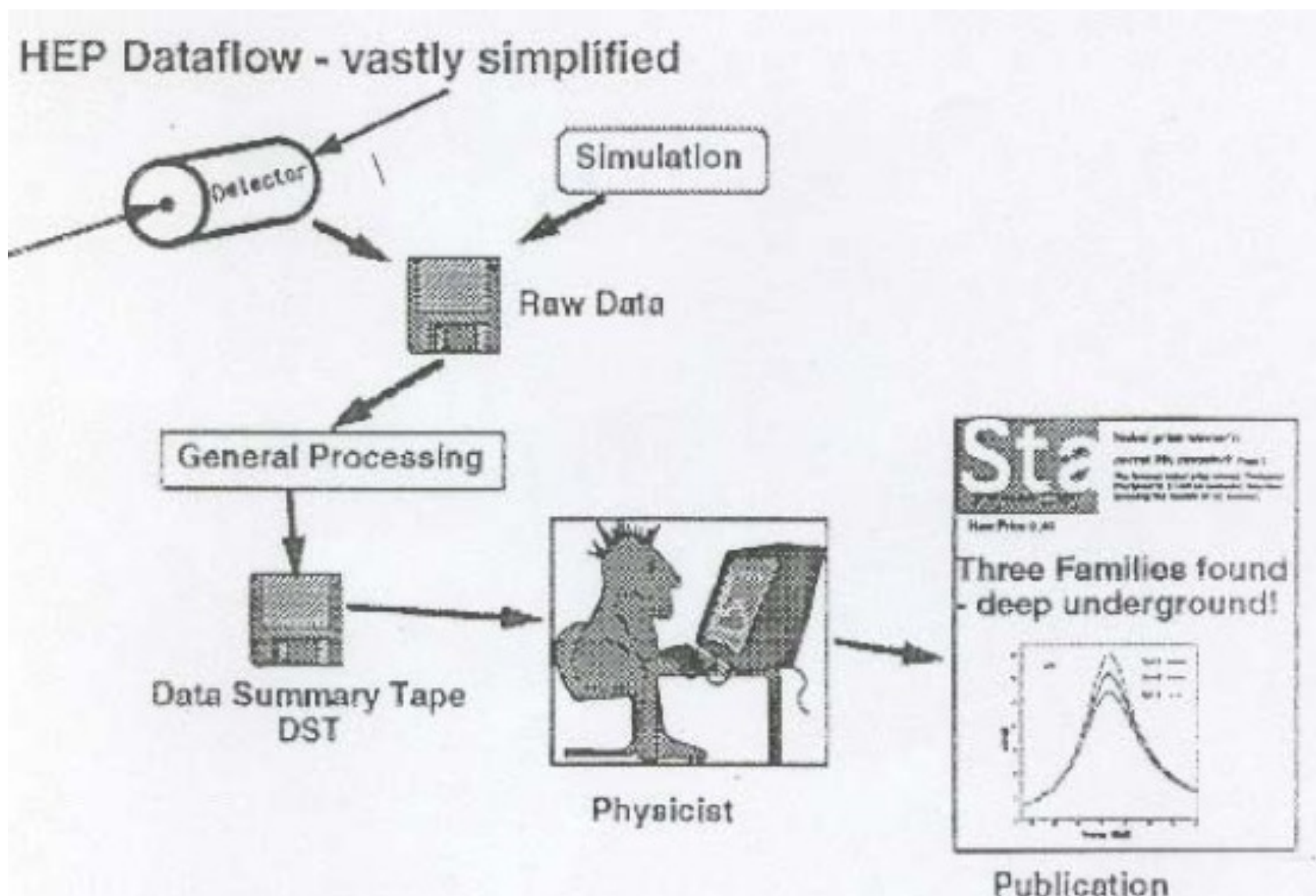


Experiments and Physics Objects

PHANC 2025
March 28, 2025

Sunanda Banerjee

- A HEP experiment provides some data read out from the detector consisting of locations of the channels and some digital information (ADC/TDC/...)

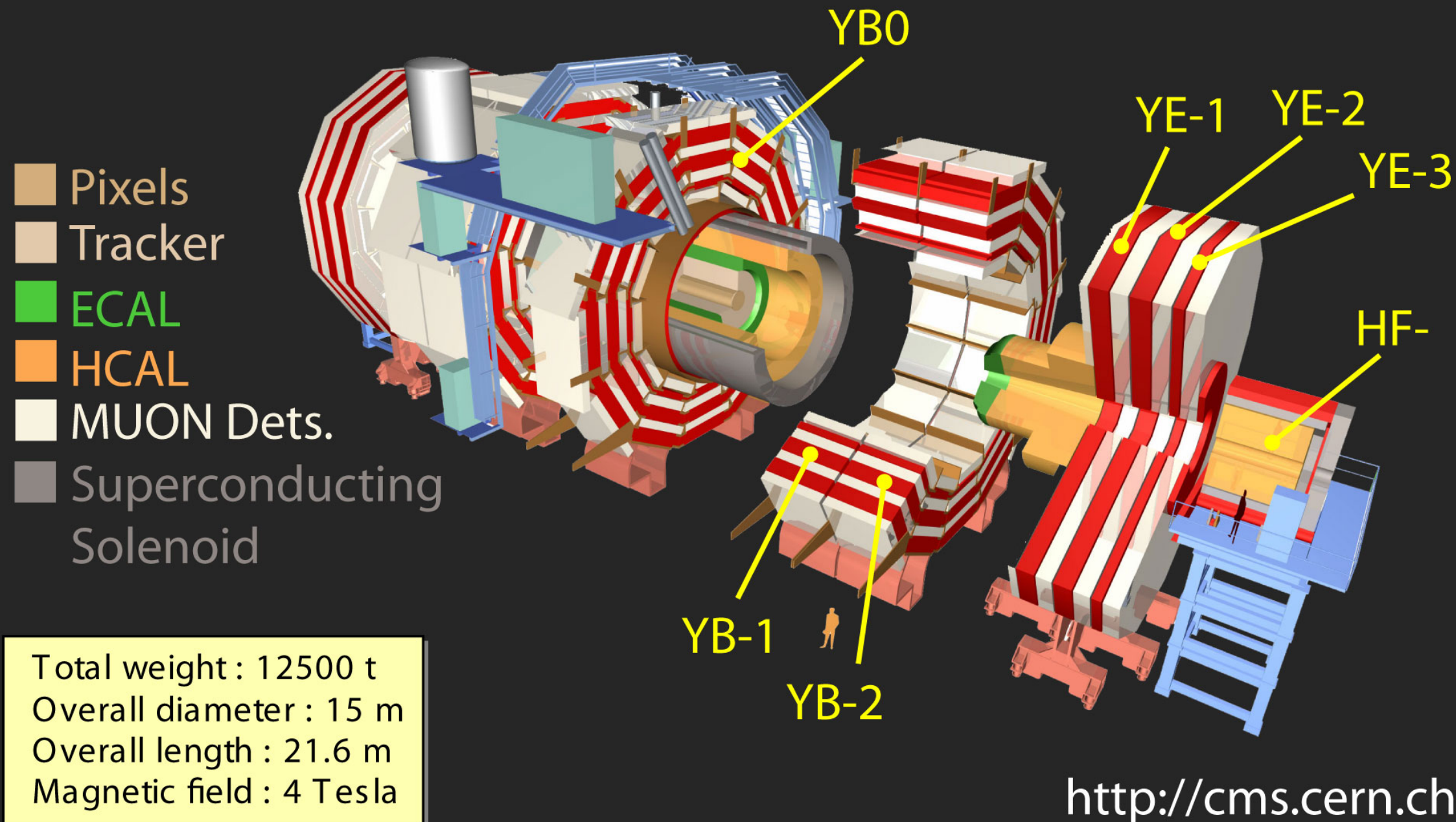


- The experimenters try to get some (Physics??) results from some derived data, known as “Physics Objects”:
 - Tracks, Vertices, Jets, Electrons, Muons, Photons,



A Detector in Action

- Beams in the Large Hadron Collider are made to collide at 4 points which are equipped with 4 detectors: ALICE, ATLAS, CMS and LHC-b



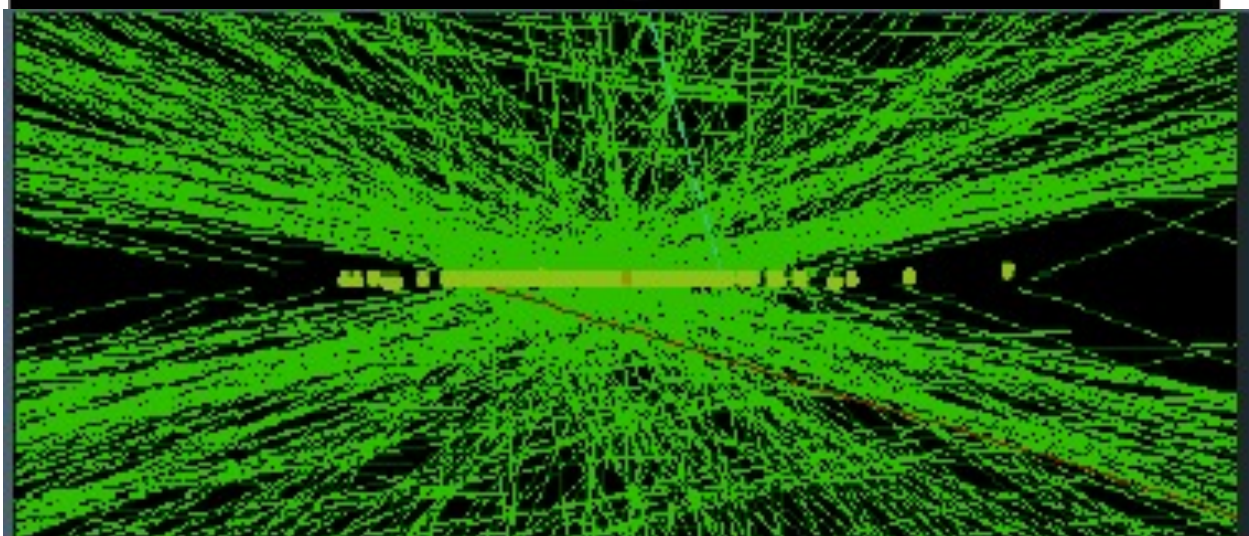
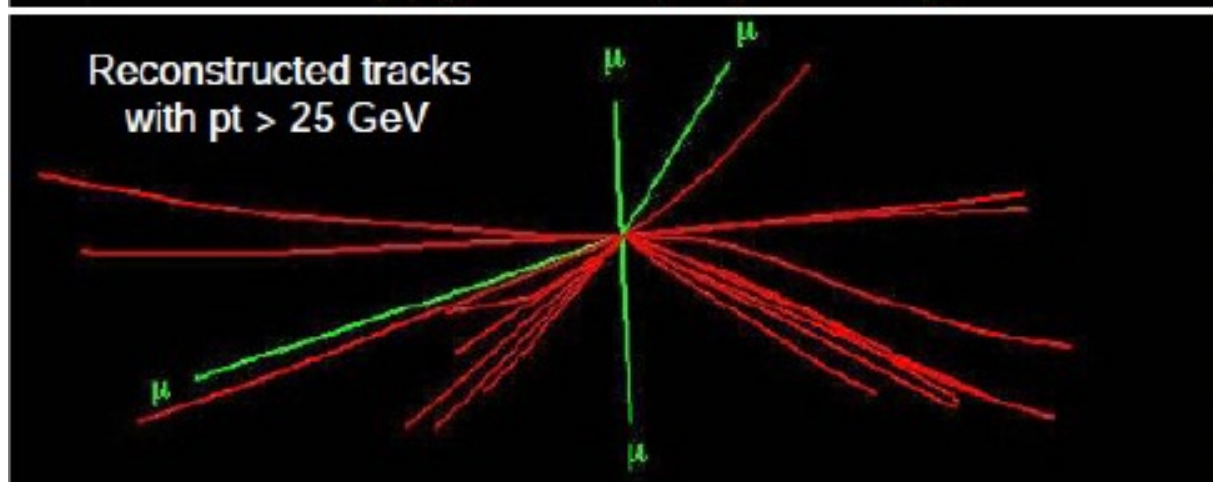
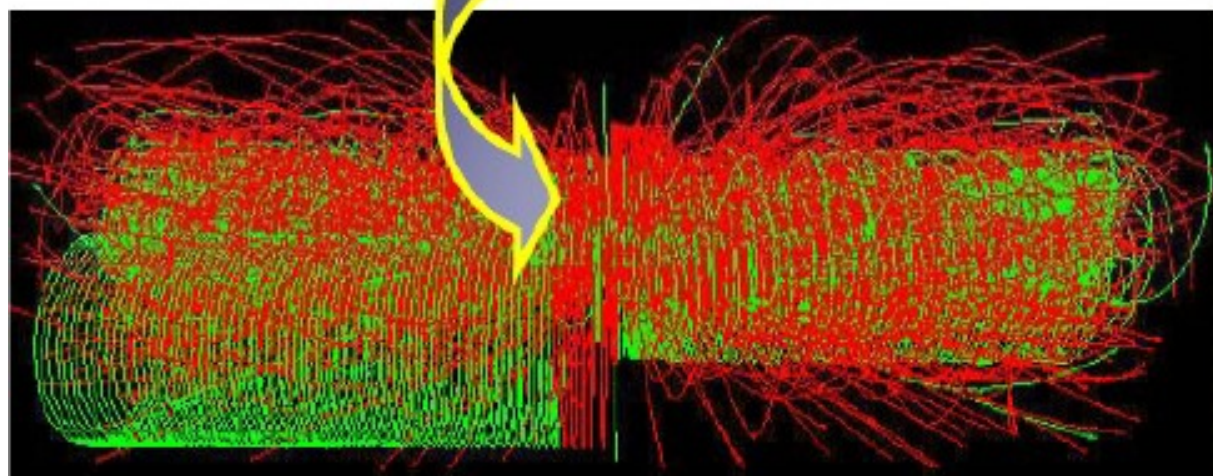
This is a Phase 1 detector about to be replaced by a more complex one



Features of a Modern Detctor



- Number of channels : > 100 Million
- Size: 15 m in diameter and 21.6 m long (also extends to ~ 140 m on either side)
- Precision of readout: $15\text{-}20\ \mu\text{m}$
- Alignment precision: $5\text{-}10\ \mu\text{m}$
- The dynamic range of energy/momentum measurement: a few 10 's of MeV to several TeV
- Bunch crossing time: 25 ns
- Data volume per event: ~ 2 MB
- Data recording rate: > 200 Hz
- Amount of data acquired per year: few peta bytes

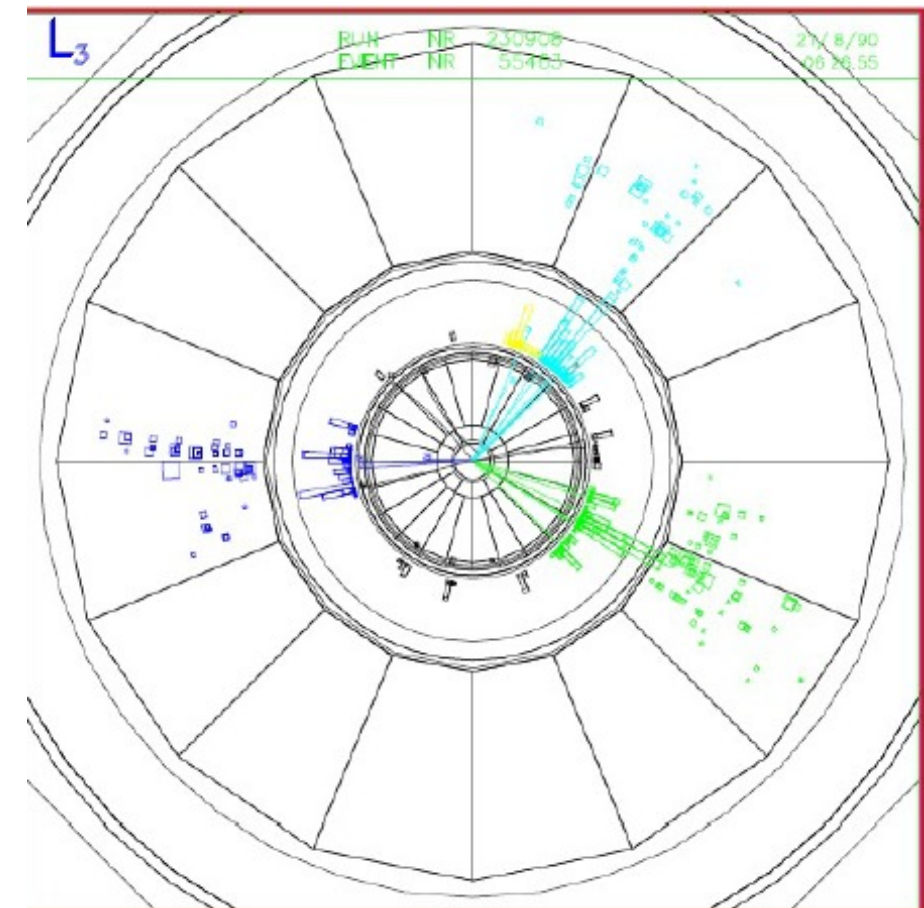


- In the presence of a **large magnetic field**, the charged particles of momenta less than 1 GeV spiralize inside the detector obscuring the image.
- Demanding to observe rare processes, there are **large number of interactions during a bunch crossing** (there are currently 50-70 interactions per bunch crossing which is expected to reach 140-200 in the near future)
- Detectors and the electronics are not as fast as the bunch crossing rate (40 MHz) resulting **interactions from many earlier and later bunch crossings** contributing to one event record

The Experiment Provides

0000000	123	cdef	8070	8070	4321	abcd	8061	8061
0000010	0	1fa4	0	0	0	8	0	1
0000020	0	31	0	2	4640	e400	0	9344
0000030	0	0	0	0	0	0	0	0
0000040	0	0	0	0	0	0	0	27
0000050	0	2	b	cb25	ffff	ff9d	0	0
0000060	0	0	0	0	20	5a38	0	0
0000070	0	0	0	0	0	0	0	0
*								
00000e0	0	0	0	0	0	0	0	9c3
00000f0	0	2	4640	e400	0	9344	0	0
0000100	0	0	0	0	0	0	0	2
0000110	0	990	17	fe61	0	27	0	2
0000120	b	cb25	0	2	5f	f2f4	2d5c	2d8b
0000130	0	0	20	5a38	3	d40	1853	600
0000140	0	10	0	0	0	180	0	0
0000150	2	180	0	0	0	0	0	0
0000160	0	0	0	0	0	4	0	0
0000170	0	4	0	0	0	0	0	0

Interpreted As



- What lies in between comes from the effort of many people
 - Size of the experiment - too many channels
 - ~hundreds of millions going to several billions
 - Complexity of the environment
 - Effort to reach the maximum precision



Constraint for Data Taking



- Though the number of sensitive elements runs in billions, not all the elements will not be fired during a bunch crossing even with out-of-time pileup
 - Store information from only those channels which are fired by the passage of particles
 - Keep maximum precision and at the same time use a minimum number of bits so that the data volume is not large
 - Raw data are chains of unsigned integers containing the channel identification and the ADC/TDC information
- Not all events are of interest. Need not store all events with larger cross-sections
 - Use a trigger table to choose only a certain fraction of events having required signatures
- Experiments like ATLAS or CMS store data at the rate of a few 100 Hz to 1 kHz with ~ 2 Mbyte data per event



First Level of Reconstruction



- Need to translate each bit of information (pair of geographical information of detector cell and ADC/TDC) into realistic information (space coordinate where the cell is and convert ADC into charge)
- Each detector has some static information - a map relating the cell identifier with a 3D position; and a conversion table of ADC/TDC to charge or time
 - These tables are not strictly static. They depend on environmental variables and are time-dependent
 - Need to align the detector - a must for all tracking detectors
 - Need to calibrate each channel - conversion of charges to energy deposits
- Momentum measurement needs the measurement of the curvature of the charged particle trajectory and knowledge of the magnetic field at each point in the detector
 - All environmental quantities are monitored and the effect of those is considered in the evaluation of the variables
- Calorimetric measurements need finer calibration to monitor the response function

Calibration and Alignment are fundamental to get reasonable Physics Objects



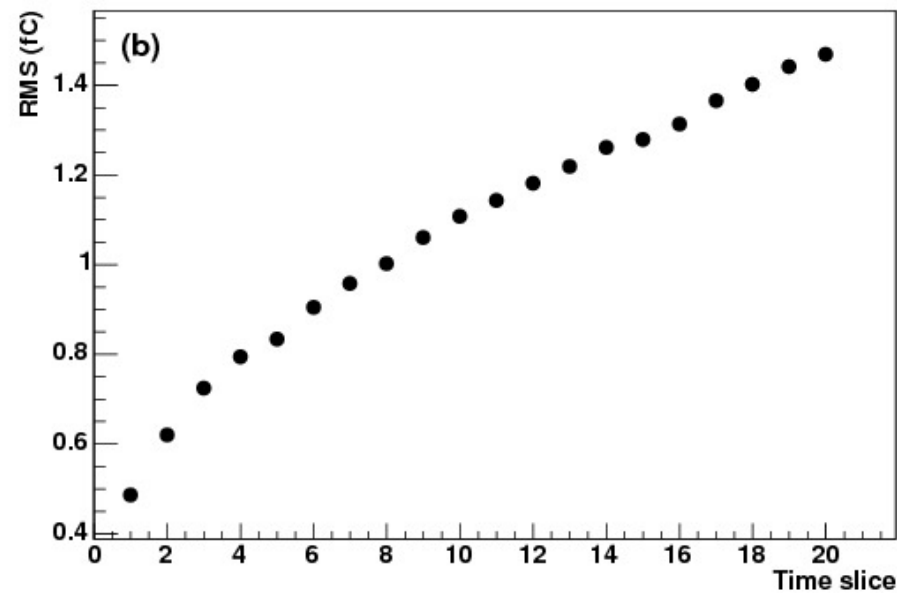
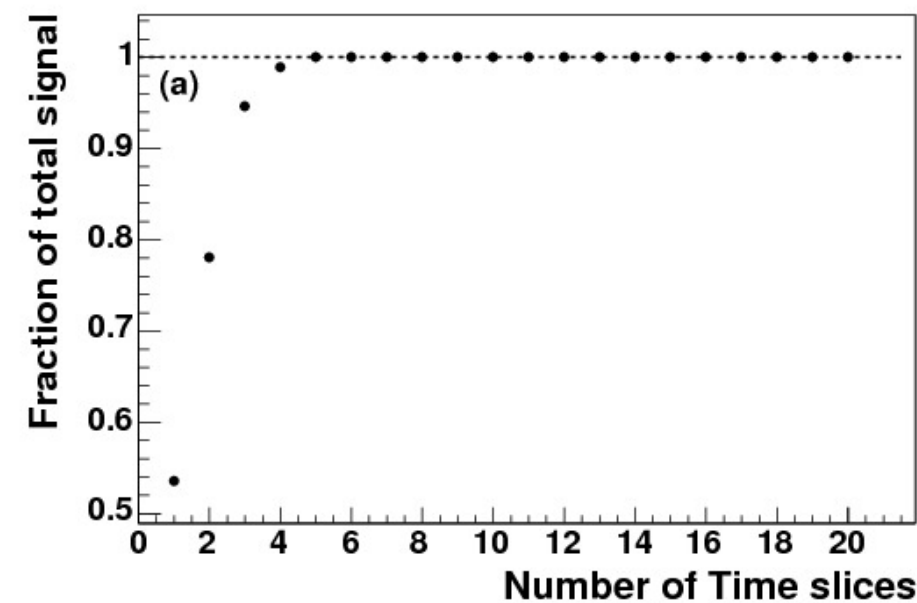
Data Processing



- This gives rise to two separate data streams:
 - Data collected during bunch crossing (possibly due to interactions of interest) **Synchronous data**
 - Data from all calibration runs, monitoring tasks, alignment devices, ... giving the status of the detector **Asynchronous data**
- Using synchronous data themselves for calibration data provides maximum constraint on data processing — cannot have a single pass reconstruction of the collected data
- Standard production schedule
 - Pass 1 in pseudo-real time (within a few hours from the data taking) known as PromptReco
 - Several re-reconstruction passes subsequently with better and better knowledge of detector conditions weeks-months-years time scale known as ReReco

The data structure is designed with this constraint in mind

An Example for Calibration



- Need to sum up energy deposit from 4 time slices to capture the entire energy
- However, the resolution gets worse with # of time slices used
- Use 2 time slices and take care of the loss in energy fraction

- The hadron calorimeter was initially calibrated using data from several dedicated test beam experiments and using signals from several built-in calibration systems using laser and/or some radioactive source.
- Take care of the loss due to radiation damage using laser run results which are done systematically in situ.
- The interactions provide azimuthal symmetric interactions. Using ϕ -symmetry inter-calibration of channels at a given is carried out for HB (barrel), HE (endcap) and HF (forward). For HO (outer), muon signals are used.
- Absolute η -dependent correction factors are estimated using isolated charged hadrons and comparing energy measurement in the hadron calorimeter with the track momentum as measured with the Tracker.
- Absolute energy measurement in the HF is done using Z decays to a pair of electrons where one of the electrons is in the ECAL and the other in the HF.



Data Processing

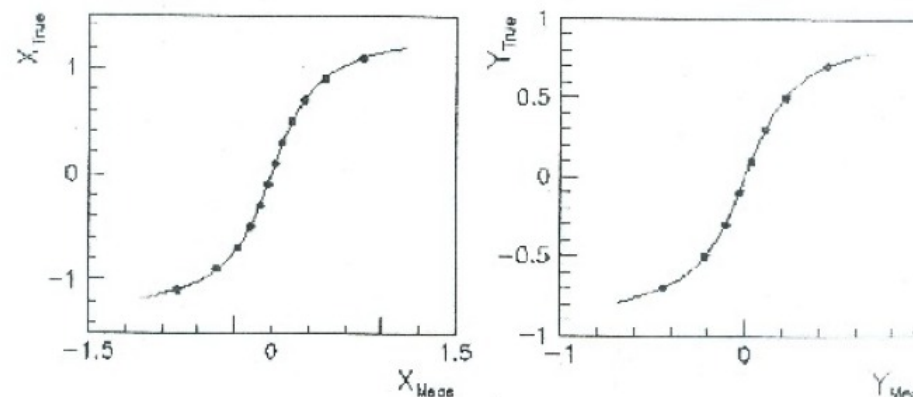
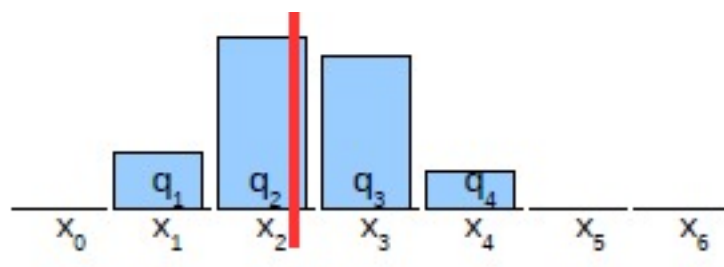


- Two types of algorithms are used in extracting information from the detector
- From tracking devices, where the particle shows its presence without losing its identity, the hit points where the particle leaves its trace within the detector are identified and joined together to define a Track. In the presence of the magnetic field, the deviation from a straight trajectory measures the momentum of the particle.

$$p \left(\frac{\text{GeV}}{c} \right) = 0.3 B(\text{T}) R(\text{m})$$
- In calorimetric devices, where the particle loses its identity by depositing all its energy to the detector material, the hit cells are joined together and the total energy of the incident particle is evaluated from the collected energy in the cells. Also, a vector sum of the hit cell positions (weighted by energy deposit in the cell) gives an idea of the direction of the particle. The centre of gravity in each detector plane is corrected by the S-curve

$$x_{COG} = \frac{\sum w_i \cdot x_i}{\sum w_i}$$

$$x_{True} = \alpha_1 \tan^{-1} [\alpha_2 \cdot (x_{meas} - x_0)] + \alpha_3 \cdot (x_{meas} - x_0)$$





Track Finding



- There are two types of pattern recognition codes in finding trajectories from Hits (one or more signals that could be related to a spatial position)
 - Global: All hits enter into an algorithm in the same way and a list of tracks is produced
 - The algorithm is linear with the number of tracks
 - Hough transformation, histogramming, template matching.,
 - Local: Select a track candidate at a time starting with a few points and then predict if additional points belong to the candidate
 - Computing time increases faster than linear with the number of hits
 - Track following, Road map, Kalman filter
- A good track-finding algorithm gives the same set of tracks irrespective of the order in which the hit points appear in the hit collection



An Example of Hough Transformation



- Principle: Define a set of n different functions of coordinates and enter the function values in a n -dimensional histogram. Tracks will appear as peaks in the n -dimensional histograms
- Measurements from the Tracker(Hits) are expressed in cylindrical (r, ϕ, z) coordinates and are quantized (particularly in r)
 - Precompute a finite number of quantities of interest

$$\phi - \sin^{-1} \left(\frac{C \cdot r}{2} \right)$$

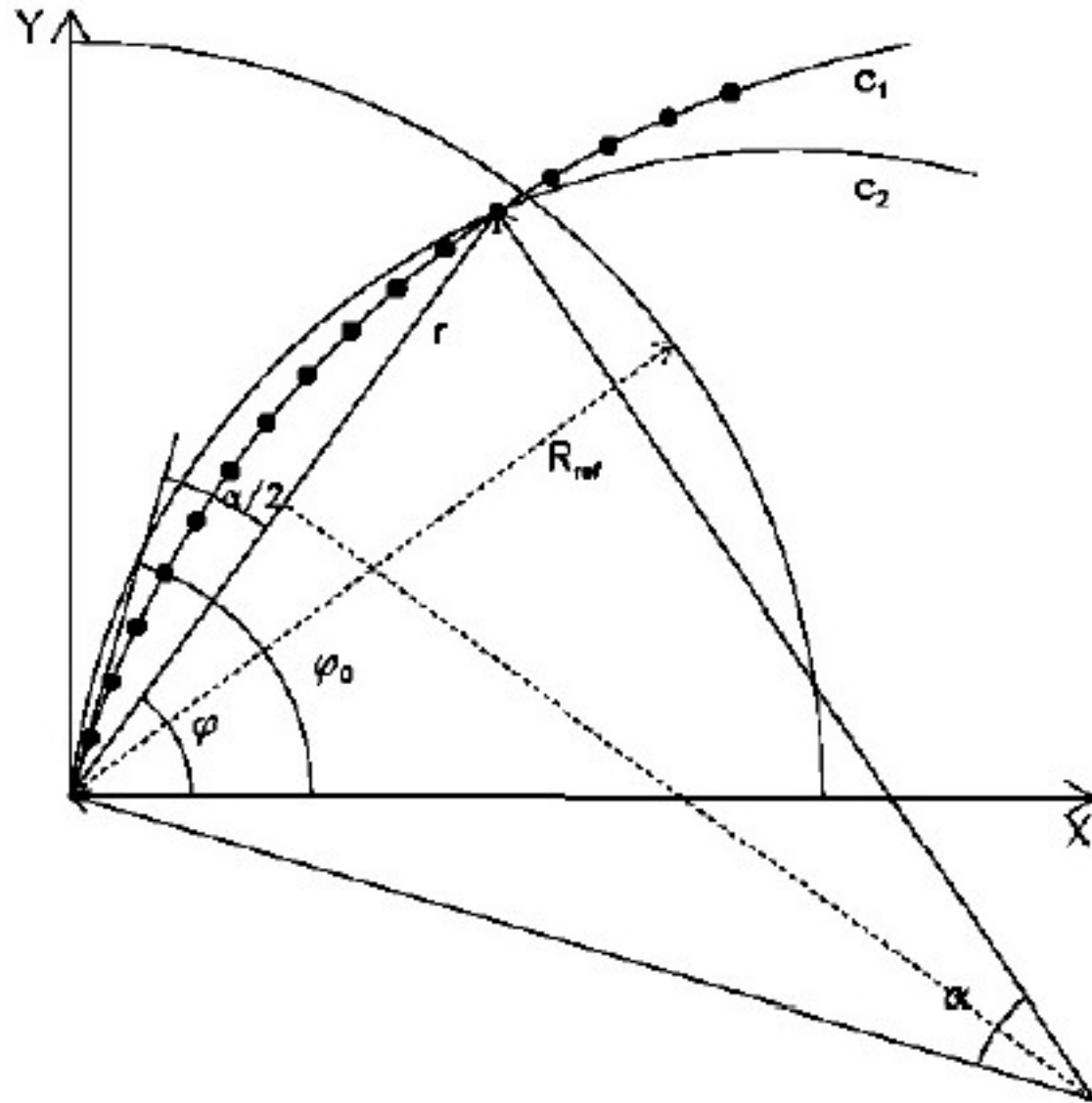
where C is the sampled signed curvature

- If C is close to the curvature of the track of interest, the quantity tends to ϕ_0 of the track at $r = 0$. It will be the same for all measurements of the track giving rise to a peak in the histogram at a fixed ϕ_0 for a given C .
- Given a solenoidal magnetic field, tracks originating from the primary vertex will follow a circle in the transverse plane. So it can be parametrized as

$$f_i(R_{ref}) = \phi_i - \sin^{-1} \left(\frac{Cr_i}{2} \right) + 2 \sin^{-1} \left(\frac{CR_{ref}}{2} \right)$$

$$f_i(R_{ref}) \equiv \text{azimuthal angle of the trajectory at } r = R_{ref}$$

Hough Method: Applied to CMS



- Sample points for all curvatures within a range. All points belonging to a single trajectory will have the same value of f , if it is computed with the closest sampled curvature giving rise to a peak in the scatter plot f vs C

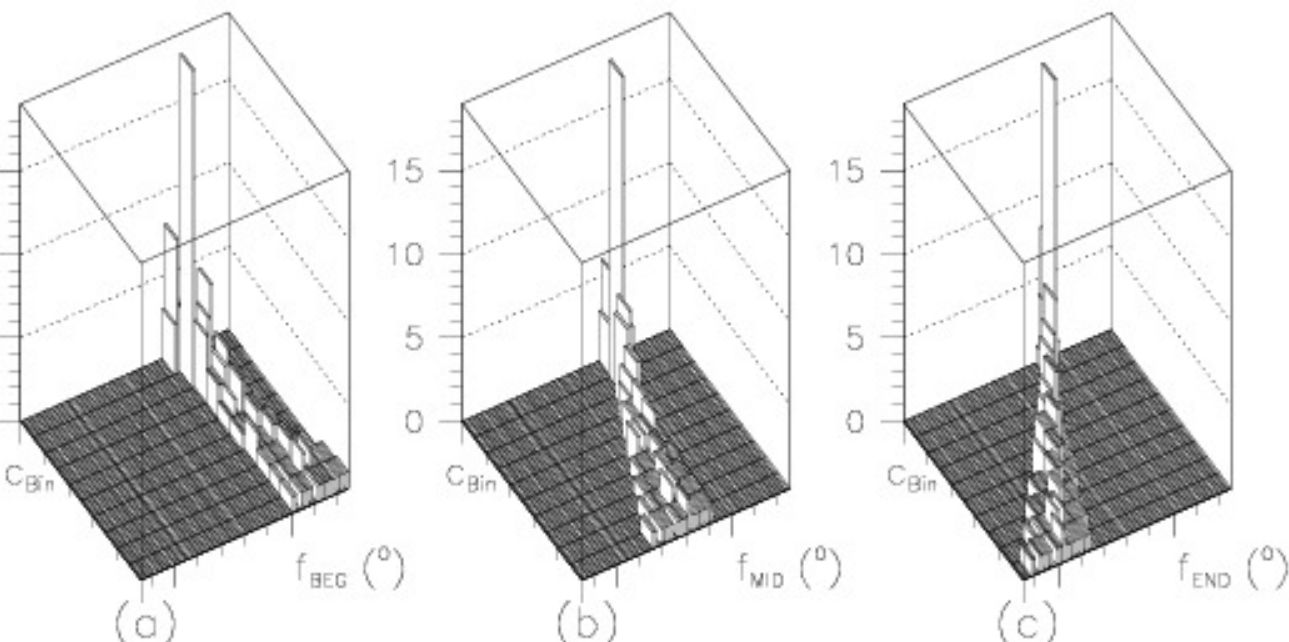
$$\phi_0 = \phi + \frac{\alpha}{2}$$

$$\sin\left(\frac{\alpha}{2}\right) = \frac{r}{2} \cdot \frac{1}{\rho}$$

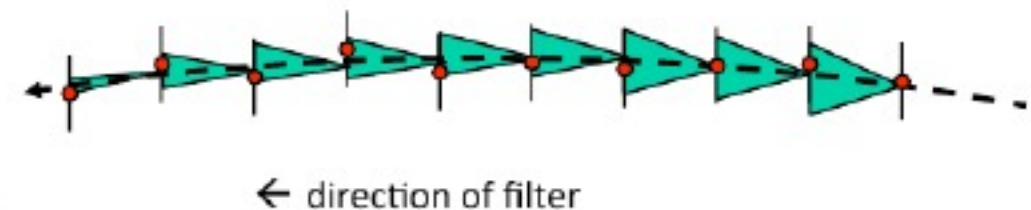
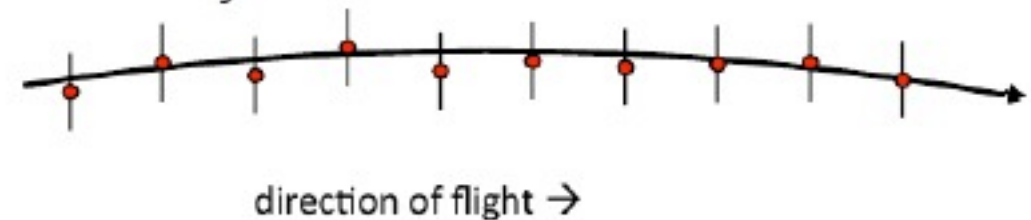
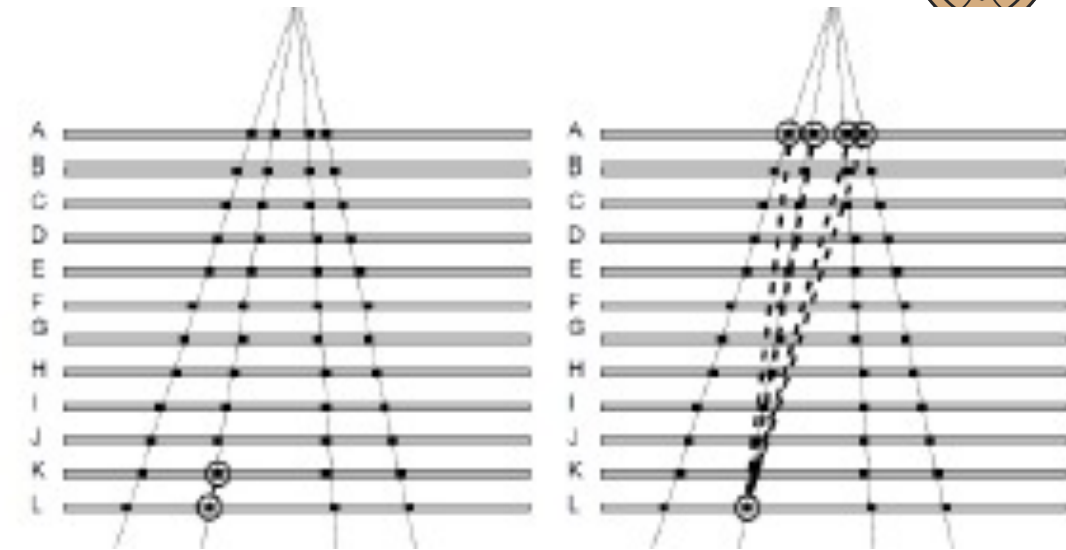
$$\frac{1}{\rho} = -C$$

$$\Rightarrow \phi_0 = \phi - \sin^{-1}\left(C \cdot \frac{r}{2}\right)$$

- For a single muon of $p_T = 41.4$ GeV, choose three values for the reference R
 - near origin,
 - at maximal R ,
 - at the half way.
- peak is observed for all three reference values.
- Tested also in Multi-track environment

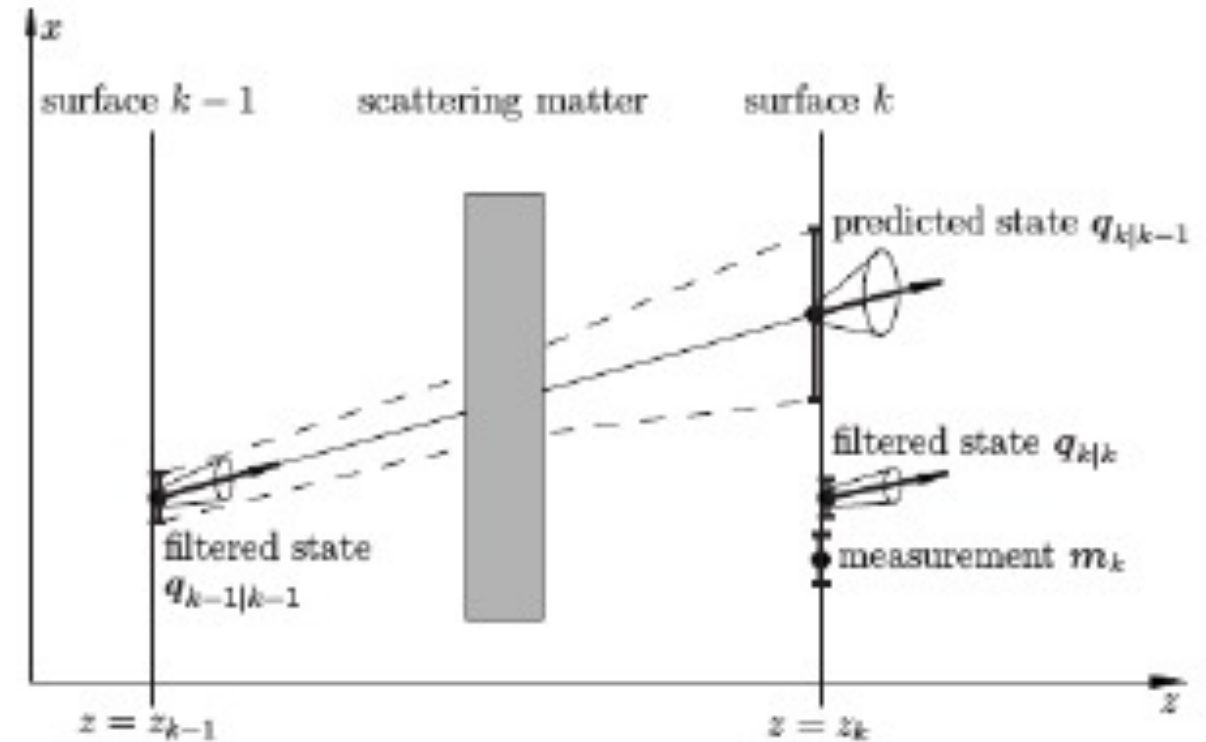


- Step 1: Find the seed track. This is often done in a combinatorial approach by using 3-4 innermost or 2-3 outermost and 1 innermost layer of measurement. The innermost ones are chosen because they have the best resolution and the outermost ones are chosen since tracks are separated maximally there. The intention is to reduce the number of possible combinations.
- Step 2: Make the track estimate based on the currently available hit points for the track and the uncertainties on them. These uncertainties come from the predictions for hits in the next layer from the current layer
- Step 3: If a hit is found in the next layer, add this hit to the track and update the track prediction for the next layer
- Continue with steps 2 and 3 till the end of the tracker



Advantage of Kalman Filter

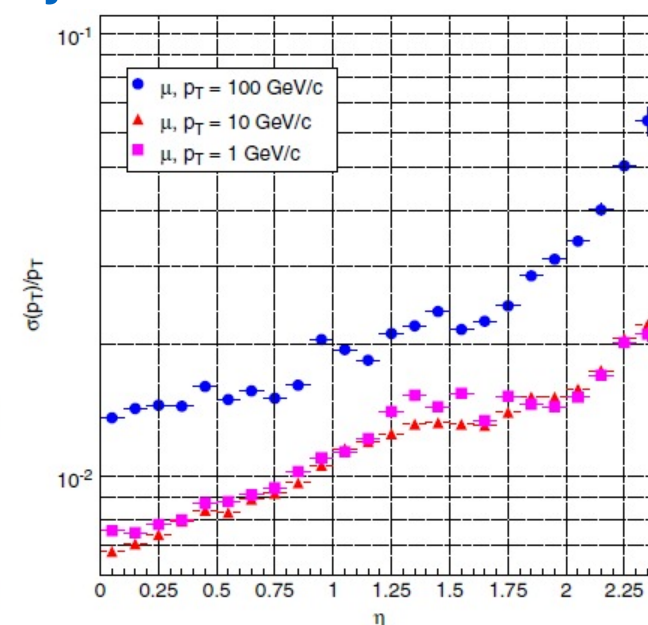
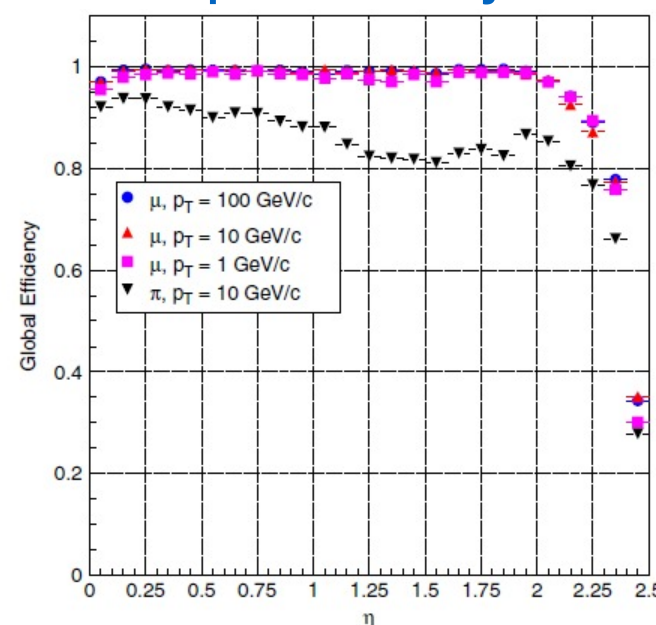
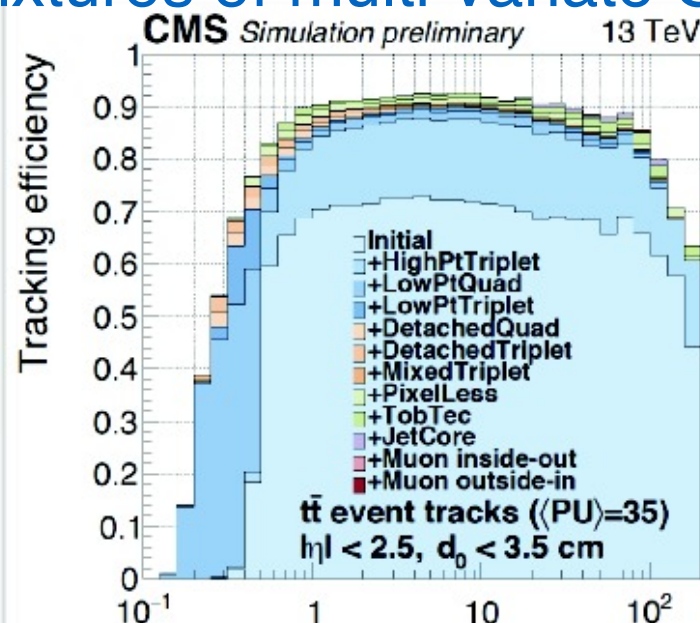
- It is easy to take detector effects into account:
 - energy loss,
 - multiple scattering,
 - bremsstrahlung.
 can be accommodated by recognising its effect and correcting it for next step



- The progressive fit is suitable for combined track finding and track fitting
- There is no large matrix to be inverted and the number of computations increases only linearly with the number of measurements
- The estimated track parameters closely follow the real path of the particle and this linear approximation of the track model needs to be valid only between two surfaces of measurements

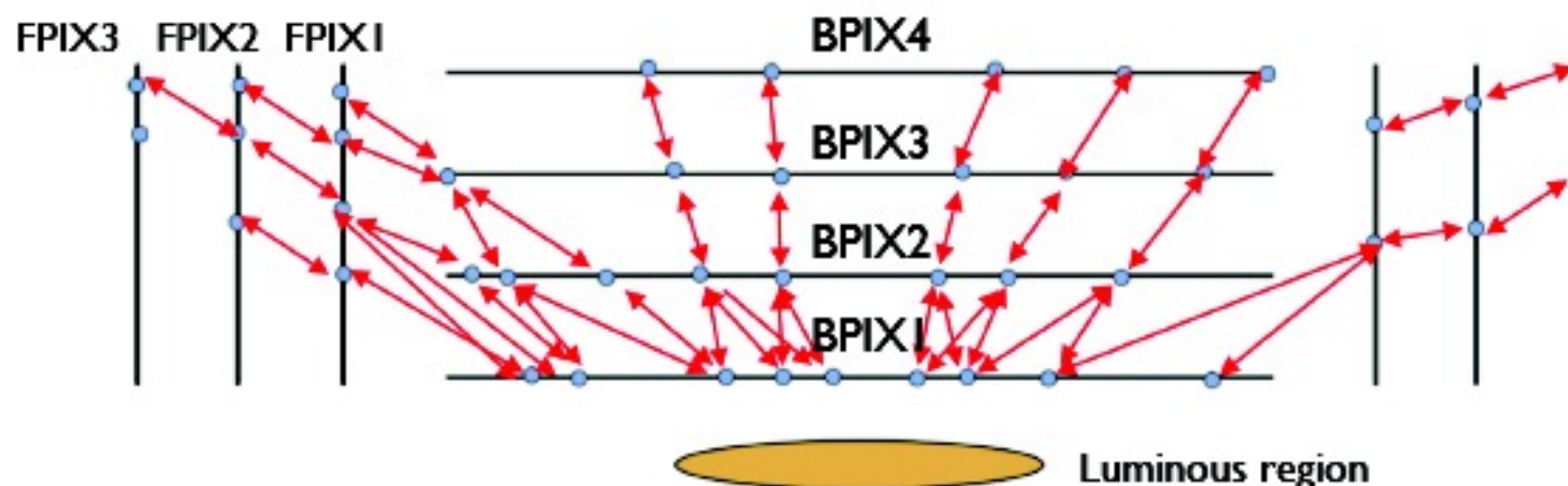
What is used during Run2/Run3

- The track reconstruction is done in several stages
 - In the first stage, the trajectory candidates are constructed for each pair of hits in two given layers compatible with the beam spot and minimum transverse momentum and are grown layer by layer from inside to the outermost silicon strip layer. Good track candidates satisfying quality criteria are kept on the basis of chi-square and the number of missing hits.
 - The hit assignment ambiguities among all the candidates are resolved by selecting a subset of compatible candidates. Next, the seeds that produce good track candidates are removed from the list and some of the criteria are relaxed (like beam spot requirement) and the same procedure. Backward propagation from outside to inside is also used.
 - In the presence of non-Gaussian probability distributions, an alternative Gaussian-sum filter is used where all involved distributions are modelled by mixtures of multi-variate Gaussian probability density functions.



What is being done for HL-LHC

- Parallelizing seeding using Cellular Automata: It requires a list of layers and their pairings
 - Create a graph of all possible connections between the layers
 - Create doublets aka cells for each pair of layers compatible with a region hypothesis
 - Make a fast computation of the compatibility between two connected cells
 - This does not require the world outside adjacent neighbouring cells. So, it is easy to parallelize the code



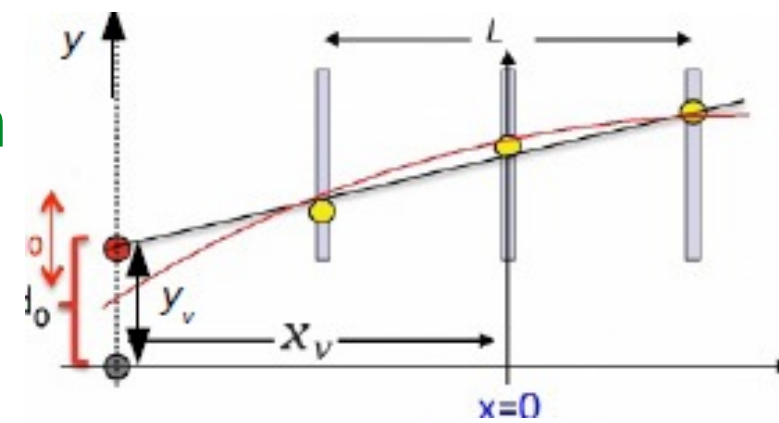
- Parallelize Kalman Filter Tracking:
 - Perform different tasks at the same time on different pieces of data
 - Splitting the workload evenly is difficult as occupancy in the detector is not uniform on a per event basis
- Quantum Annealing ???

Vertex Reconstruction

- Many of the particles produced in the interactions have small but finite life-time. Since they are produced at high energies, the decay length is large enough to be measured in the experiment

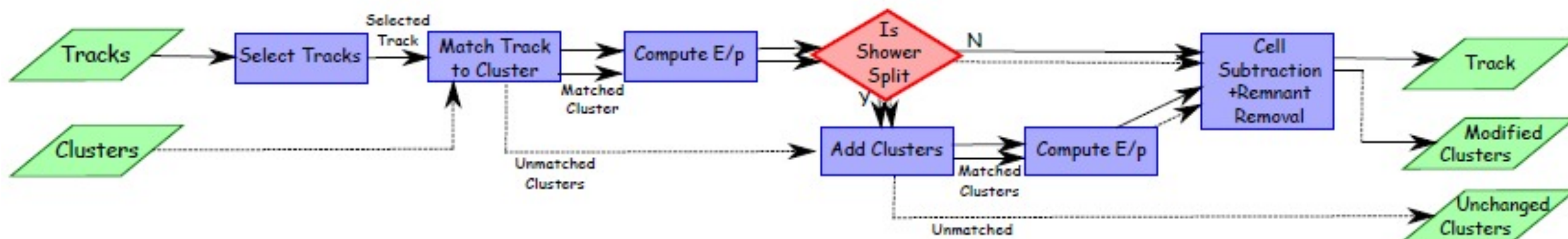
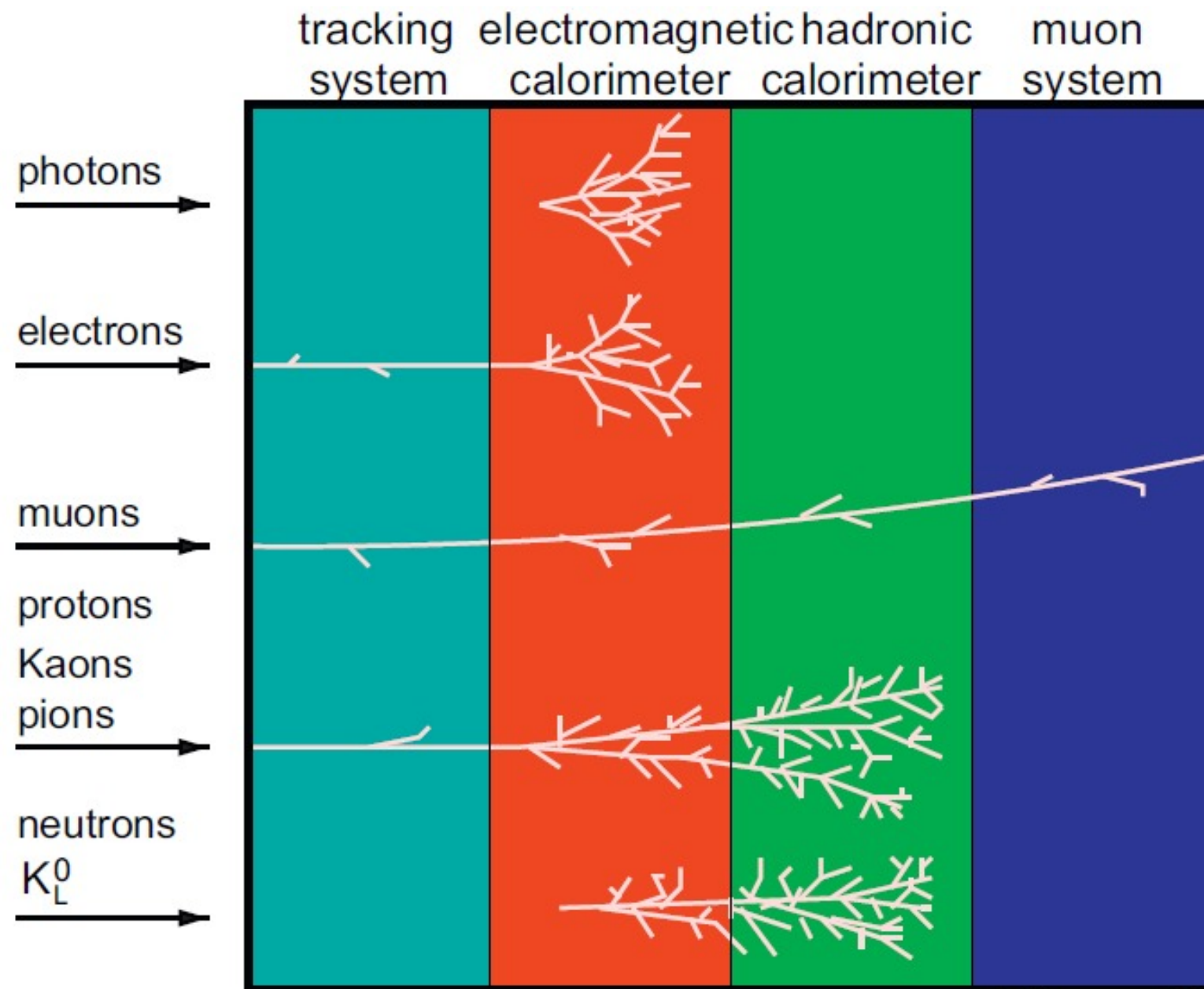
Particle	m (GeV/ c^2)	τ (10^{-12} s)	$l(p_T = 10 \text{ GeV})$
τ^+	1.776	0.290	500 μm
D^0	1.865	0.410	700 μm
D^+	1.869	1.040	1700 μm
Λ_c^+	2.286	0.200	300 μm
B_S^0	5.367	1.512	800 μm
B^+	5.279	1.641	900 μm
Λ_b	5.619	1.425	800 μm

- Identifying the primary vertex (PV), secondary (SV), tertiary (TV) vertex helps to understand the decay chain
- Impact parameter d_0 qualifies the mismatch to the PV. The impact parameter significance $S_{d0} = d_0/\sigma_{d0}$ serves as a selection criterion to distinguish tracks from primary or secondary vertices
- Tracks as measured are extrapolated to the vertices using quadratic extrapolation due to the presence of B-field



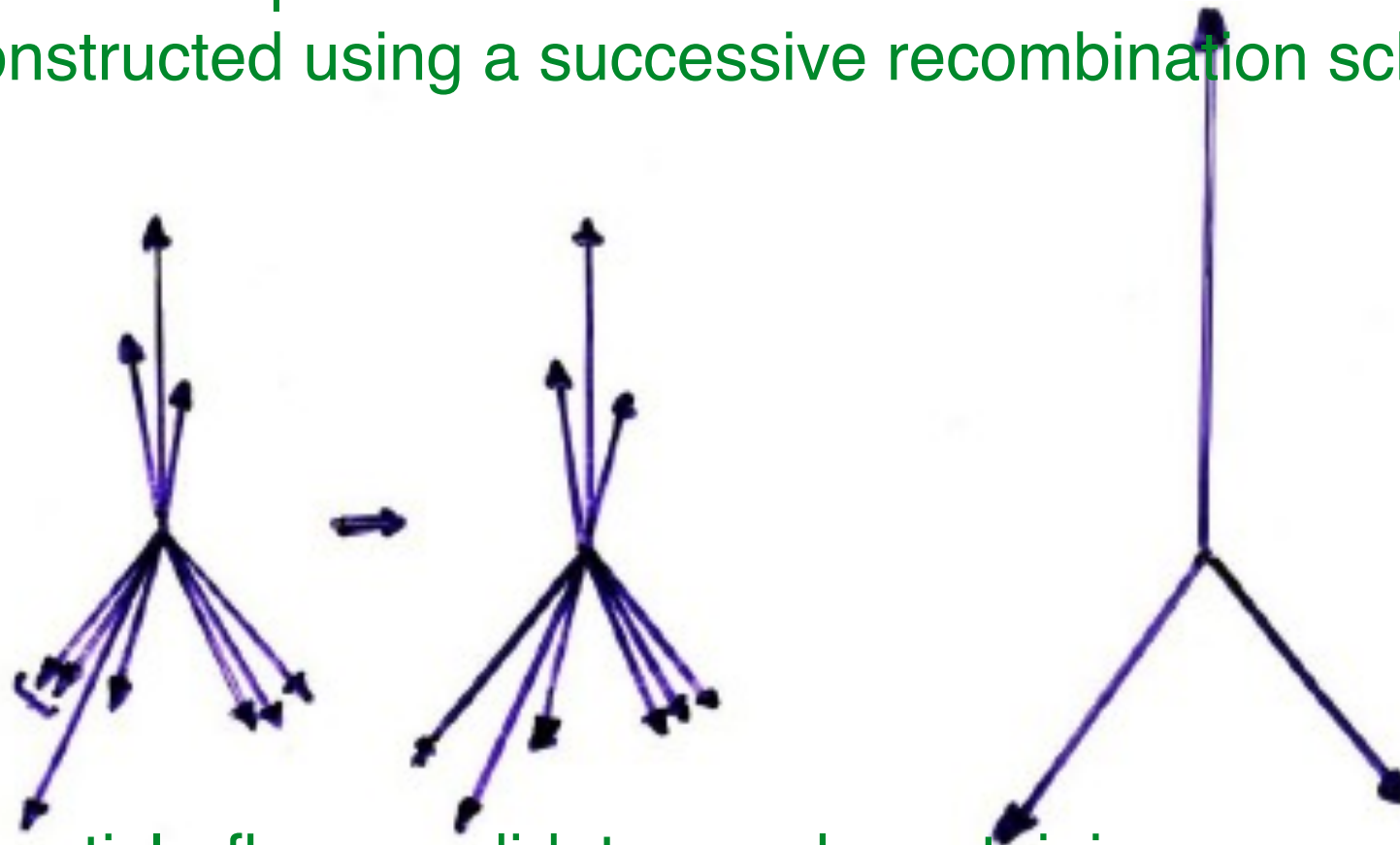
Combination of Detector Information

- Different particles behave differently in the detector.
- Combining information from different detectors, it is possible to identify individual particles and this process improves the quality of measurements
- Tracks reconstructed from the tracker information and clusters reconstructed using calorimetric measurements are combined in the Particle Flow approach:



Jet Reconstruction

- Stable hadrons observed in the interactions are due to fragmentation/hadronization of hard partons $\rightarrow$ Jets
- Jets are reconstructed using a successive recombination scheme:



- Start with N particle-flow candidates each containing one particle
- Compute ρ_{ij} , a distance measurement between i and j
- Find the pair i,j for which ρ_{ij} is the smallest
- Regroup $i,j \rightarrow l$. Remove i,j from the list of pflow candidates; insert l with its 4-momentum computed from i,j , recompute
- Repeat this process till all ρ_{ij} 's exceed a cut-off (jet resolution parameter)



Jets in Colliders



- Ambiguity lies in
 - Definition of ρ
 - Recombination of the four momenta of the two (pseudo)particles
- In jet algorithms for hadron colliders, an extra distance ρ_{iB} , the distance from the beam is computed in addition to the distance between two particles i, j (ρ_{ij}). The minimum is determined among all ρ_{ij} 's and ρ_{iB} 's. If the minimum distance found is between two particles, they are combined. If one of the ρ_{iB} 's is found to be the minimum the (pseudo)particle i is declared as a jet and is taken out of the list

Name	ρ_{ij}	ρ_{iB}	Recombination Scheme
k_T	$\min(k_{Ti}^2, k_{Tj}^2) \frac{\Delta R_{ij}^2}{R^2}$	k_{Ti}^2	$p_l = p_i + p_j$
Aachen-Cambridge	$\frac{\Delta R_{ij}^2}{R^2}$	1	$p_l = p_i + p_j$
Anti- k_T	$\min(k_{Ti}^{-2}, k_{Tj}^{-2}) \frac{\Delta R_{ij}^2}{R^2}$	k_{Ti}^{-2}	$p_l = p_i + p_j$

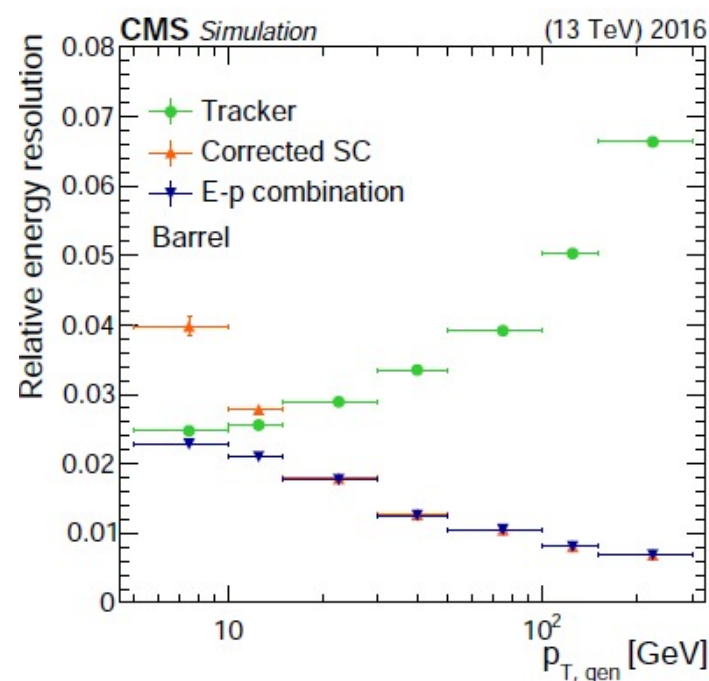
where $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$
 and R is a jet radius parameter and R in many applications



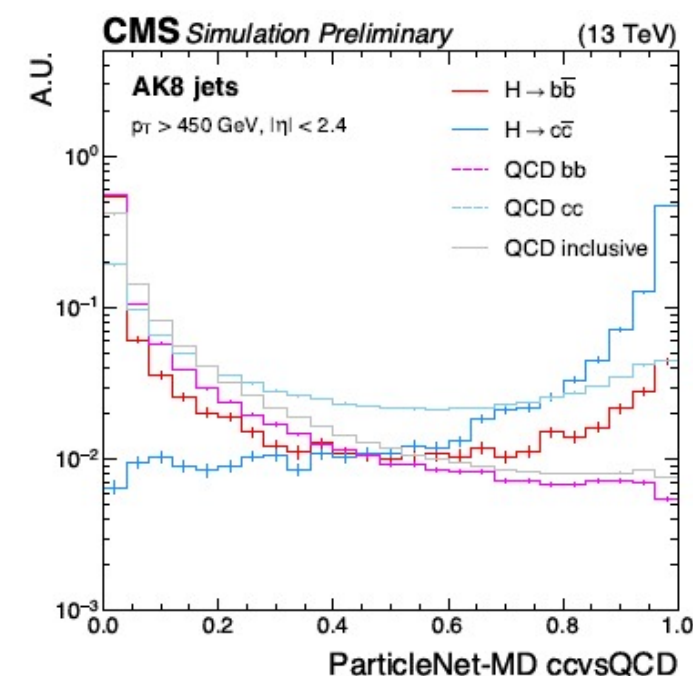
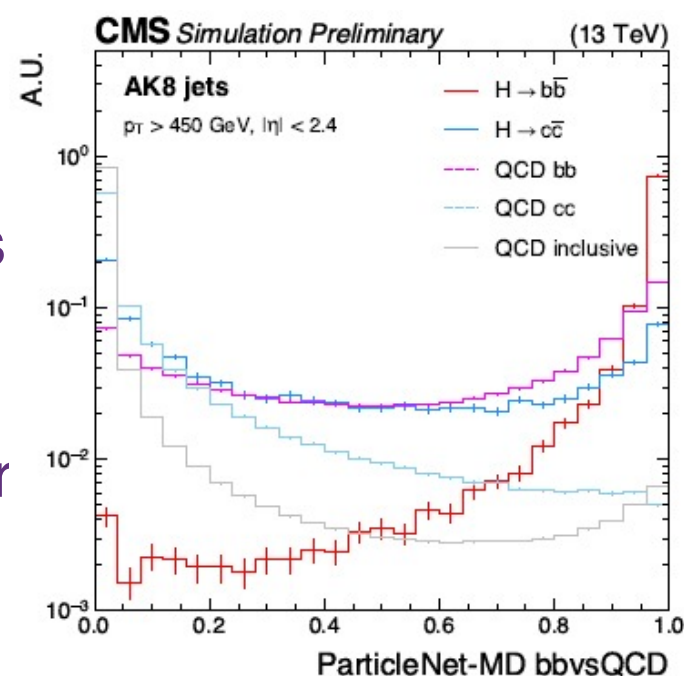
Multivariates and ML



- Categorization of particle type or jet classification often involves many independent or related variables which makes some contribution in the selection process. The best way to use these is to combine them using some machine learning technique
- A set of regression fits based on BDTs is applied to correct the energy of electron and photon candidates. The regression targets the ratio between reconstructed and true energy as well as the resolution.
- Identification of b-jet and c-jet makes use of DNN. The main features that distinguish a heavy boosted jet from light quark jets are the invariant mass and the distribution of particles within the jet.



Energy regress
resolution
improves the
energy of electr





Summary



- Extraction of physics results from a high-energy hadron machine is no joke. It requires a well-designed detector built with the latest technology, a sophisticated trigger and data acquisition system and utilisation of classical as well as modern software technology to hand the analyst a set of physics objects having the most precise attributes
- **Nothing is free in this world.**