	2455 g
full detector	6350 kg	6173 kg

ATLAS	estimated from measurements	simulation
Pixel package	201 kg	197 kg
SCT detector	672 ±15 kg	672 kg
TRT detector	2961 ±14 kg	2962 kg

Simulation (Geant4)

- Geant4 is based upon:
  - stack to keep track of all particles produced and stack manager
  - extrapolation system to propagate each particle
    - → transport engine with navigation
    - → geometry model
    - → B-field
  - set of physics processes describing interaction of particles with matter
  - a user application interface

Same concepts as track reconstruction



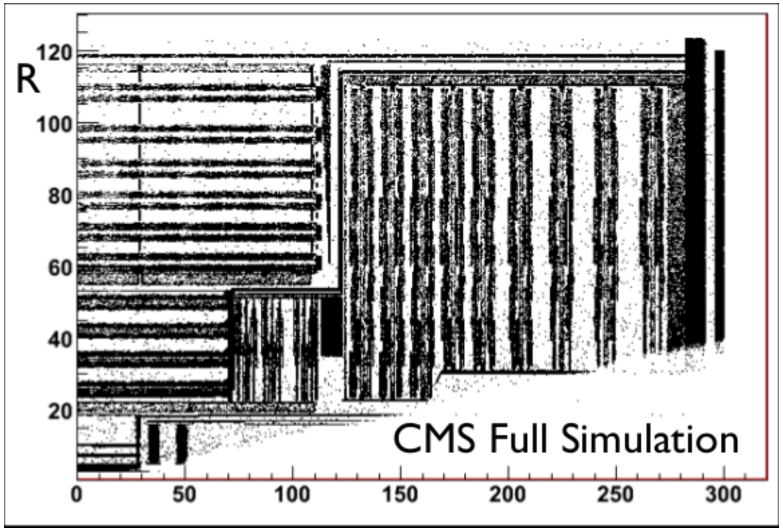
Fast Simulation

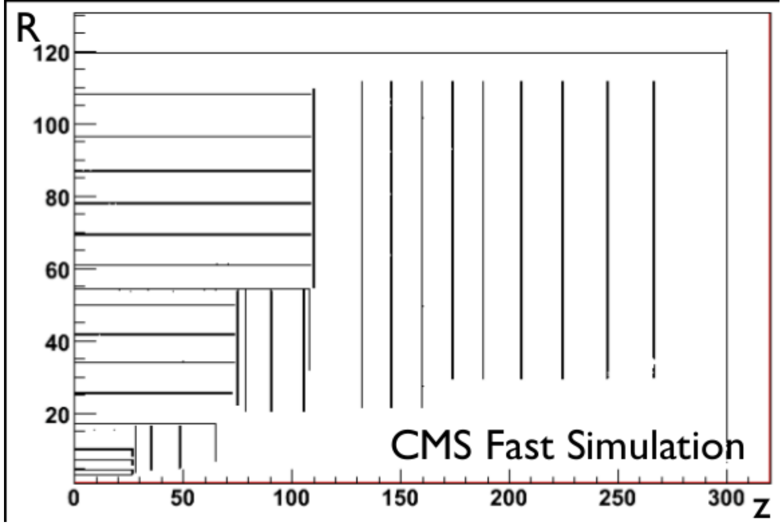
- CPU needs for Geant4:
  ➡ simulation strategies of experiments mix full G4 and fast simulation

	G4	fast sim.
CMS	360	0.8
ATLAS	1990	7.4

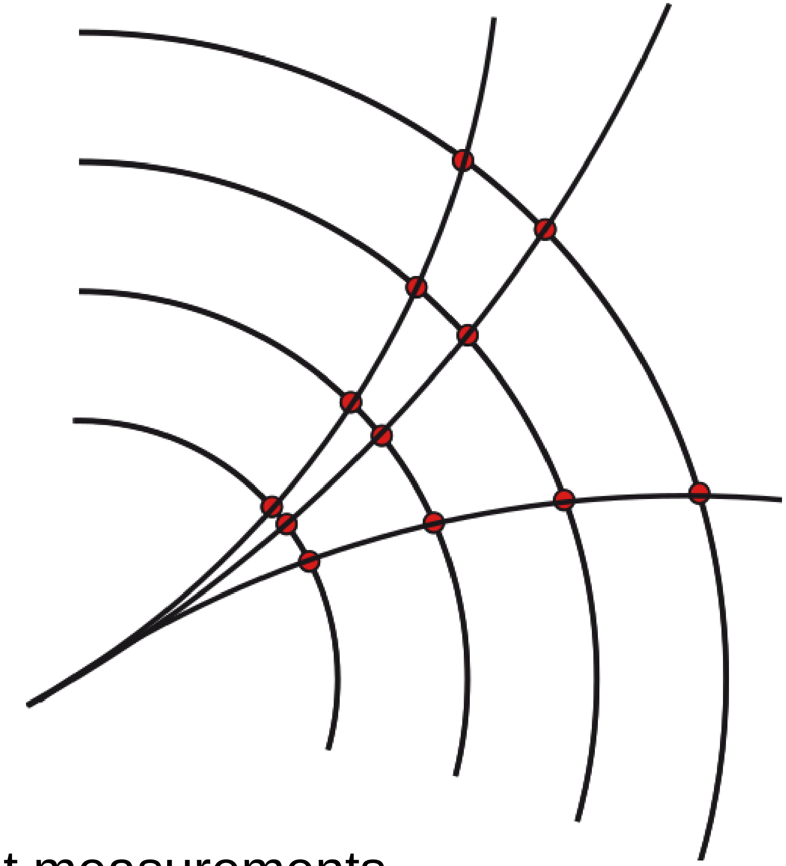
(ttbar events, in seconds)

- fast simulation engines:
➡ fast calo. simulation (parameterisation, showers libraries, ...)
➡ simplified tracking geometries
➡ simplify physics processes w.r.t. G4
➡ output in same data model as full sim.
➡ able to run full reconstruction (trigger)





From Measurements to Track Fitting



- A measurement model is like:

$$\boldsymbol{m}_k = \boldsymbol{h}_k(\boldsymbol{q}_k) + \boldsymbol{\gamma}_k$$

with: $\boldsymbol{h}_k$ ~ functional dependency of measurement on e.g. track angle

$\boldsymbol{\gamma}_k$ ~ error (noise term)

$$\boldsymbol{H}_k = \frac{\partial \boldsymbol{m}_k}{\partial \boldsymbol{q}_k}$$ ~ Jacobian, often contains only rotations and projections

Measurements mk. In practice those mk are clusters, drift circles …

- Task of track fit:
  - estimate the track parameters from a set measurements
- Examples of fitting techniques:
  - Least Square; Kalman Filter
  - Gaussian Sum Filter or Deterministic Annealing Filters

Classical Least Square Track Fit

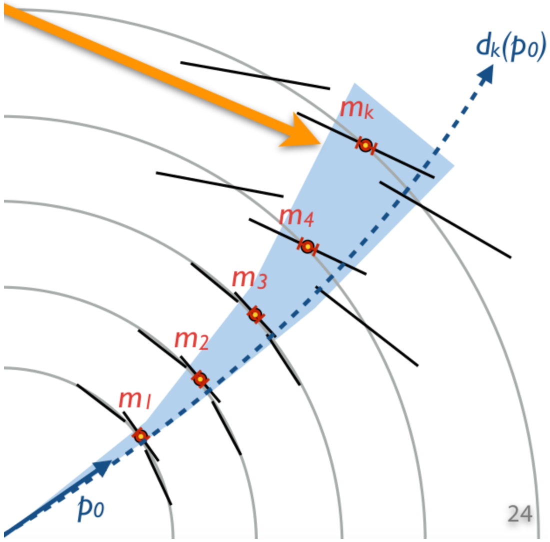
Construct and minimise the χ^2 function:

- Carl Gauss is credited with developing the fundamentals of the basis for least-squares analysis in 1795 at the age of eighteen

- Legendre was the first to publish the method, however

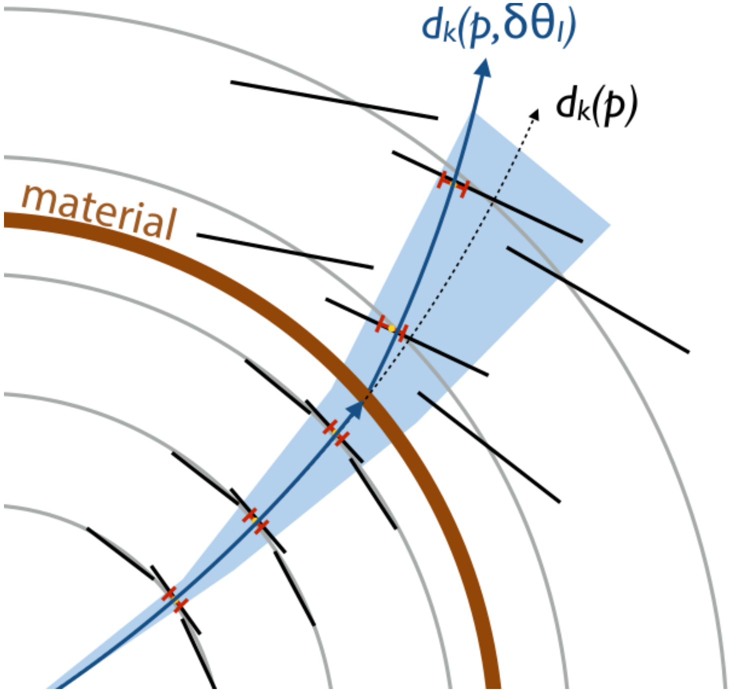
➡Write down Least Square function:

$$\chi^2 = \sum_k \Delta m_k^T G_K^{-1} \Delta m_k \quad \text{with:} \quad \Delta m_k = m_k - d_k(p)$$



Classical Least Square Track Fit

- Allowing for material effects in fit:
  - ➡ can be absorbed in track model, provided effects are small
  - ➡ for substantial multiple scatting, allows for scattering angles in the fit



- Introduce scattering angles on material surfaces:
  - ➡ on each material surface, add 2 angles δθ i as fee parameters to the fit

- Results in additional term in χ2 equations:

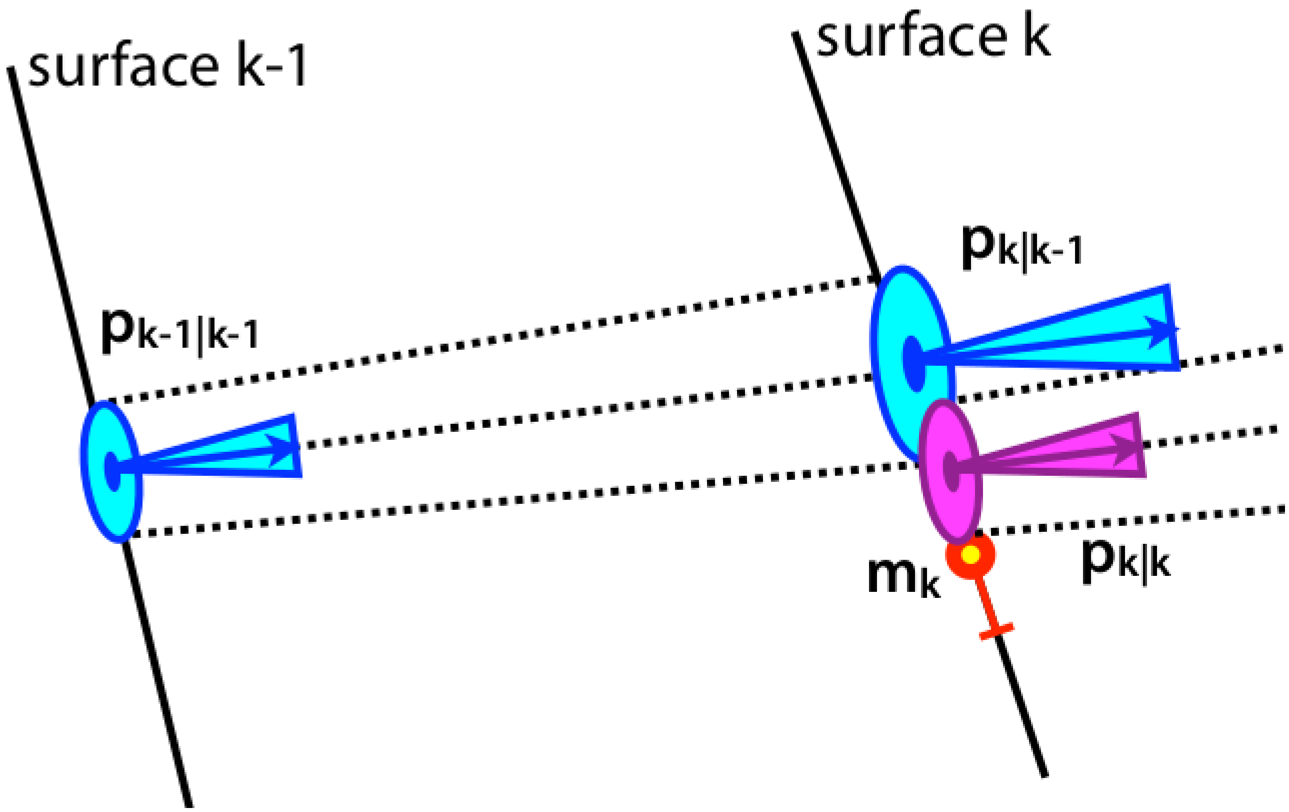
$$\chi^2 = \sum_k \Delta m_k^T G_K^{-1} \Delta m_k + \sum_i \delta\theta_i^T Q_i^{-1} \delta\theta_i$$

with: $\Delta m_k = m_k - d_k\left(p, \delta\theta_i\right)$

The Kalman Filter Track Fit

A Kalman Filter is a progressive way of performing a least square fit

➡ can be shown that it is mathematically equivalent

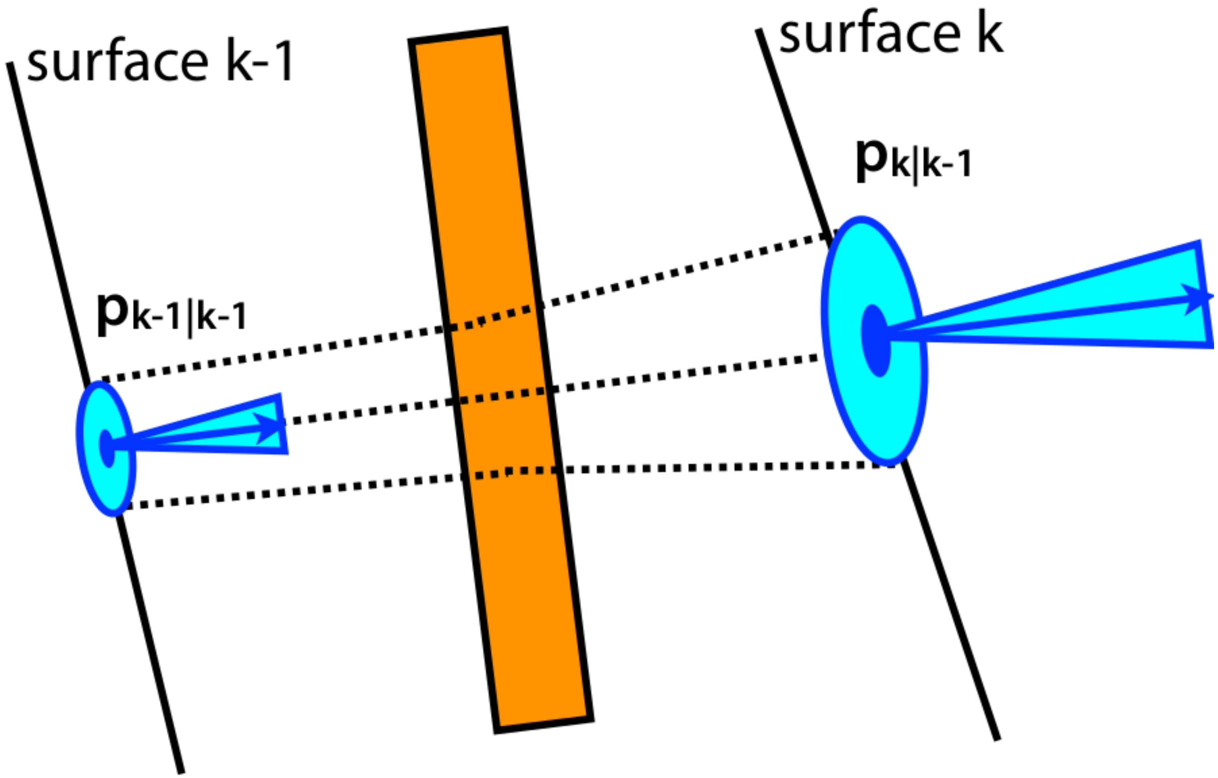


The Kalman Filter Track Fit

Material effects (multiple scattering and energy loss):

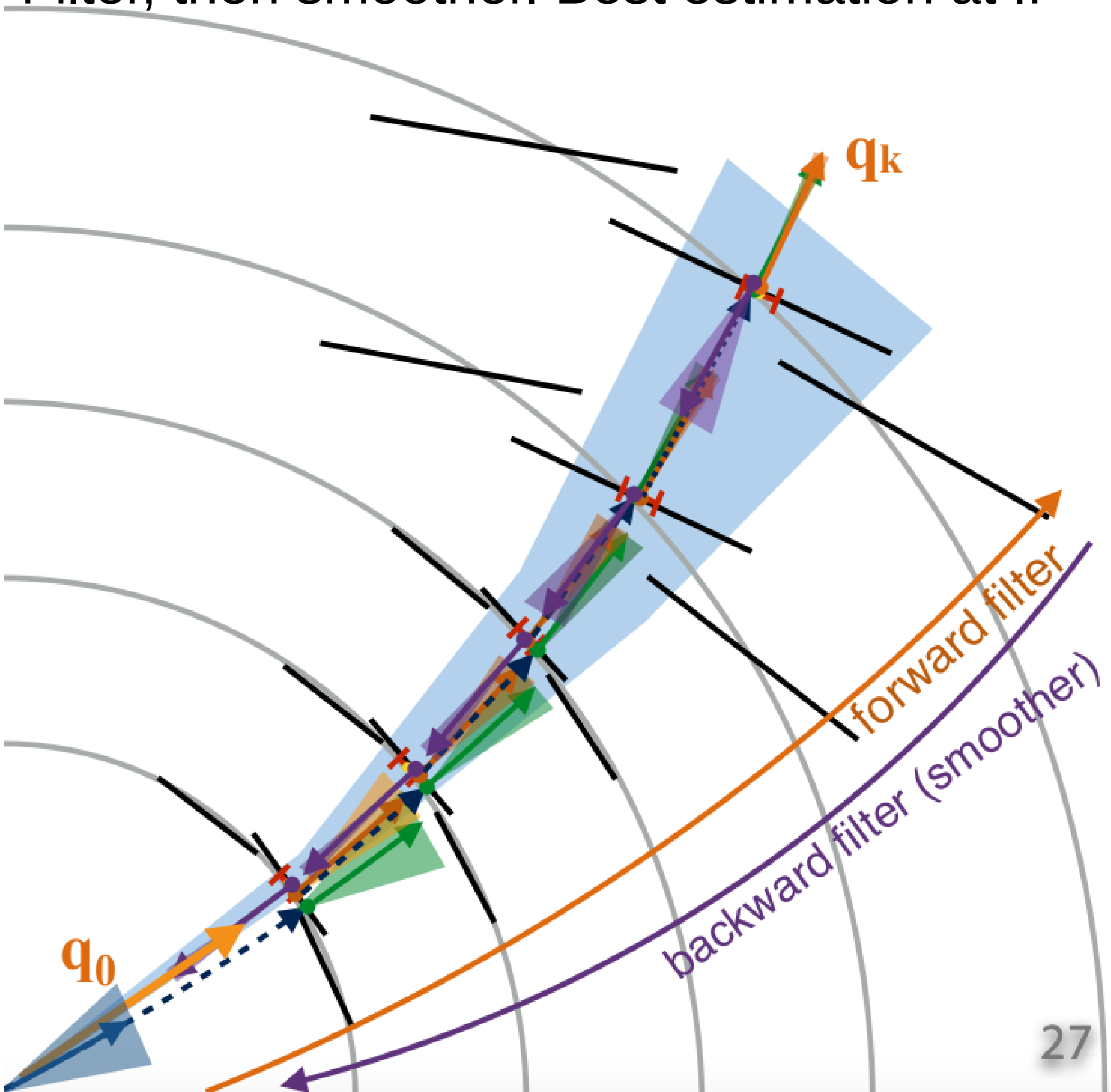
➡ incorporated in the propagated parameters p k|k-1 (extrapolated prediction)

➡ and therefore enters automatically in the updated parameters p k|k at point k



Filter and Smoother

- Initial parameters could be a bit arbitrary
- Filter, then smoother. Best estimation at IP



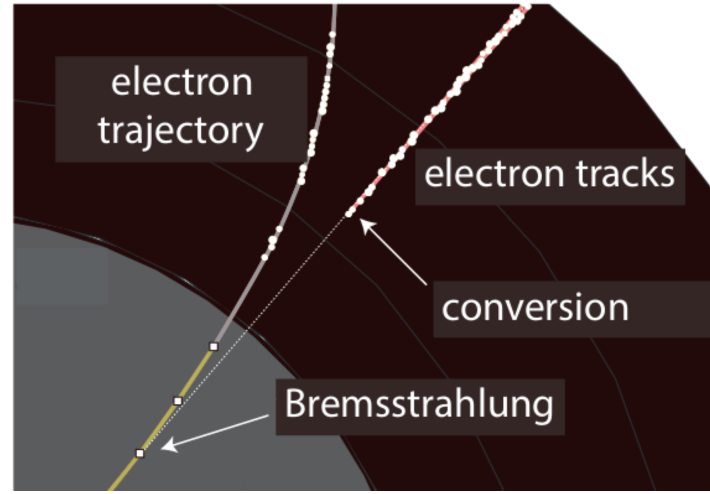
Fitting for Electron

material in tracker:

➡ e-Bremsstrahlung and γ-conversions

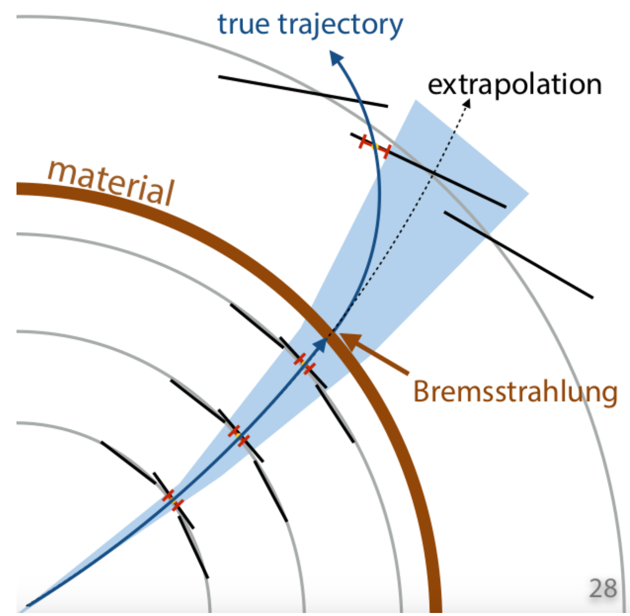
Electron efficiency limited:

momentum loss due to Bremsstrahlung leads to sudden large changes in track curvature
➡ loosing hits after Brem. leads to inefficiency
➡ fit either biased towards small momenta or fails completely because of bad $\chi 2$



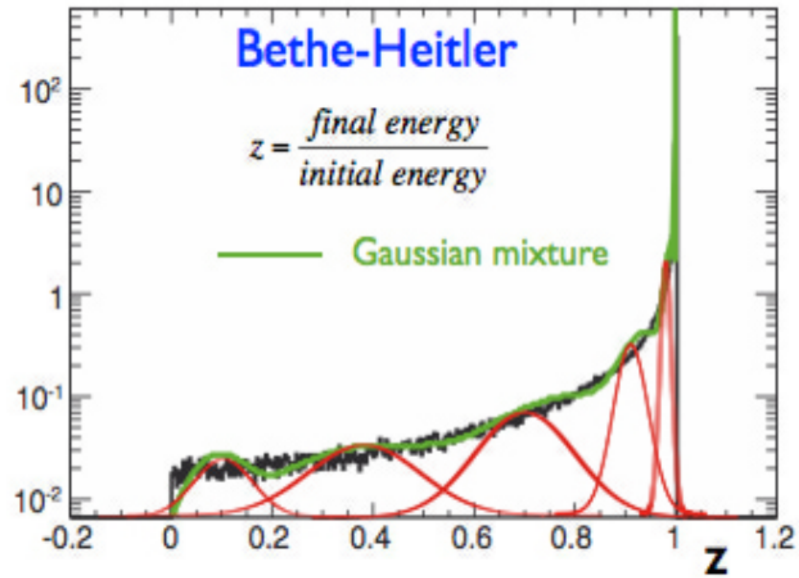
Techniques to allow for Bremsstrahlung in track fitting:

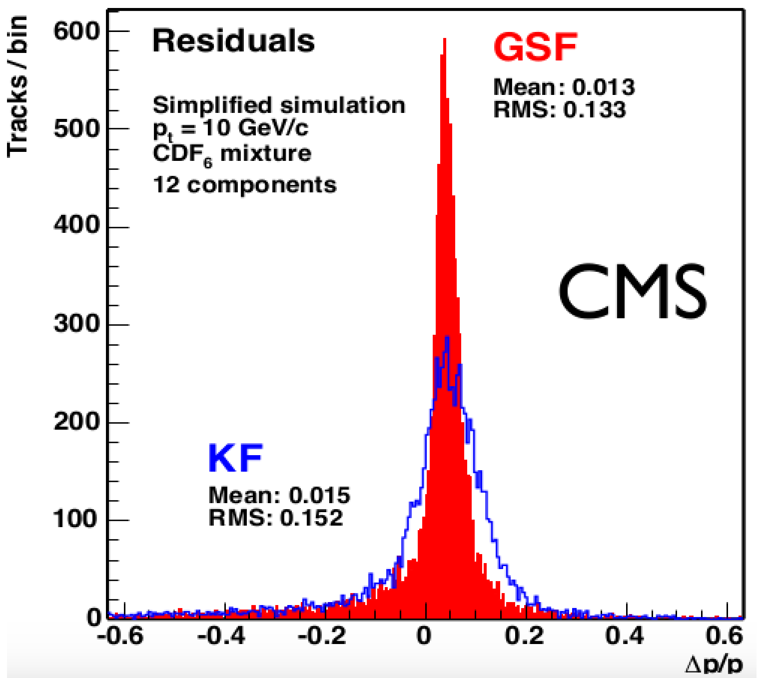
➡ for Least Square track fit
allow Brem. effect to change curvature, additional term similar is to scattering angle
➡ for Kalman Filter
increase correction for material effects in propagation to allow for Brem.
➡ better: Gaussian Sum Filter

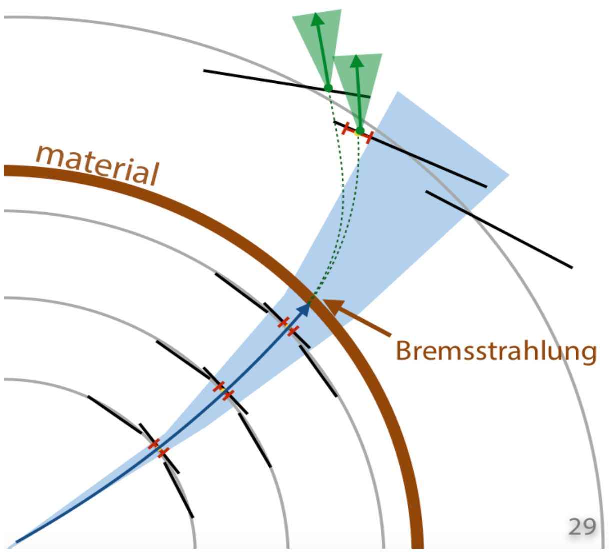


The GSF

- Approximate Bethe-Heitler distribution as Gaussian mixture
- GSF step resembles set of parallel Kalman Filters computationally expensive !

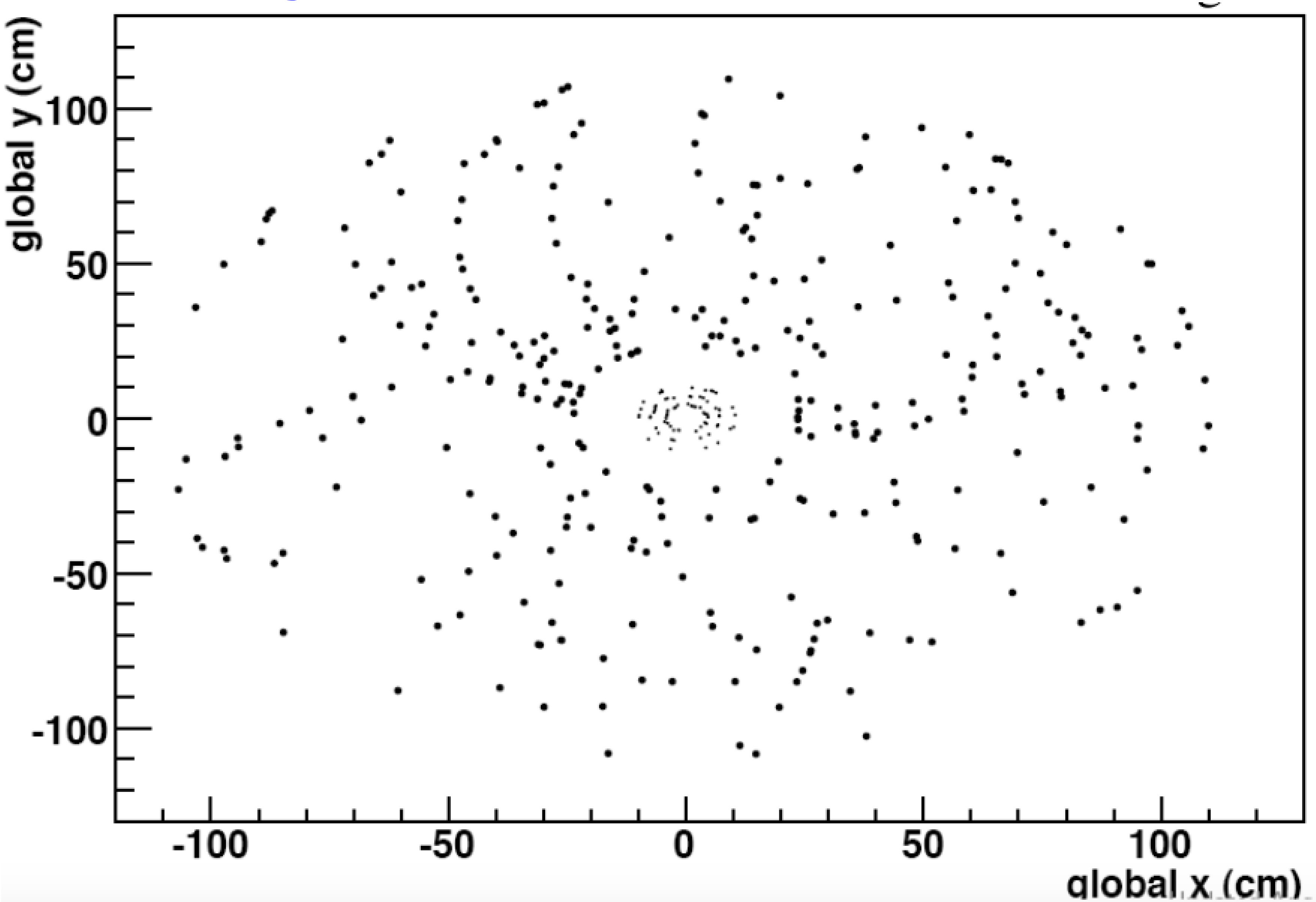




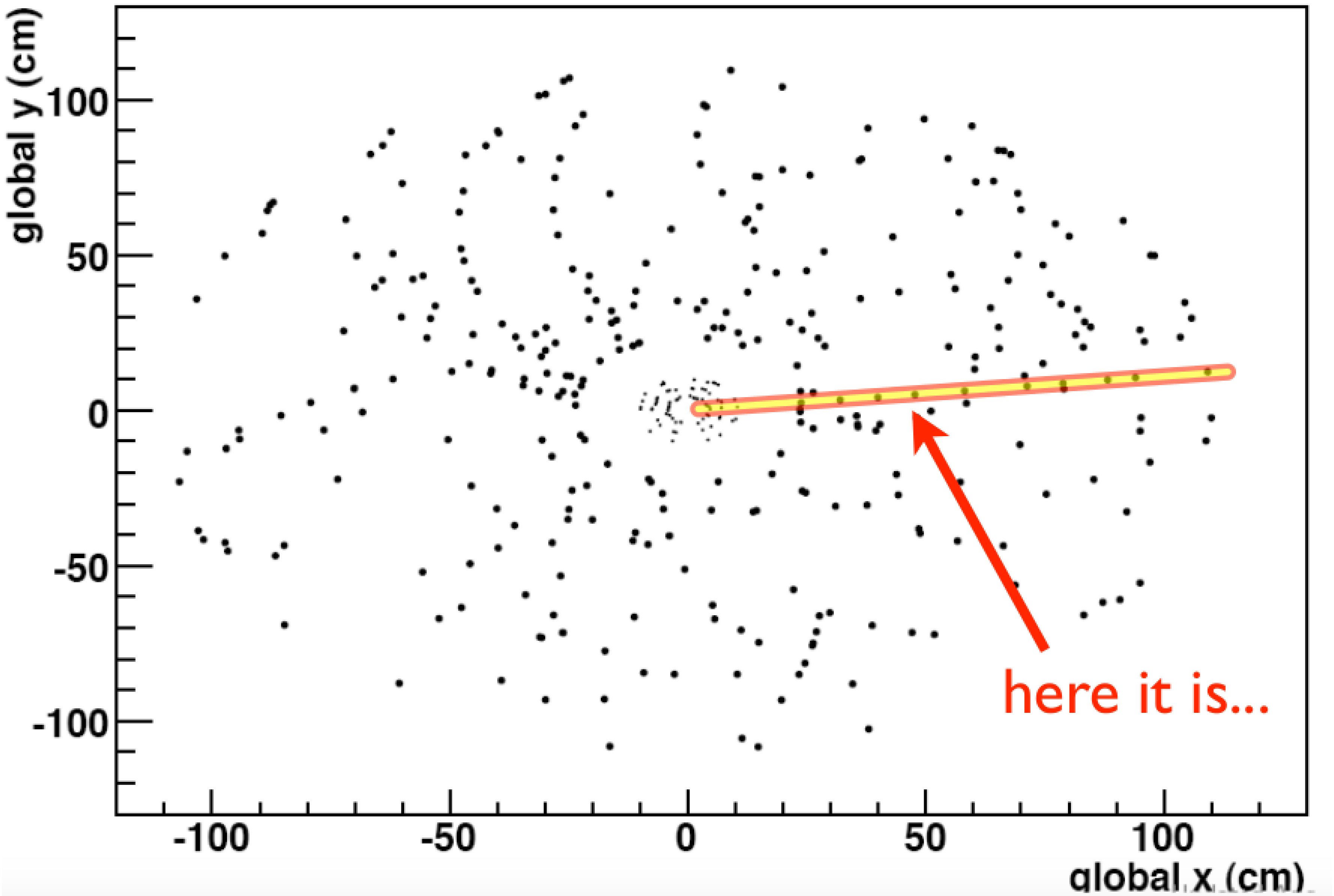


Track Finding

Can you find a 50GeV track ??



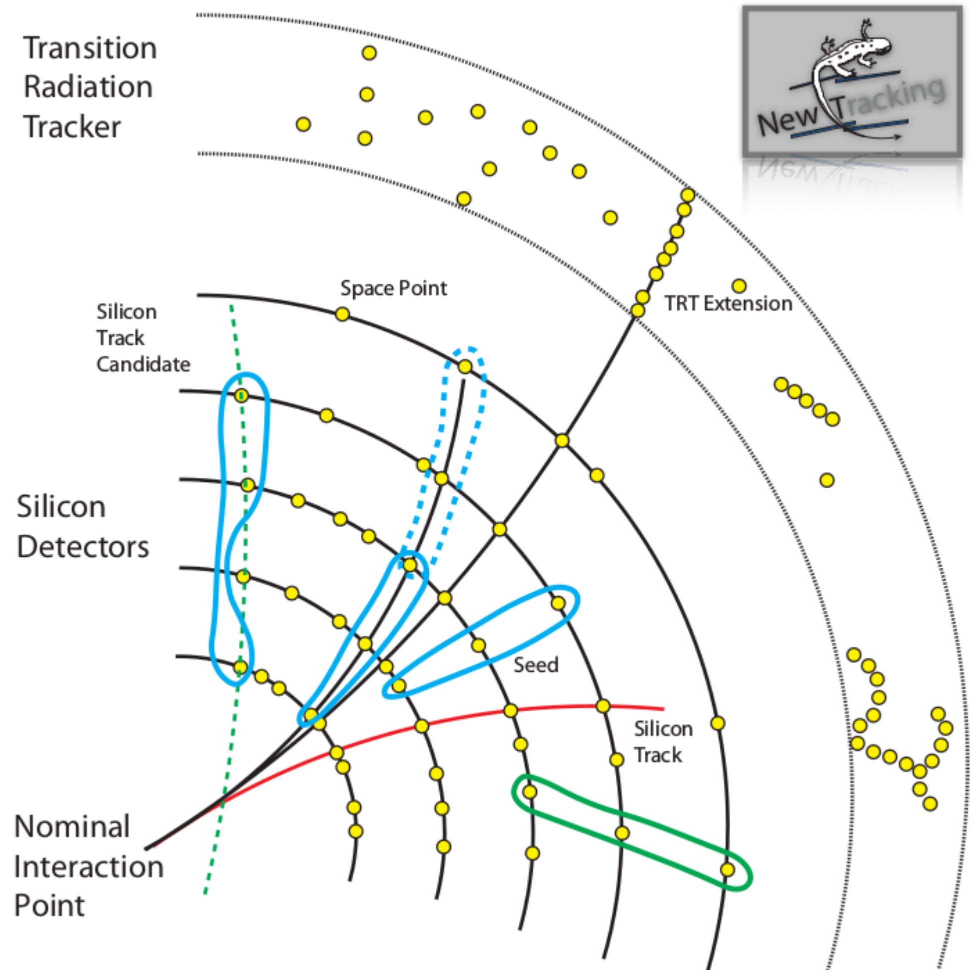
Can you find a 50GeV track ??



Track Finding

The task of the track finding
➡ identify track candidates in event
➡ cope with the combinatorial explosion of possible hit

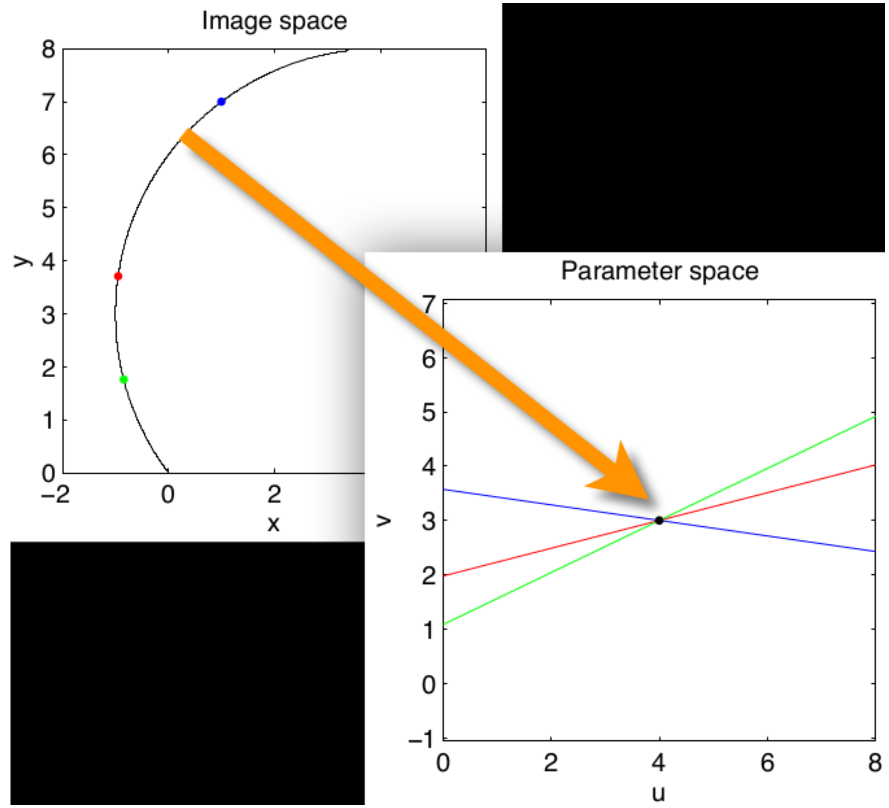
Different techniques:
➡ rough distinction: local/sequential and global/parallel methods
➡ local method: generate seeds and complete them to track candidates
➡ global method: simultaneous clustering of detector hits into track candidates



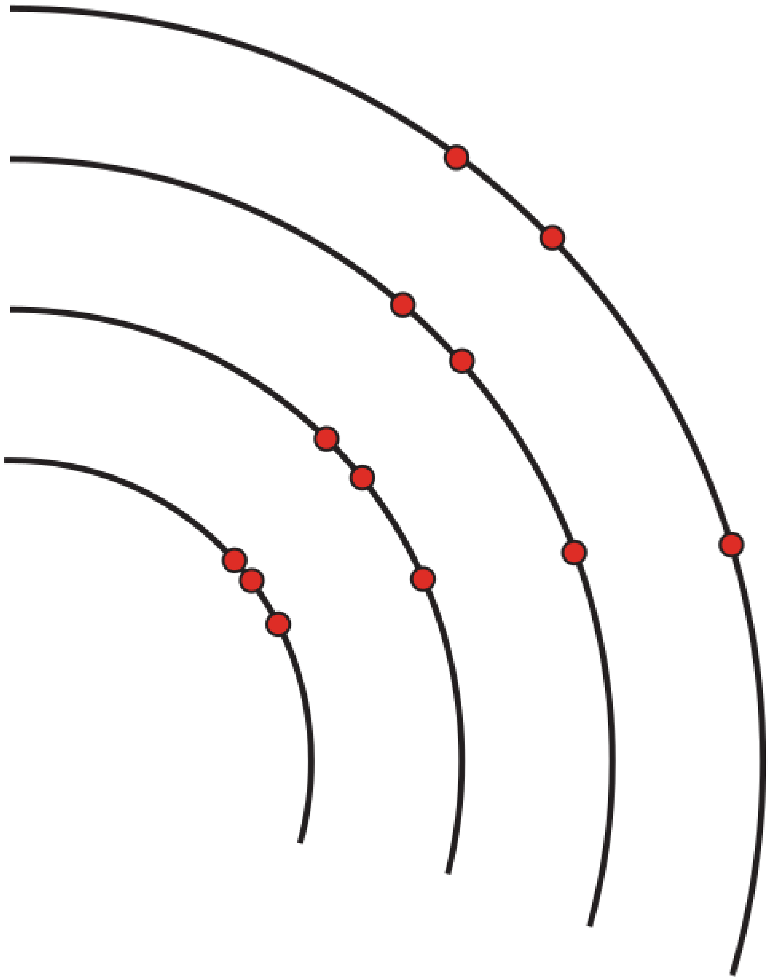
Conformal Mapping

- Hough transform:
  - cycles through the origin in x-y transform into point in u-v
  - each hit becomes a straight line

- Search for maxima in parameter space to find track candidates

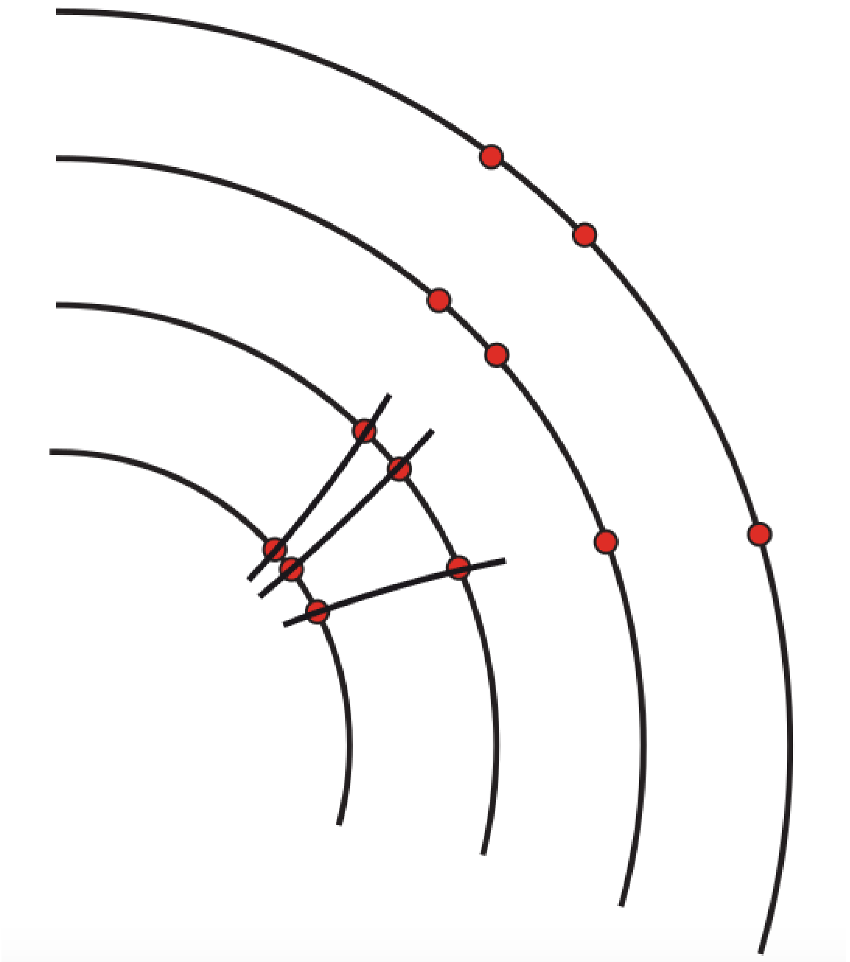


Local Track Finding



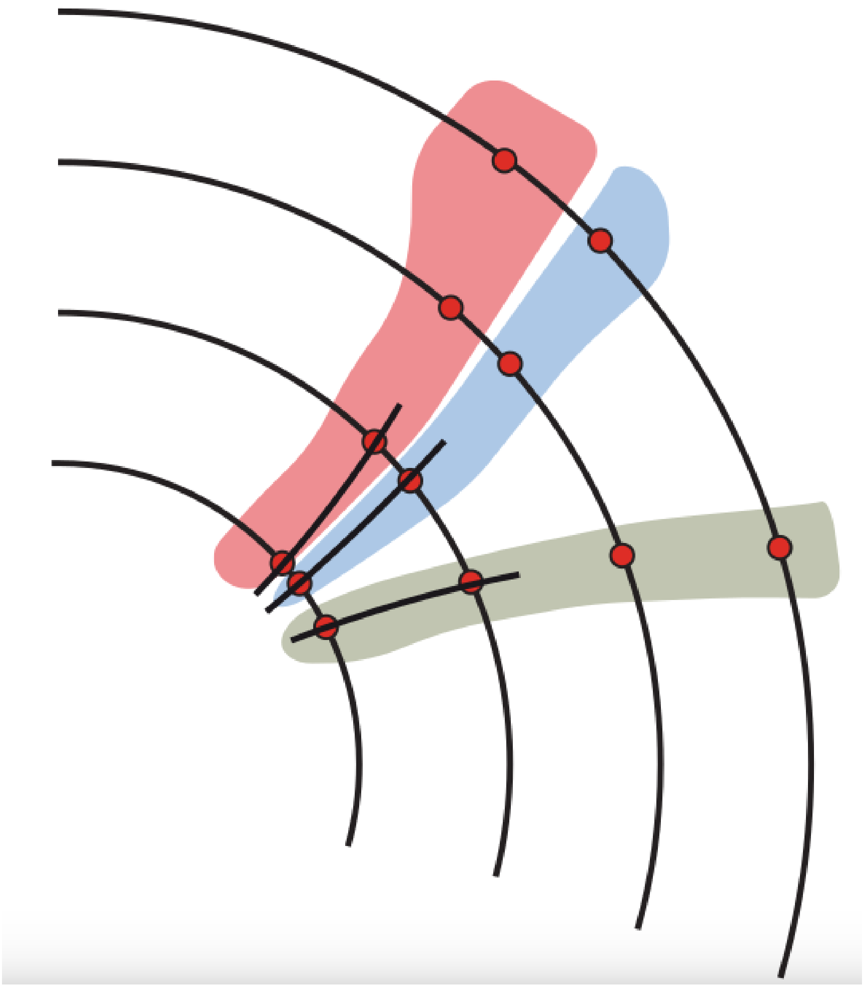
Local Track Finding

find seeds ~ combinations of 2-3 hits



Local Track Finding

Build roads along the likely trajectory

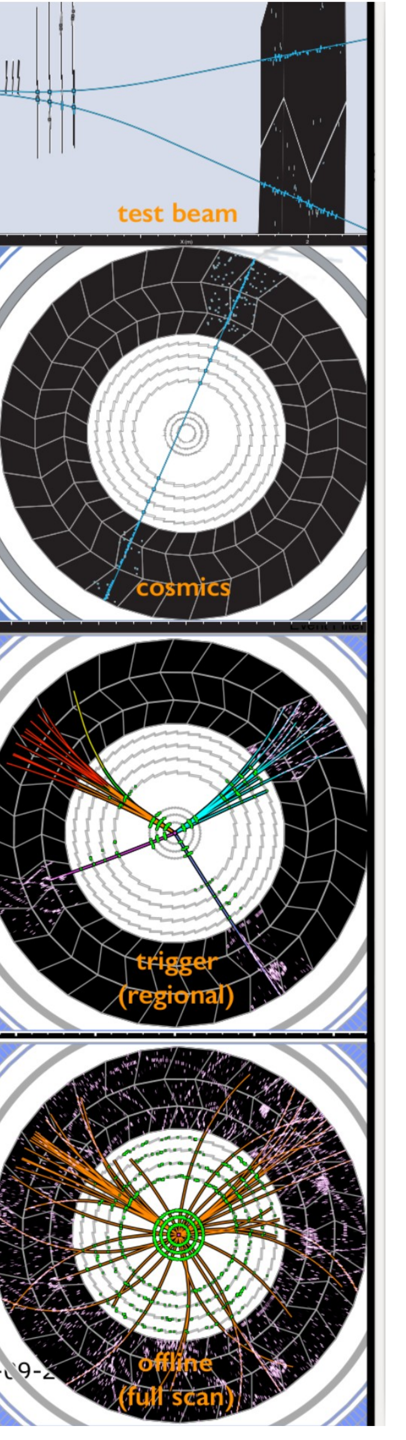


In practice

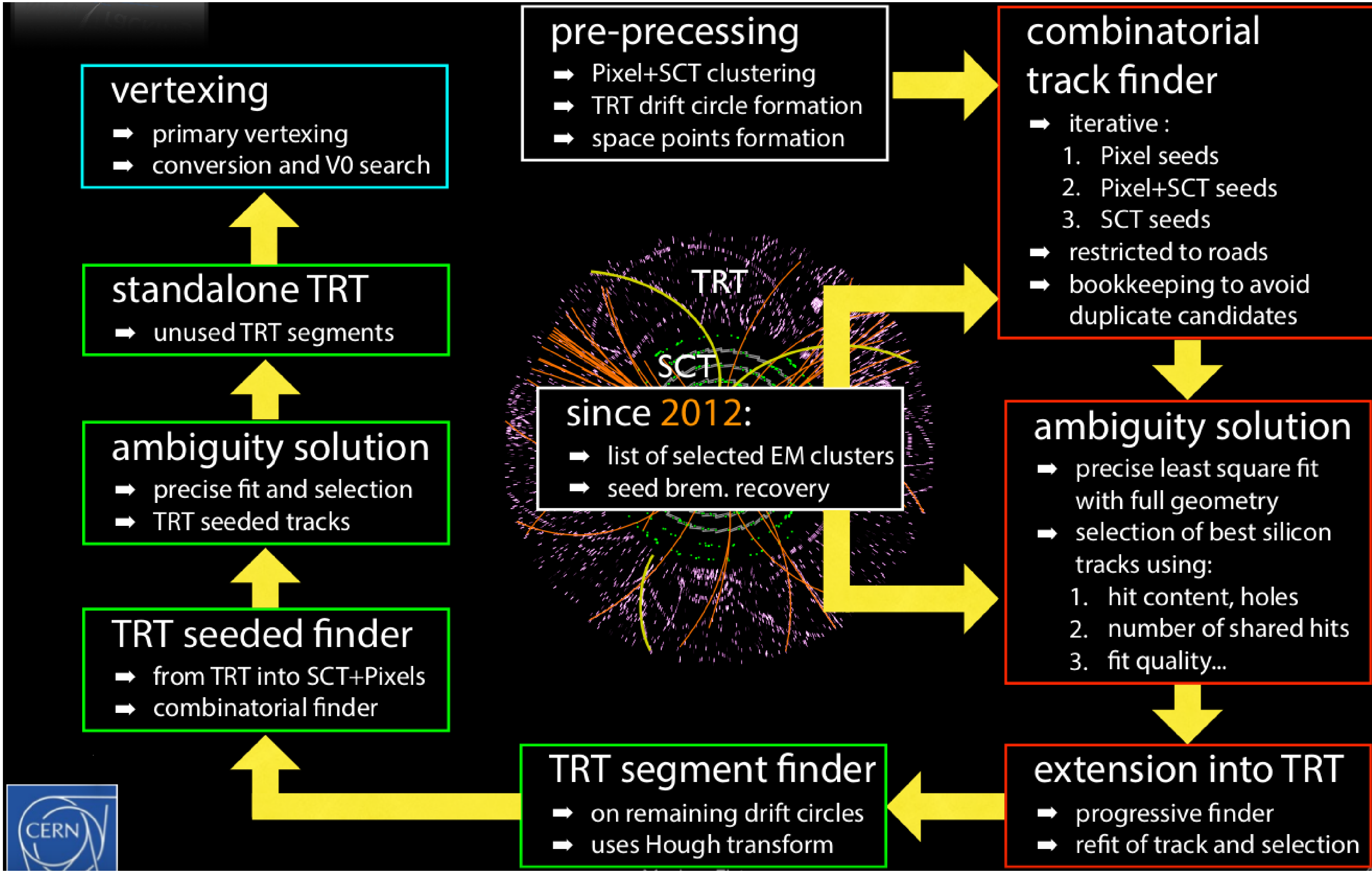
- Reconstruction strategy depends on:
  - detector technologies
  - physics/performance requirements
  - occupancy and backgrounds
  - geometry
  - technical constraints (CPU, memory)

- Even for same detector setup one looks at different types of events

- Track reconstruction used by experiments:
  - Usually apply a combination of different techniques
  - Often iterative ~ different strategies run on after the other to obtain best possible performance within
  - resource constraints



ATLAS Tracking Chain

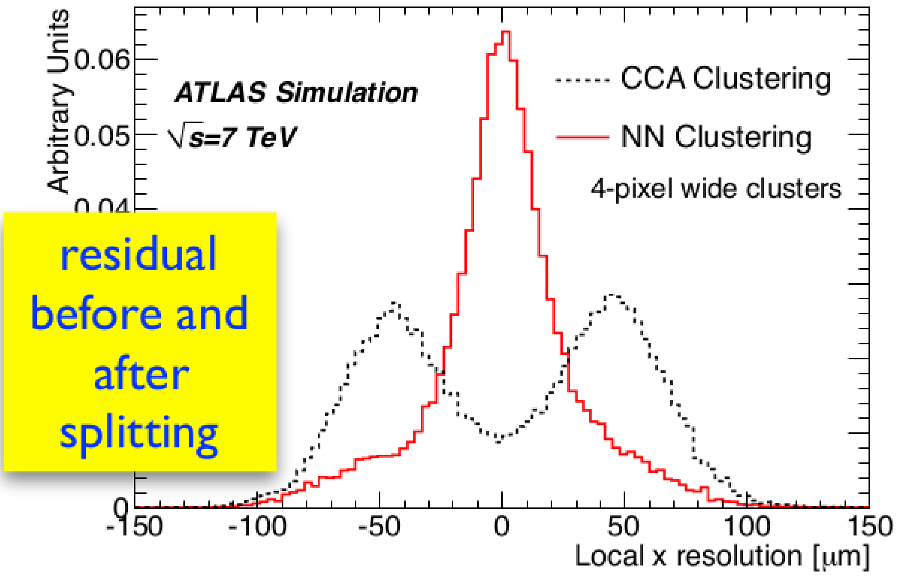


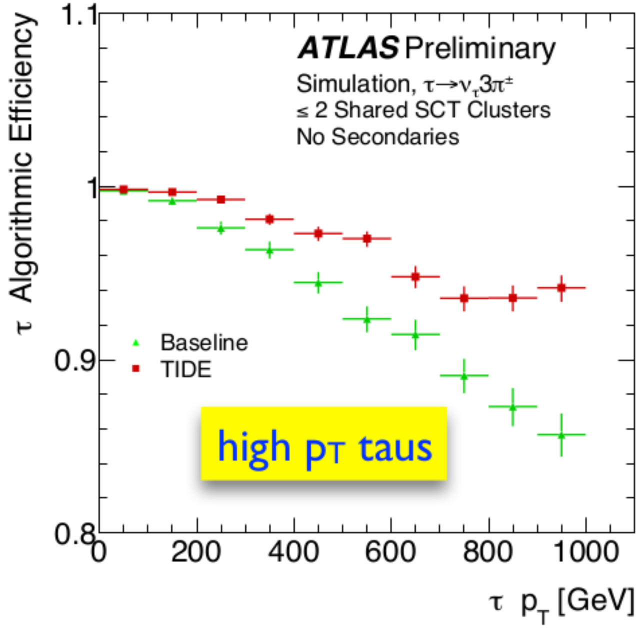
Tracking in Dense

- problem of cluster merging:
  - merging when track separation reaches single Pixel size

- Neural network (NN) Pixel clustering
  - identify merged clusters and splitting them
  - identify merge clusters, split them and correct positions

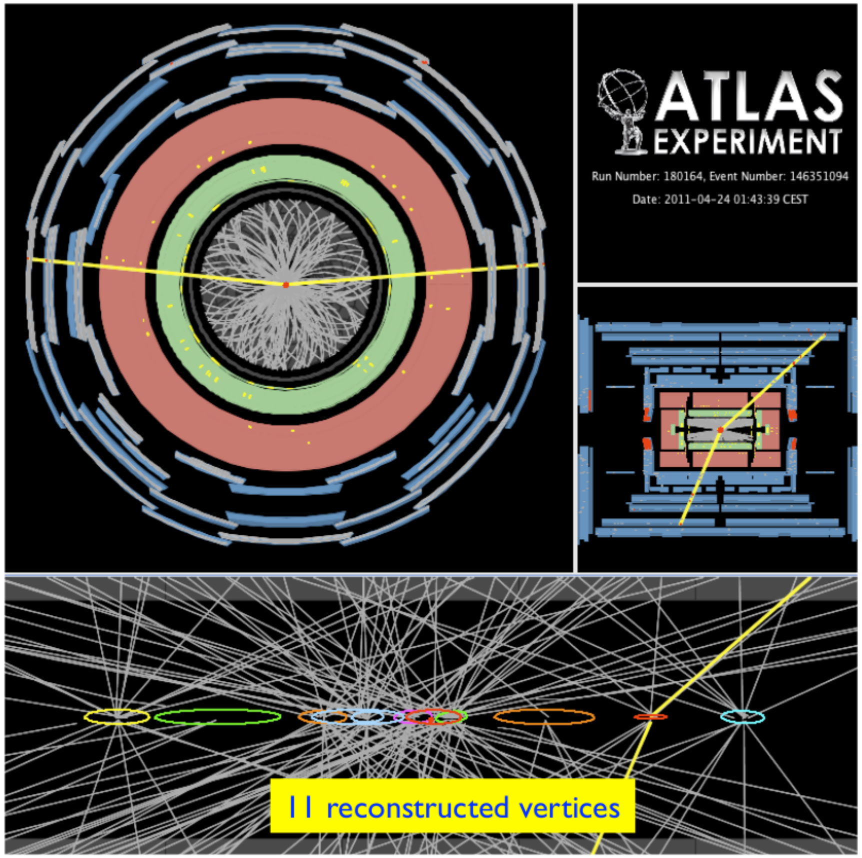
- Crucial in many areas:
  - ➡ b-tagging (especially at high momenta)
  - ➡ jet calibration and particle flow
  - ➡ 3-prong τ identification





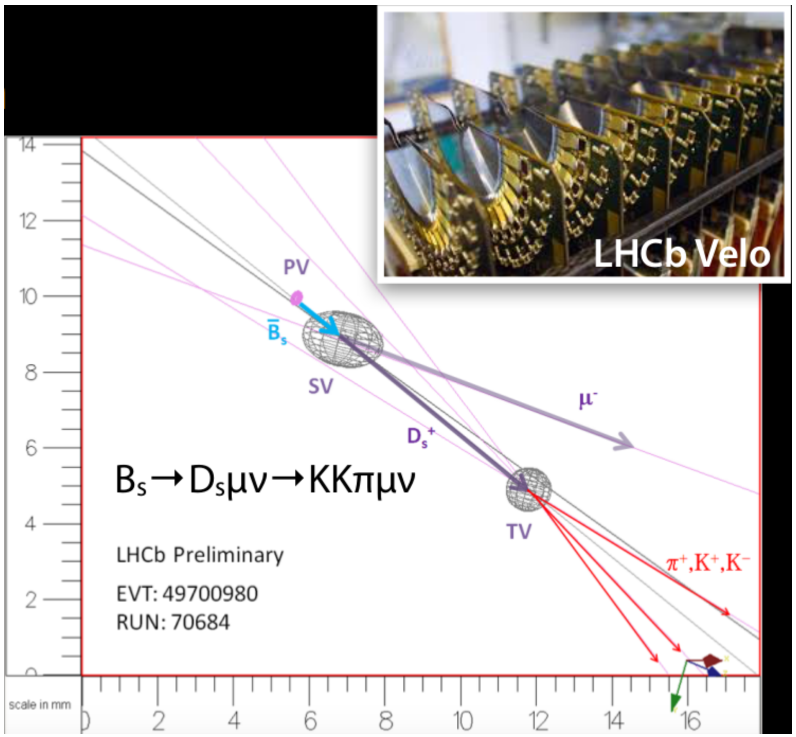
Vertexing

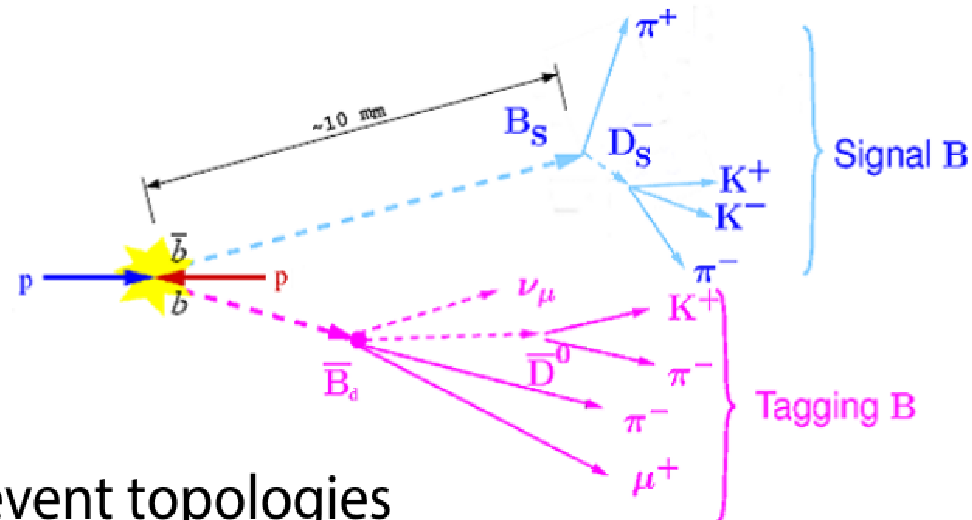
- Vertex fitting techniques play an important role:
  - in reconstruction chain following track reconstruction
  - primary interaction vertex reconstruction and identification
  - in time pileup estimation and pileup mitigation in particle flow reconstruction
  - secondary vertex finding for b-/c-jet identification, τ-reconstruction, photon conversions finding



Vertexing Application

- Explores b- and c-hadron lifetime:

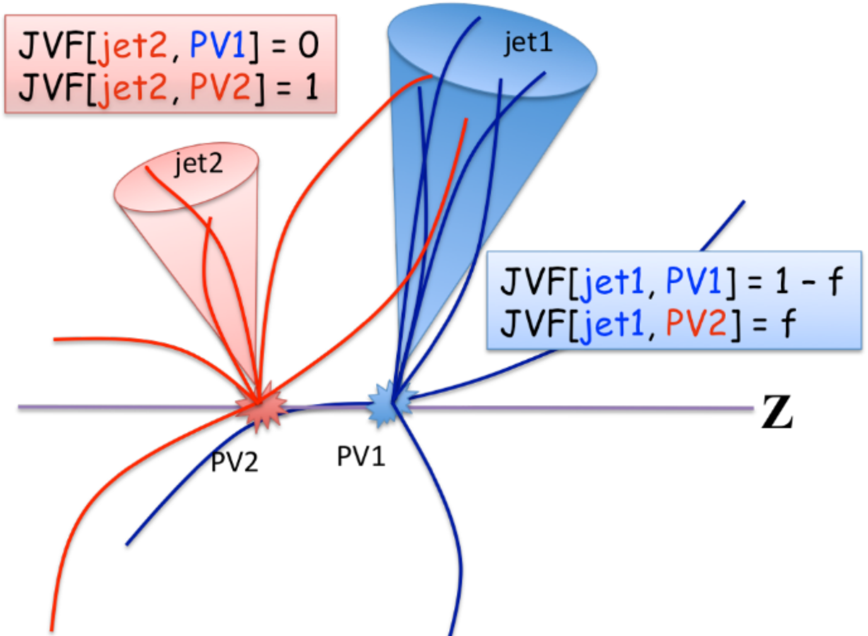


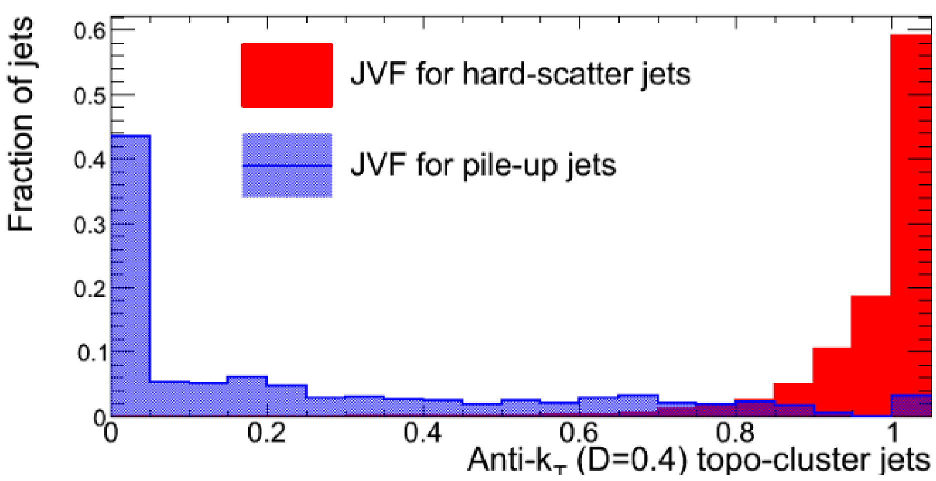


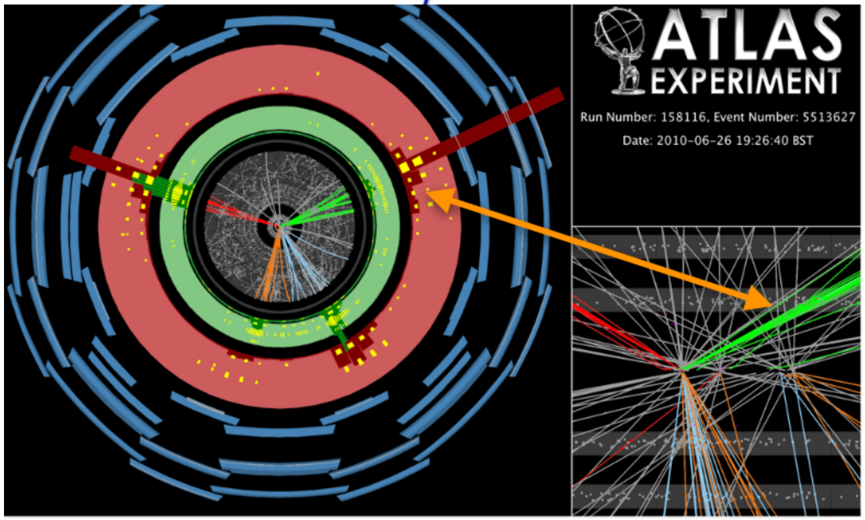
event topologies

Vertexing Application

$$JVF(\text{jet}_i, \text{vtx}_j) = \frac{\sum_k p_T(\text{trk}_k^{\text{jet}_i}, \text{vtx}_j)}{\sum_n \sum_l p_T(\text{trk}_l^{\text{jet}_i}, \text{vtx}_n)}$$

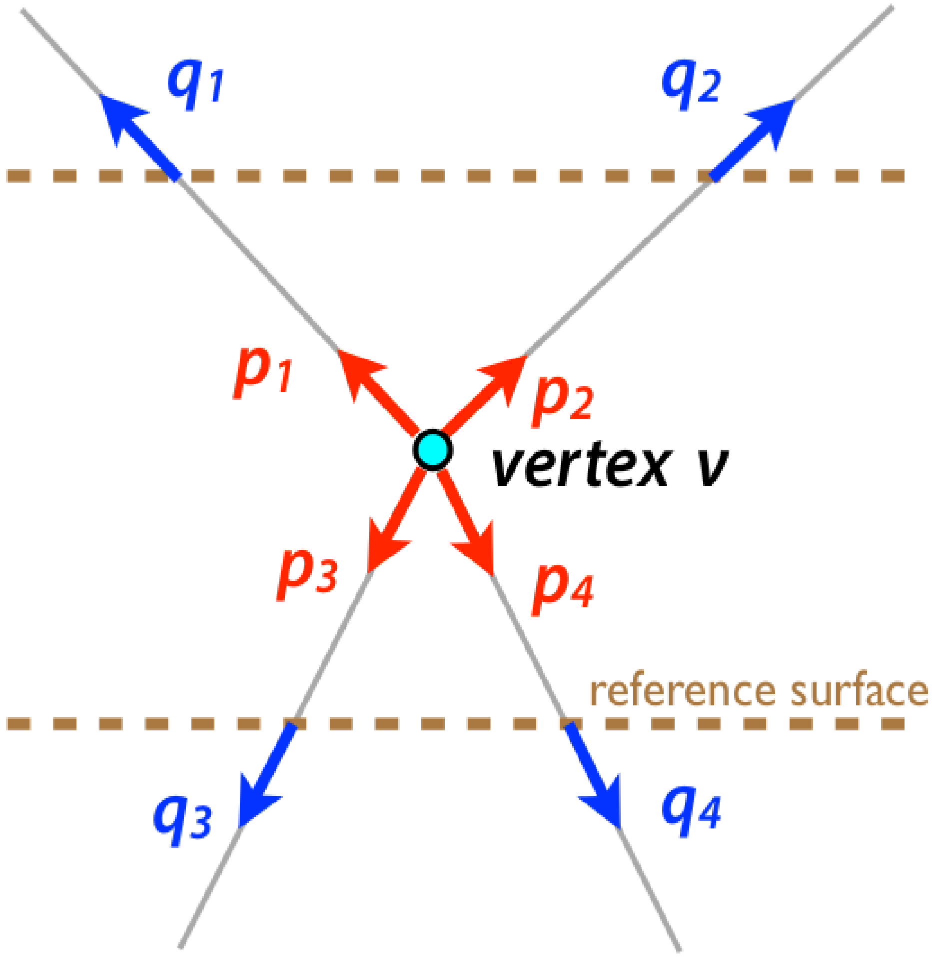






Vertex Fitting

- Task of a vertex fit:
  - ➡ start from a set of measured track parameters qi
  - ➡ estimate the vertex position ν
  - ➡ and the parameters pi at the vertex



$$q_i = h_i(v, p_i) + \varepsilon_i$$

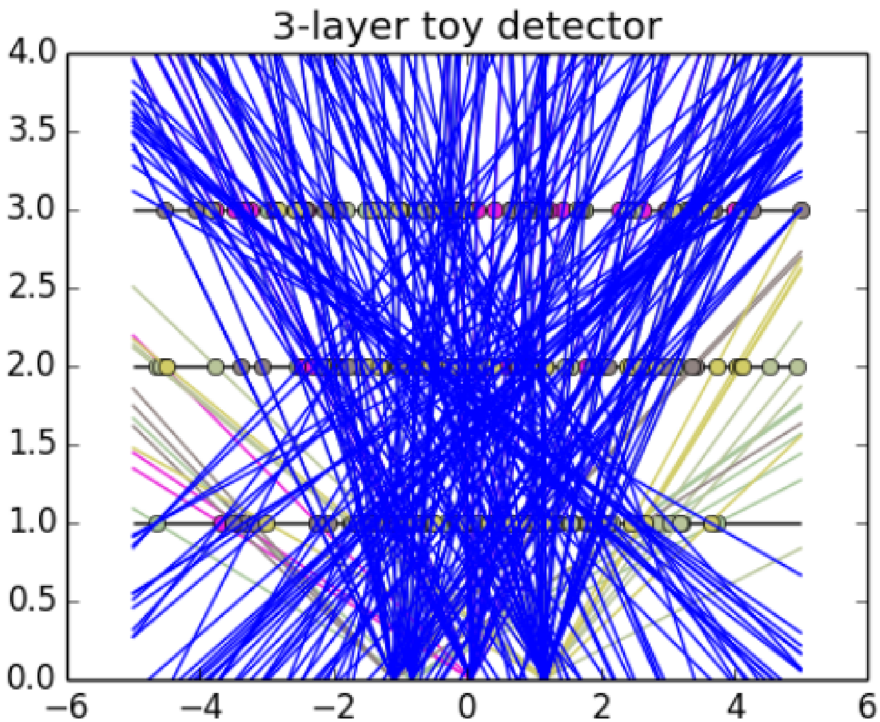
with: h_i ~ dependency of track parameters on vertex V and parameters q_i at vertex

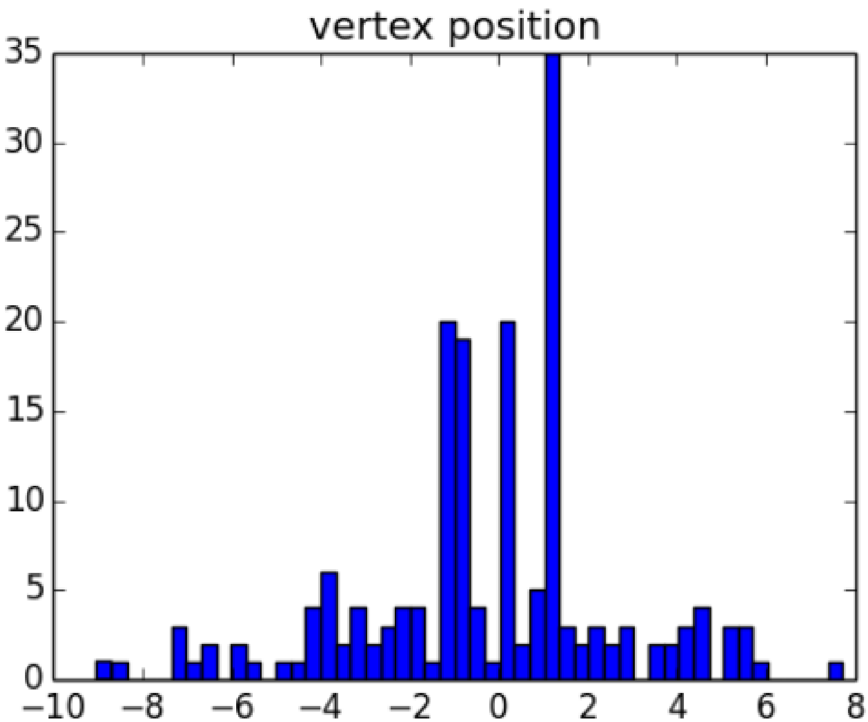
ε_i ~ error of q_i (noise term)

Jacobians: $A_i = \frac{\partial h_i(v, p_i)}{\partial v}$ $B_i = \frac{\partial h_i(v, p_i)}{\partial p_i}$

Vertex Finding

- Vertex z-scan on beam line:
  - ➡ histogram technique that searches for peaks in z0 of hit combinations extrapolated to beam line
  - ➡ used e.g. to seed primary vertex finding or to constrain HLT tracking to point to primary vertex





Let‘s Summarize

- I have briefly went through the techniques for trigger and tracking.

- Time limited, still miss many recent and interesting developments like:
  - Tracker trigger. Explore tracking at L0/L1
  - Machine learning for tracking
  - PF reconstruction
  - etc …

- High Energy Physics could tightly connect with all kinds of fields ...