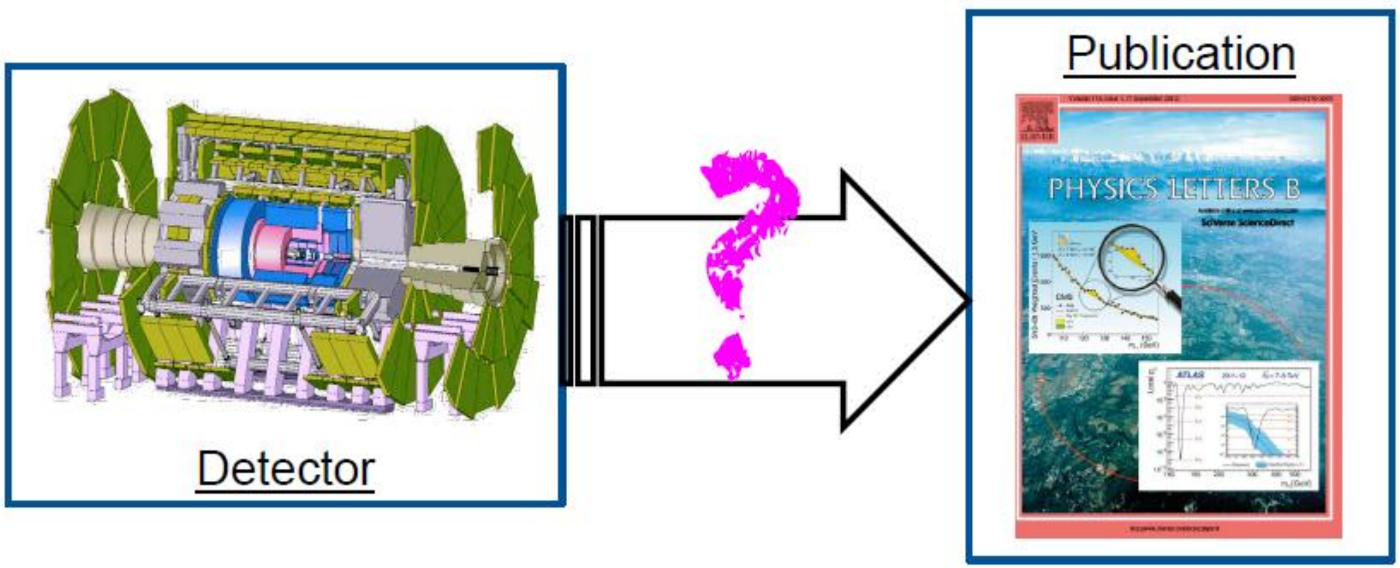
Introduction to particle physics: experimental part

From raw events to physics

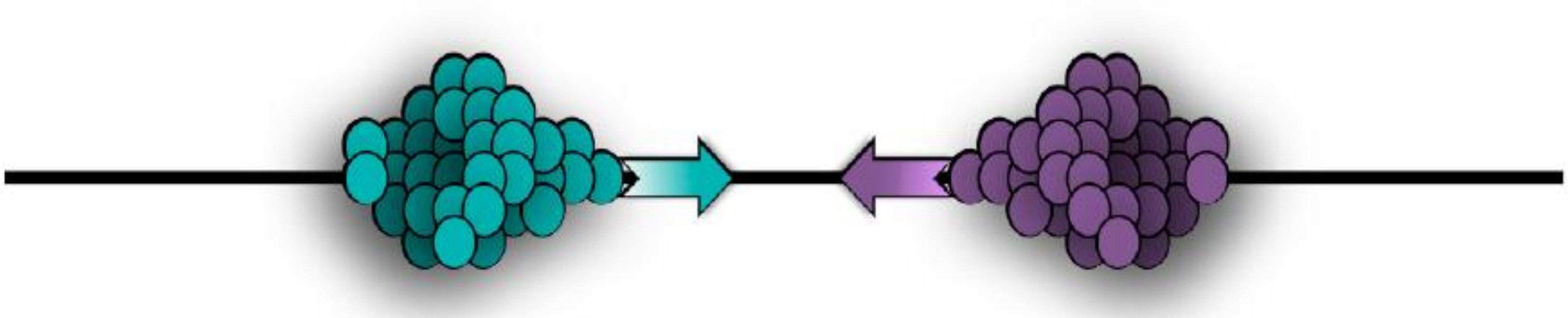
Large fraction of slides from A. Sfyrla lectures at CERN Summer School 2017

Prof. dr hab. Elżbieta Richter-Wąs

How do we deal with physics events from when they leave the detector till when they make it into our publications?

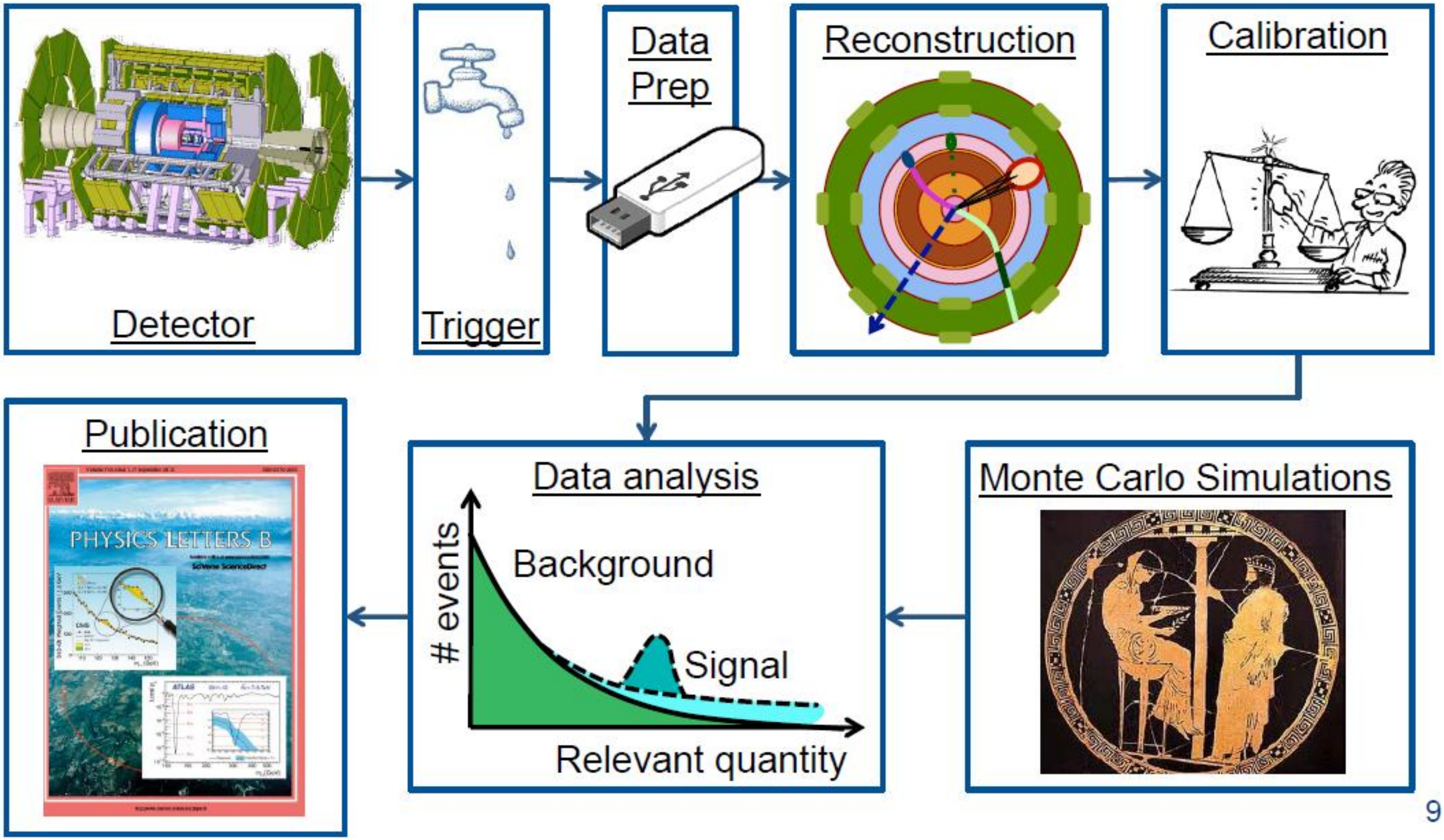


What is an event?



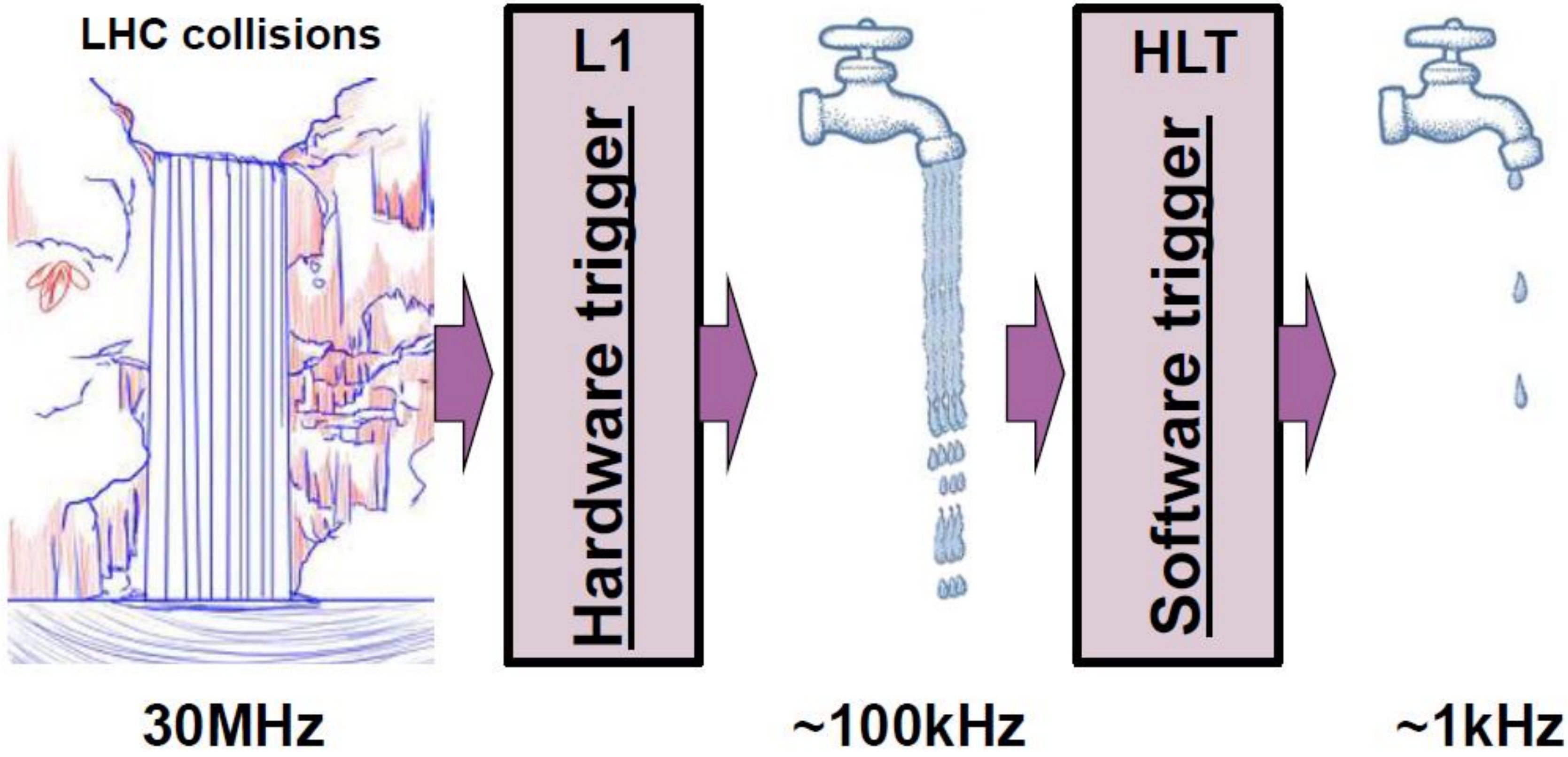
Proton bunches
>10^{11} protons/bunch
colliding at **13TeV** and at ~**30MHz in Run-2**
collided at **7/8TeV** and at ~**20MHz in Run-1**

An event’s lifetime

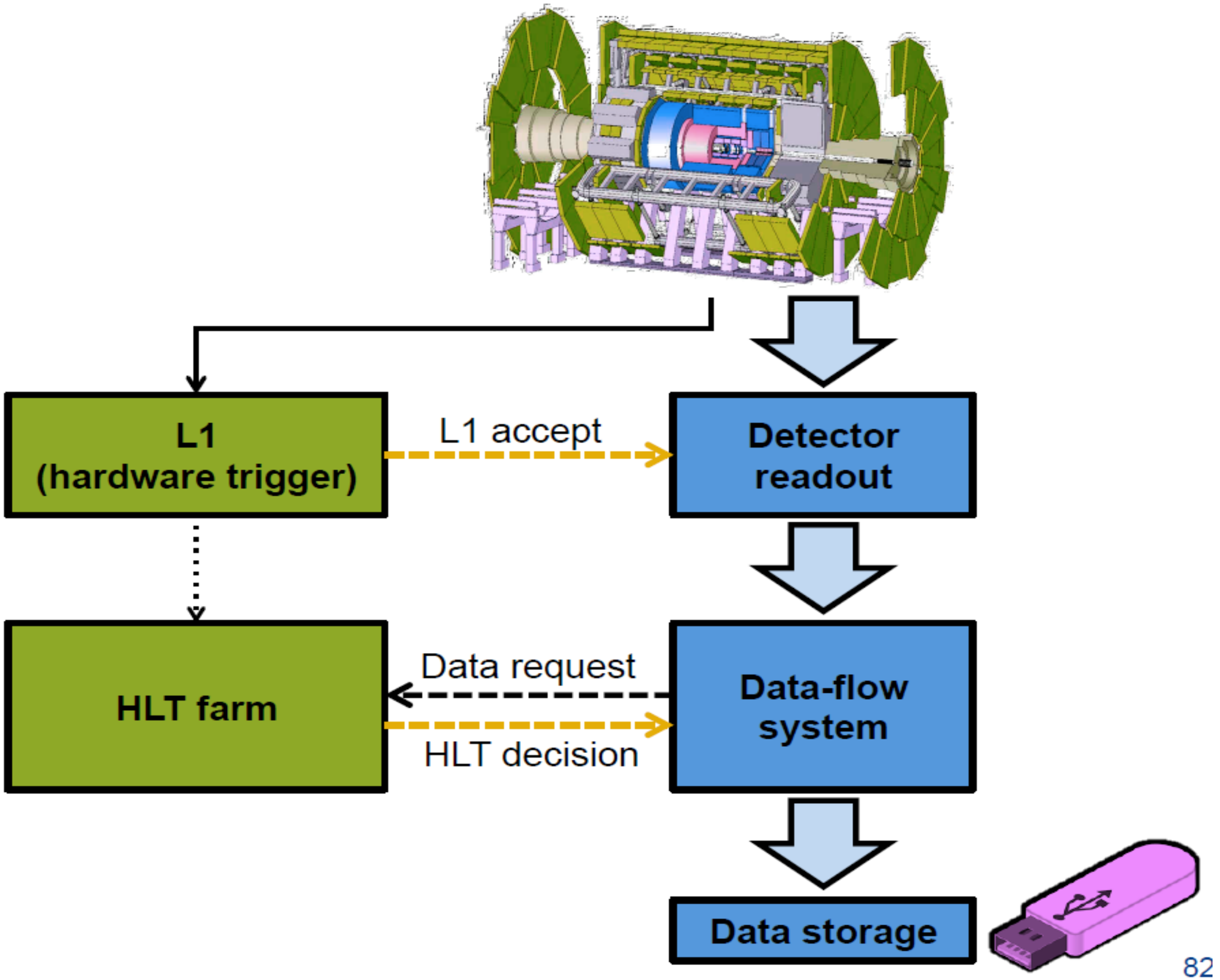


Triggering on physics

The ATLAS / CMS paradigm



The Data Acquisition (DAQ)



What does raw contain?

A simple example from the trigger on ATLAS (run1 data)

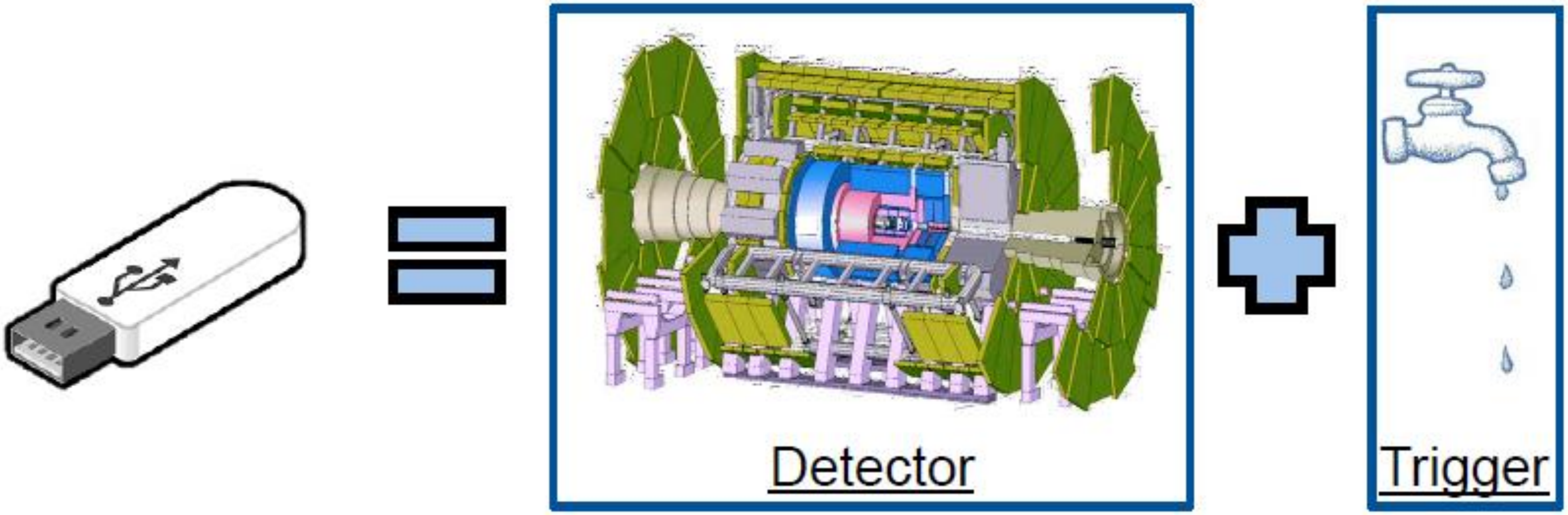
0x00000015	0x20000e3f	536874559	lvl1 trigger info[0]	L1 Trigger Bits Before Prescale
0x00000016	0x100000c0	268435648	lvl1 trigger info[1]	
0x00000017	0x8000043f	2147484735	lvl1 trigger info[2]	
0x00000018	0x00021007	135175	lvl1 trigger info[3]	
0x00000019	0x00000e10	3600	lvl1 trigger info[4]	
0x0000001a	0x00080000	524288	lvl1 trigger info[5]	
0x0000001b	0x02c00400	46138368	lvl1 trigger info[6]	
0x0000001c	0x00020001	131073	lvl1 trigger info[7]	
0x0000001d	0x00000816	2070	lvl1 trigger info[8]	L1 Trigger Bits After Prescale
0x0000001e	0x100000c0	268435648	lvl1 trigger info[9]	
0x0000001f	0x80000018	2147483672	lvl1 trigger info[10]	
0x00000020	0x00021001	135169	lvl1 trigger info[11]	
0x00000021	0x00000e10	3600	lvl1 trigger info[12]	
0x00000022	0x00000000	0	lvl1 trigger info[13]	
0x00000023	0x02c00400	46138368	lvl1 trigger info[14]	
0x00000024	0x00020000	131072	lvl1 trigger info[15]	
0x00000025	0x00000010	16	lvl1 trigger info[16]	L1 Trigger Bits After Veto
0x00000026	0x00000000	0	lvl1 trigger info[17]	
0x00000027	0x00000008	8	lvl1 trigger info[18]	
0x00000028	0x00000000	0	lvl1 trigger info[19]	
0x00000029	0x00000810	2064	lvl1 trigger info[20]	
0x0000002a	0x00000000	0	lvl1 trigger info[21]	
0x0000002b	0x00000400	1024	lvl1 trigger info[22]	
0x0000002c	0x00000000	0	lvl1 trigger info[23]	

What does raw contain?

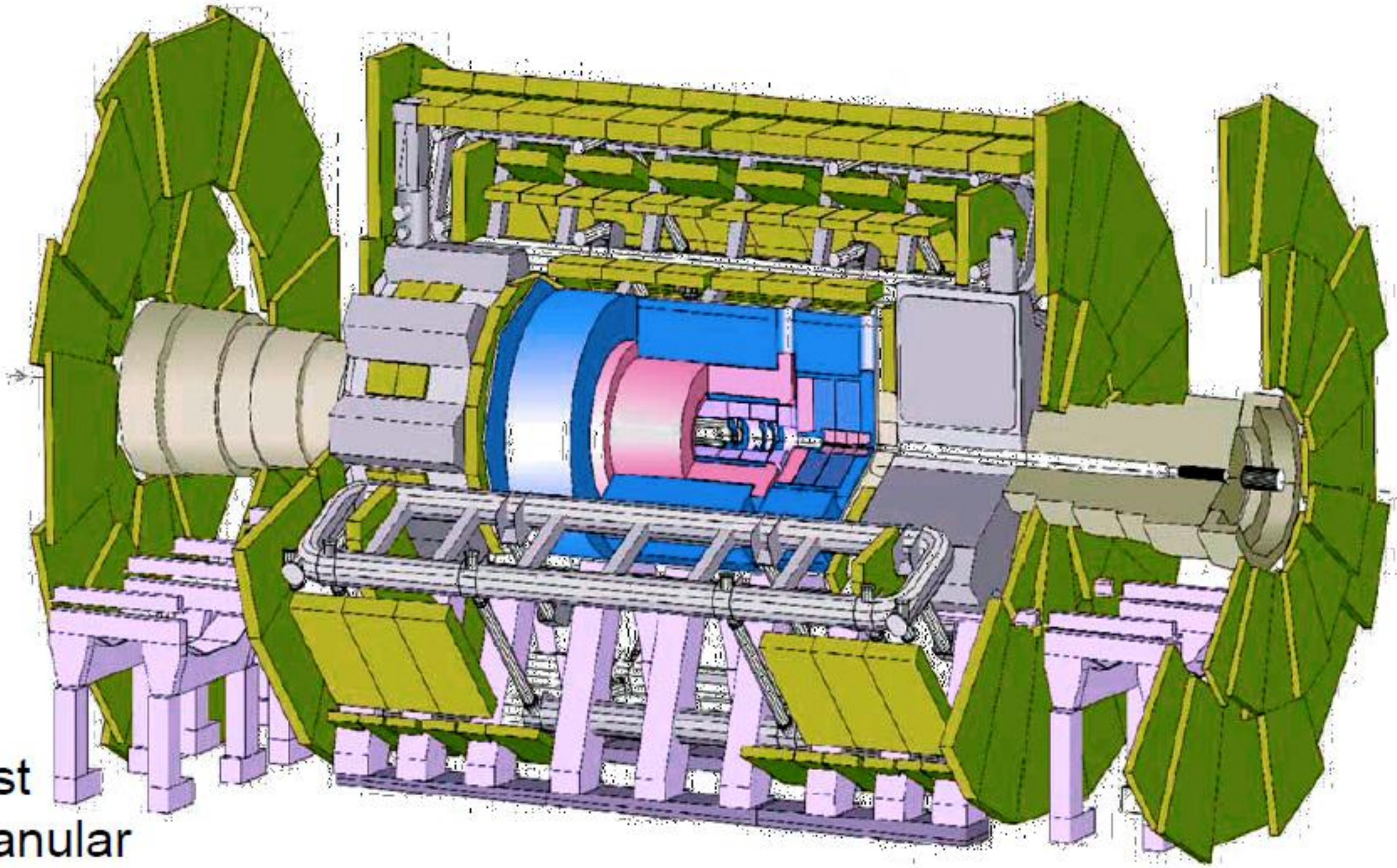
0x00000015	0x20000e3f	536874559	lvl1 trigger info[0]
0x00000016	0x100000c0	268435648	lvl1 trigger info[1]
0x00000017	0x8000043f	2147484735	lvl1 trigger info[2]
0x00000018	0x00021007	135175	lvl1 trigger info[3]
0x00000019	0x00000e10	3600	lvl1 trigger info[4]
0x0000001a	0x00080000	524288	lvl1 trigger info[5]
0x0000001b	0x02c00400	46138368	lvl1 trigger info[6]
0x0000001c	0x00020001	131073	lvl1 trigger info[7]
0x0000001d	0x00000816	2070	lvl1 trigger info[8]
0x0000001e	0x100000c0	268435648	lvl1 trigger info[9]
0x0000001f	0x80000018	2147483672	lvl1 trigger info[10]
0x00000020	0x00021001	135169	lvl1 trigger info[11]
0x00000021	0x00000e10	3600	lvl1 trigger info[12]
0x00000022	0x00000000	0	lvl1 trigger info[13]
0x00000023	0x02c00400	46138368	lvl1 trigger info[14]
0x00000024	0x00020000	131072	lvl1 trigger info[15]
0x00000025	0x00000010	16	lvl1 trigger info[16]
0x00000026	0x00000000	0	lvl1 trigger info[17]
0x00000027	0x00000008	8	lvl1 trigger info[18]
0x00000028	0x00000000	0	lvl1 trigger info[19]
0x00000029	0x00000810	2064	lvl1 trigger info[20]
0x0000002a	0x00000000	0	lvl1 trigger info[21]
0x0000002b	0x00000400	1024	lvl1 trigger info[22]
0x0000002c	0x00000000	0	lvl1 trigger info[23]

- More than 300K such words in each event, corresponding to the full data from all the detector components.
- Data size: 1-1.5MB / event depending on the compression. Pretty consistent between ATLAS and CMS.
- **Challenge: make sense out of all these numbers!!**

What does raw contain?

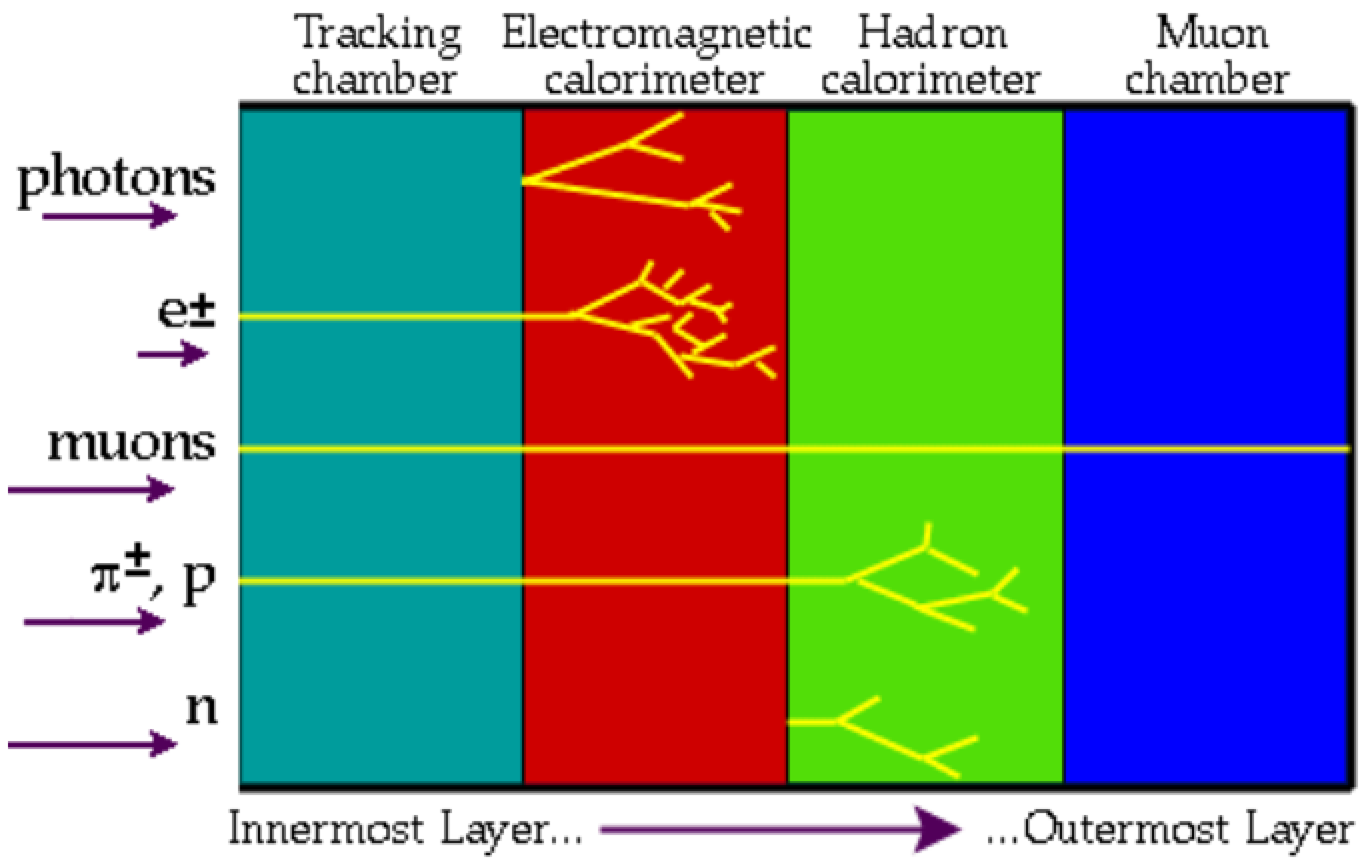


A detector (e.g. ATLAS)

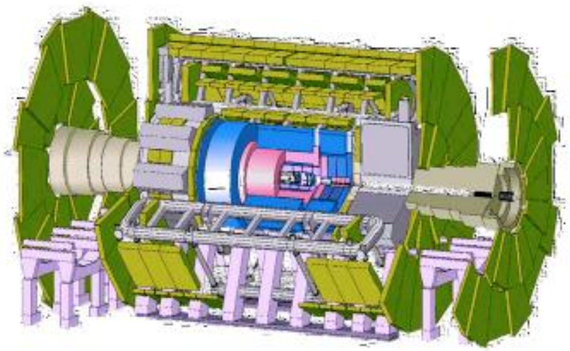


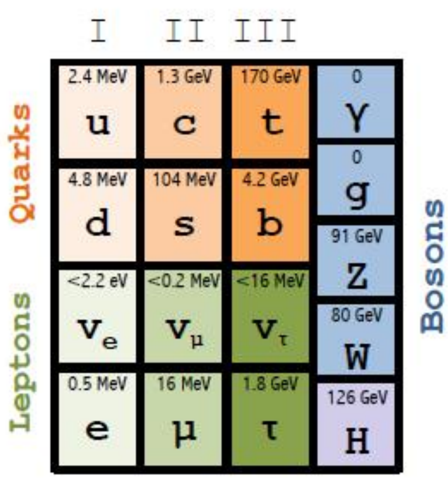
- ✓ Fast
- ✓ Granular
- ✓ Resistant to radiation

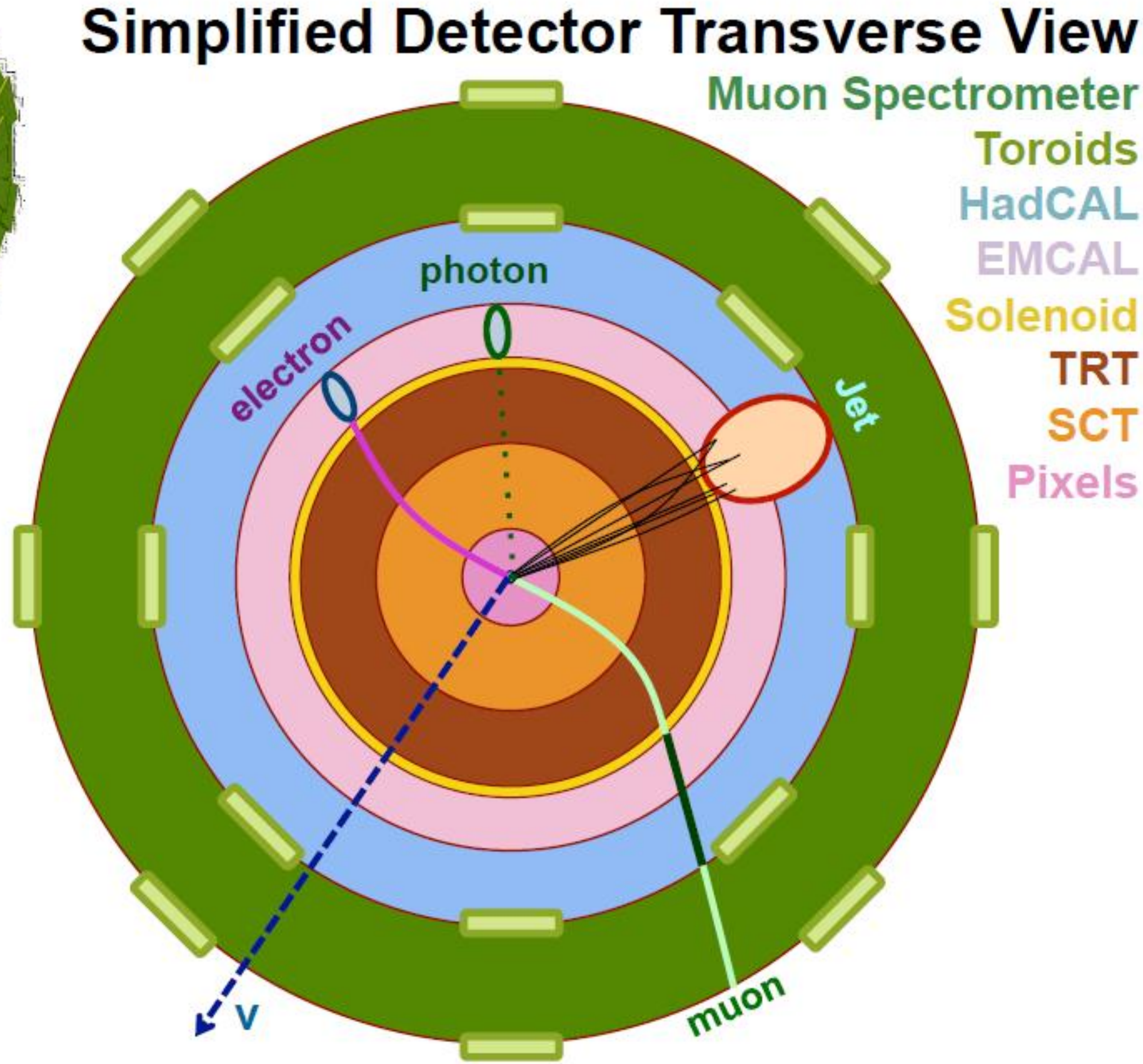
Particles through matter



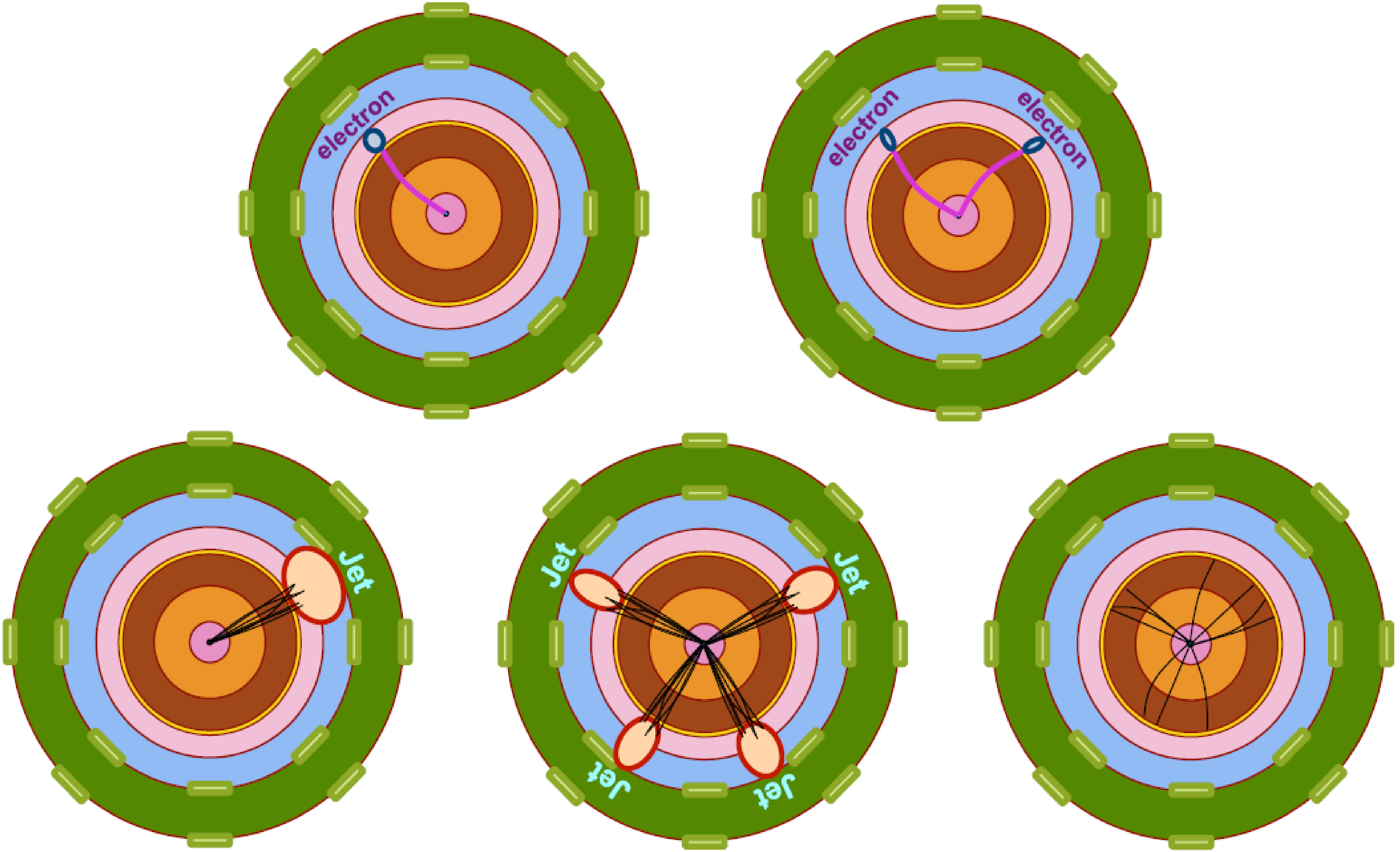
A detector (eg. ATLAS)



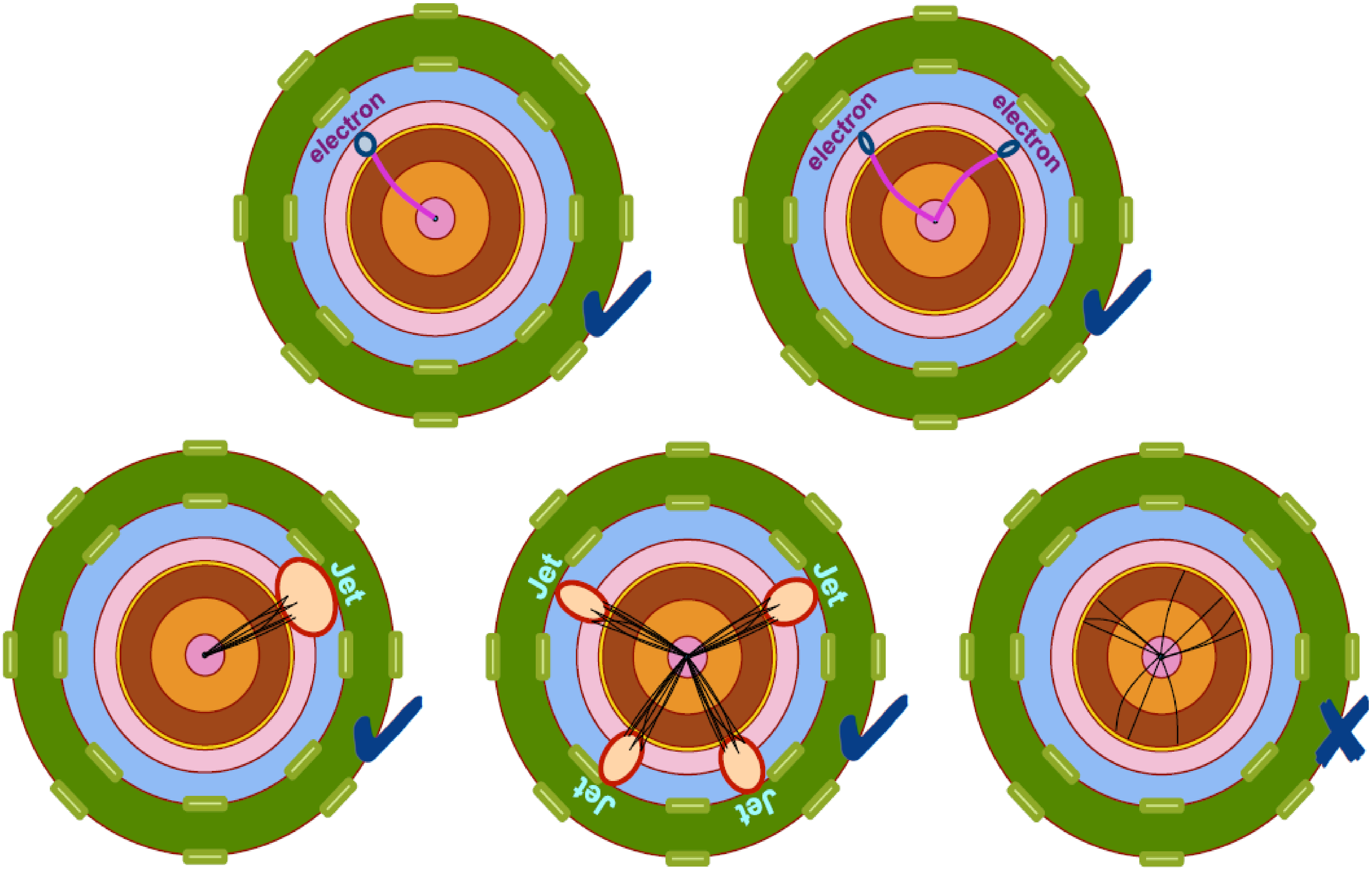




Online reconstruction



Triggering on physics



Streaming

- Streaming is based on trigger decisions at all stages
- The Raw Data physics streams are generated at the HLT output level

Debug Streams

events for which a trigger decision has not been made, because of failures in parts of the online system

Physics Streams

data for physics analyses

Express Stream

full events for fast reconstruction

Calibration Streams

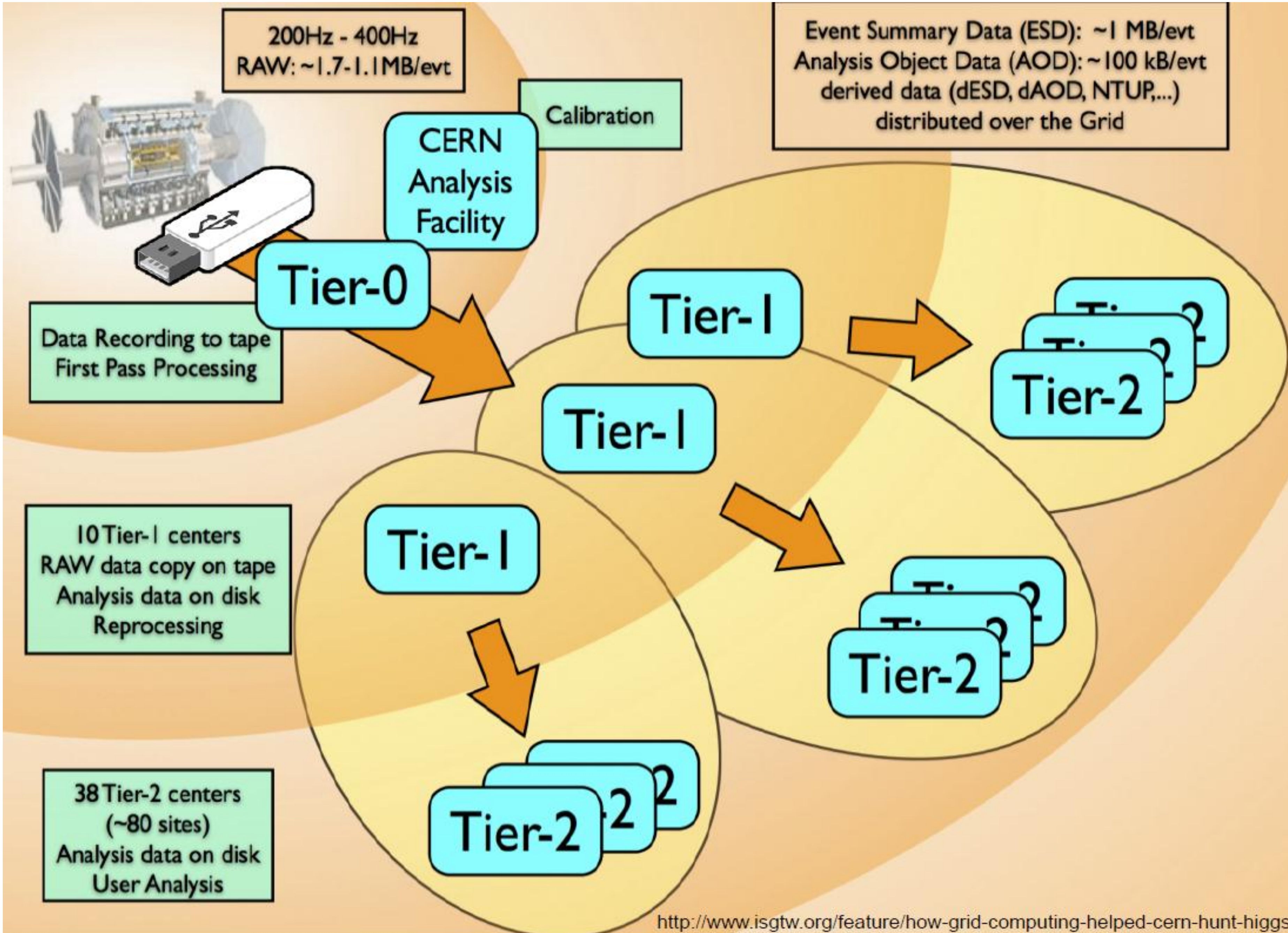
events delivering the minimum amount of information for detector calibrations at high rate

Huge amount of data …

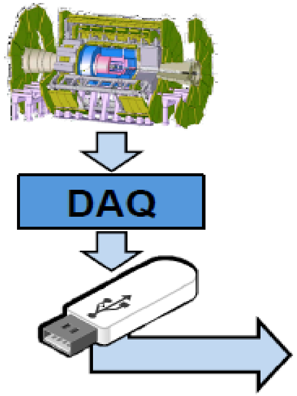
LHC delivered billions of recorded collision events to the LHC experiments from proton-proton and proton-lead collisions in the Run 1 period (2009-2013).

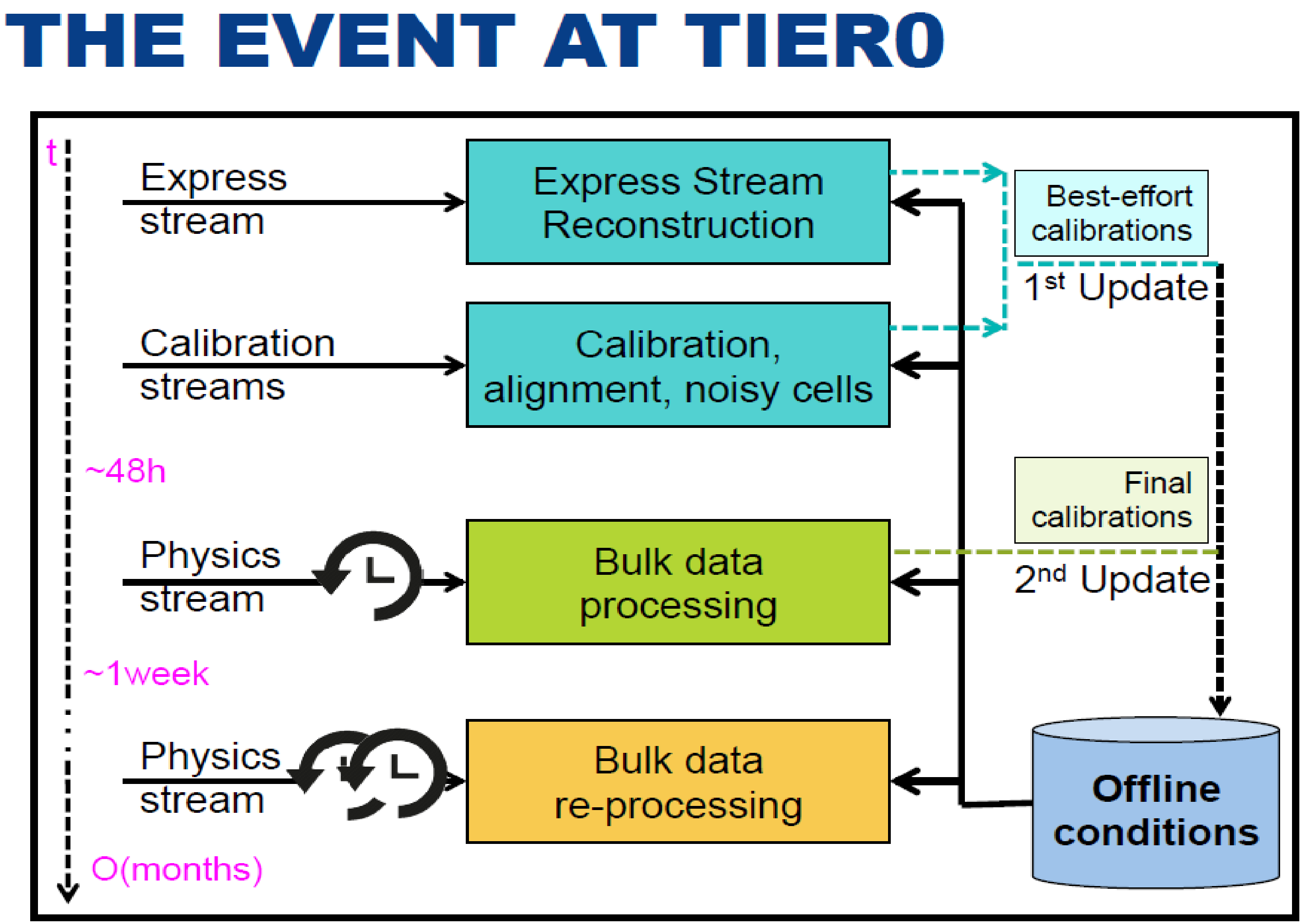
- This translates to ~ 100 PB of data recorded at CERN.

The challenge how to process and analyze the data and produce timely physics results was substantial but in the end resulted in a great success.


200Hz - 400Hz
RAW: ~1.7-1.1MB/evt
Calibration
CERN Analysis Facility
Tier-0
Data Recording to tape
First Pass Processing
Event Summary Data (ESD): ~1 MB/evt
Analysis Object Data (AOD): ~100 kB/evt
derived data (dESD, dAOD, NTUP,...)
distributed over the Grid
Tier-1
Tier-1
Tier-1
Tier-2
Tier-2
Tier-2
10 Tier-1 centers
RAW data copy on tape
Analysis data on disk
Reprocessing
38 Tier-2 centers
(~80 sites)
Analysis data on disk
User Analysis
http://www.isgtw.org/feature/how-grid-computing-helped-cern-hunt-higgs

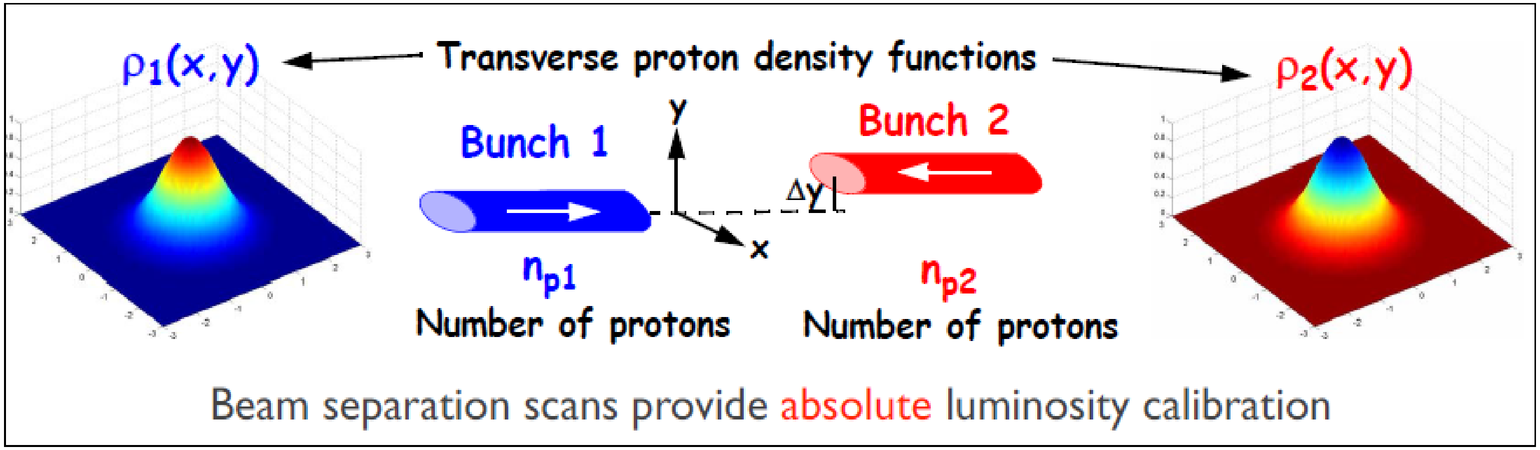
Huge amount of data …





Luminosity determination

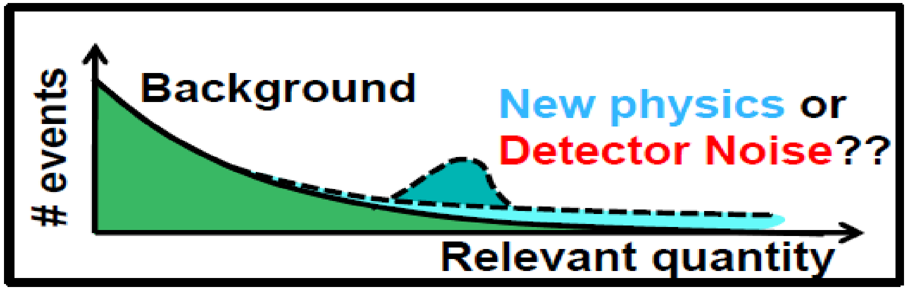
- A measurement of the number of collisions per cm^2 and second.
- Multiple methods used for determining luminosity: reducing uncertainties.
- Normalization is done with beam-separation scan (Van-der-Meer scan). Requires careful control of beam parameters.



From http://cds.cern.ch/record/1490292/files/ATL-DAPR-SLIDE-2012-627.pdf

- Result: luminosity measurement with very small uncertainties (order of few %) with very fast turn-around time.

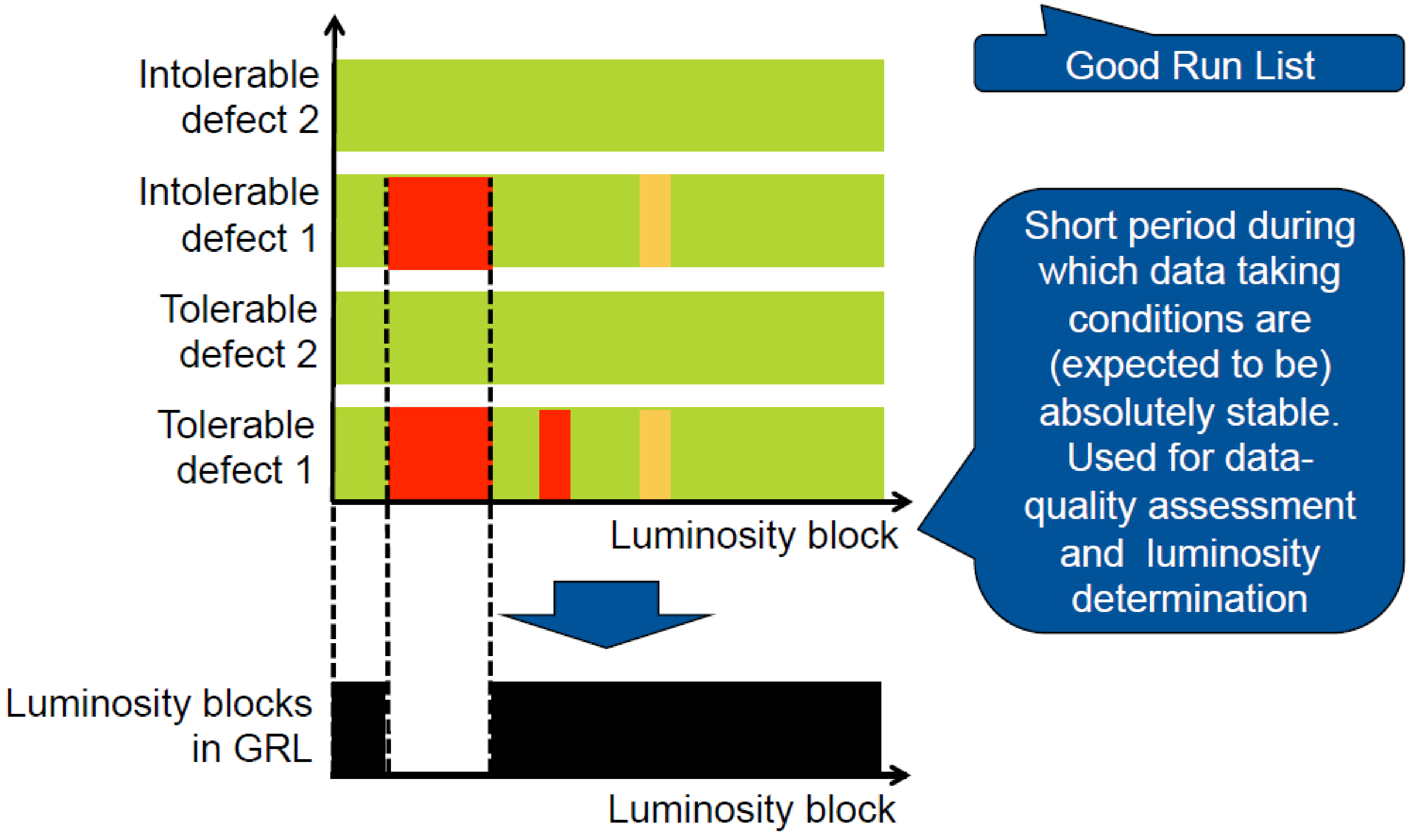
Data quality



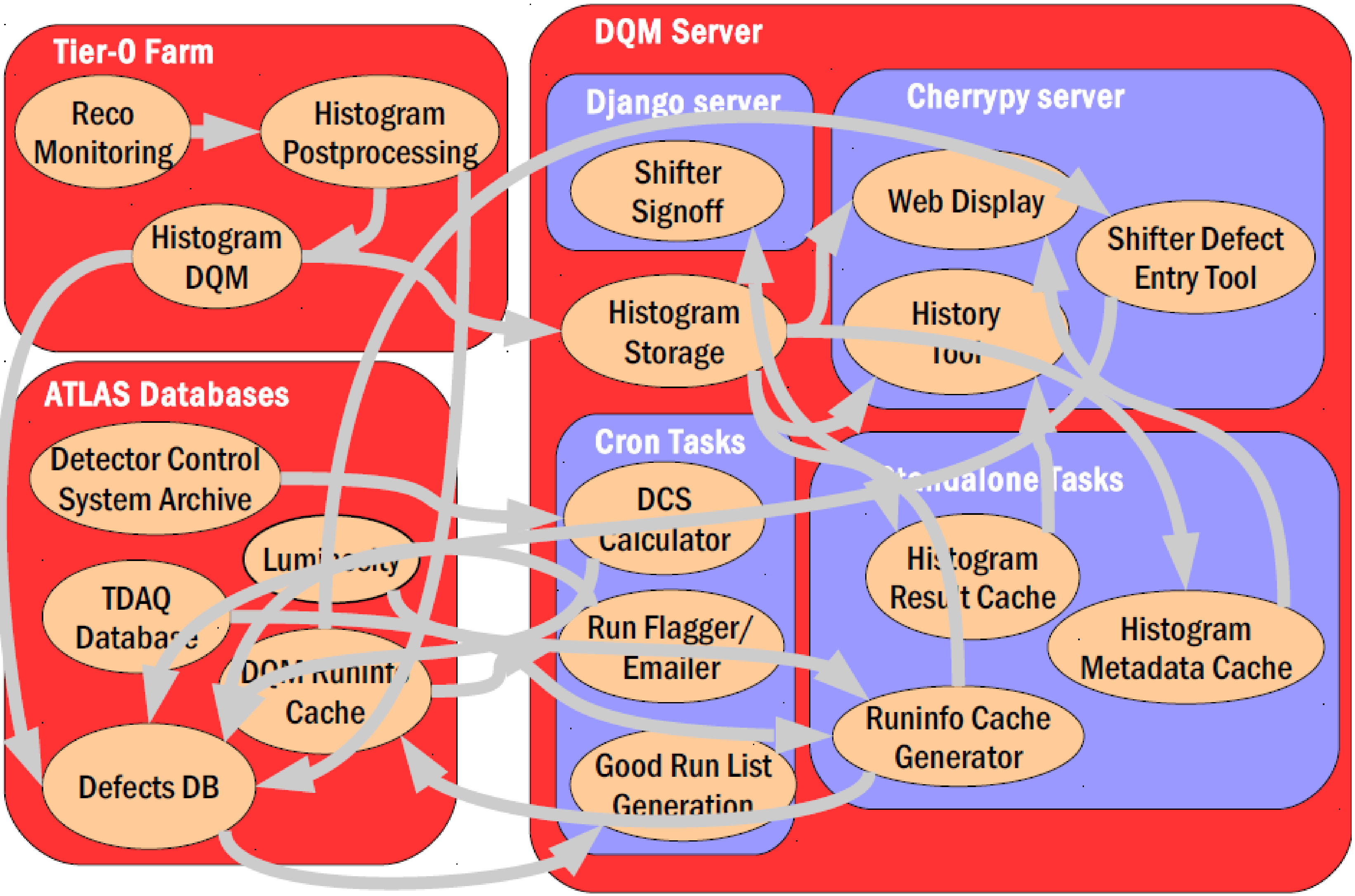
The data we analyze has to follow norms of quality such that our results are trustable.

- **Online: Fast monitoring of detector performance during data taking, using dedicated stream, "express stream".**
- **Offline: More thorough monitoring at two instances:**
  - Express reconstruction; fast turn-around.
  - Prompt reconstruction: larger statistics.
- **What is monitored?**
  - Noise in the detector.
  - Reconstruction (tracks, clusters, combined objects, resolution and efficiency).
  - Input rate of physics.
  - All compared to reference histograms of data that has been validated as "good".

Data quality and „GRL"

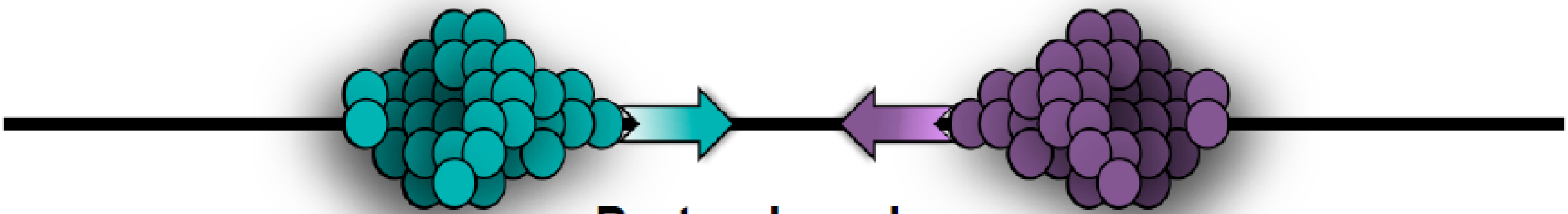


Data quality



https://cds.cern.ch/record/2008725/files/ATL-SOFT-SLIDE-2015-179.pdf

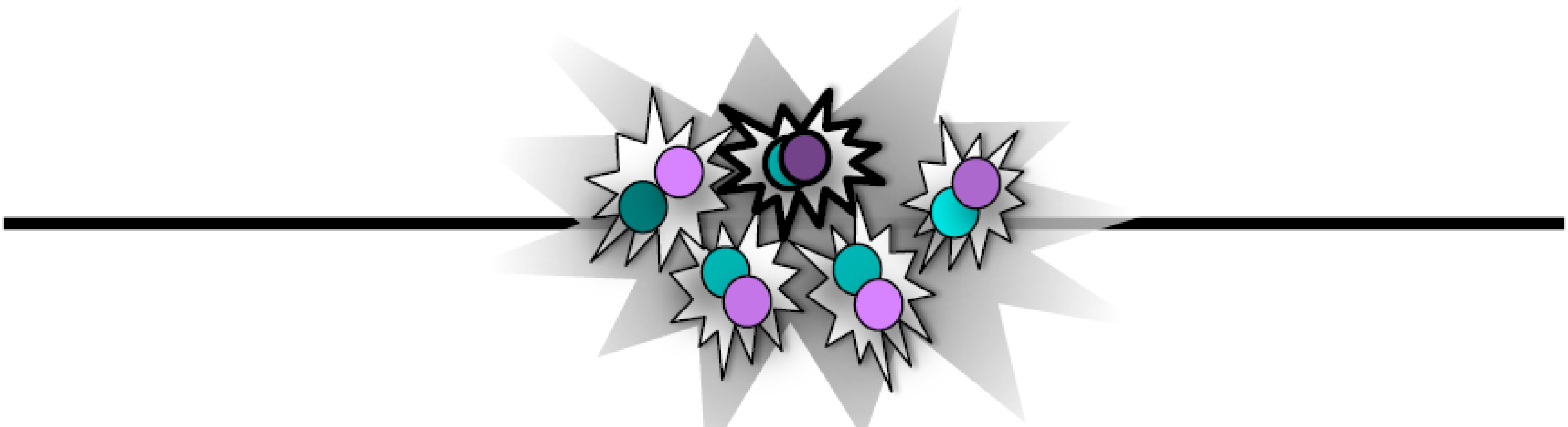
Pile-up



Proton bunches

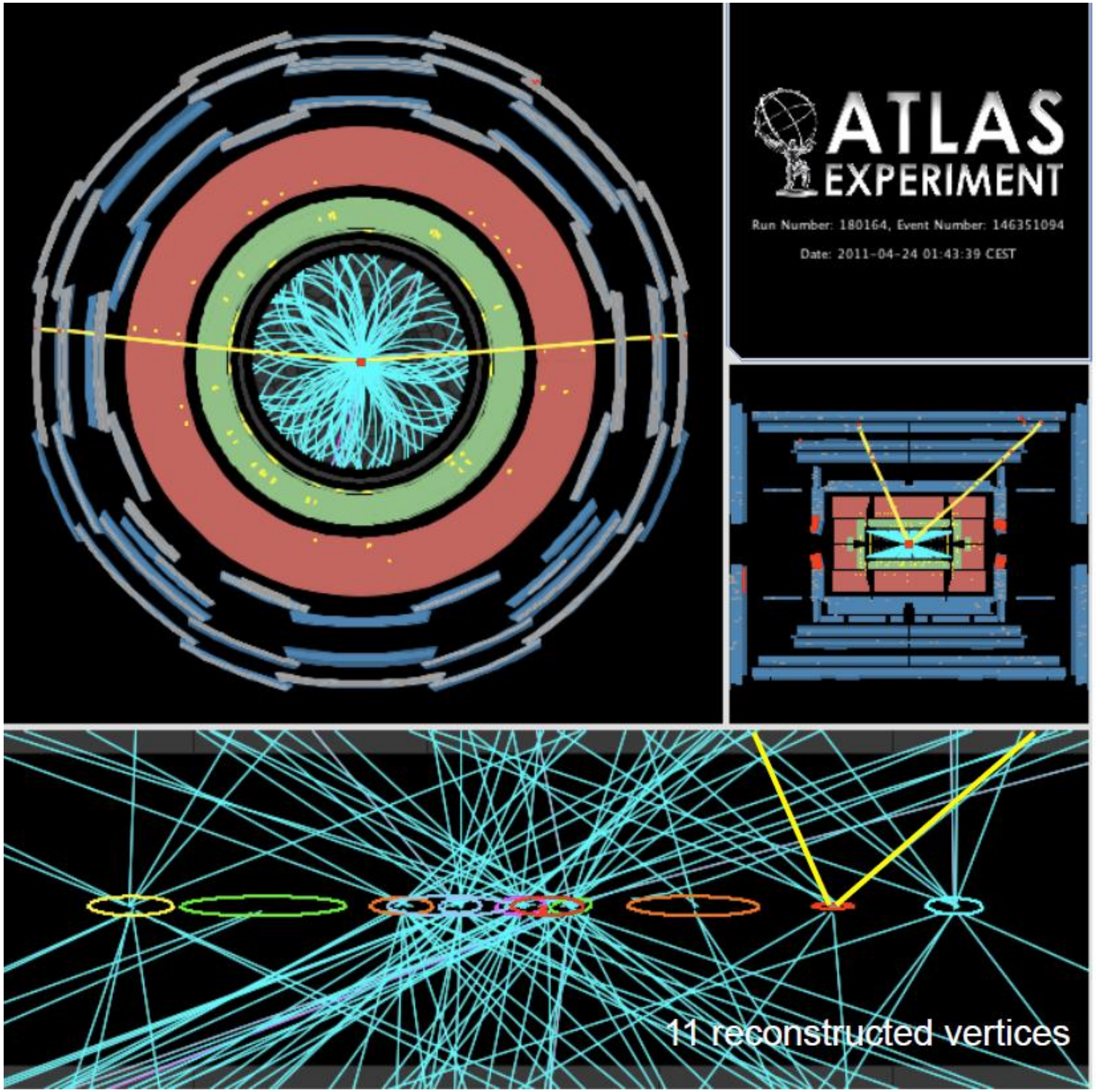
>10^{11} protons/bunch

(colliding at ~**30MHz in run2)**



~25 p-p collisions / bunch crossing

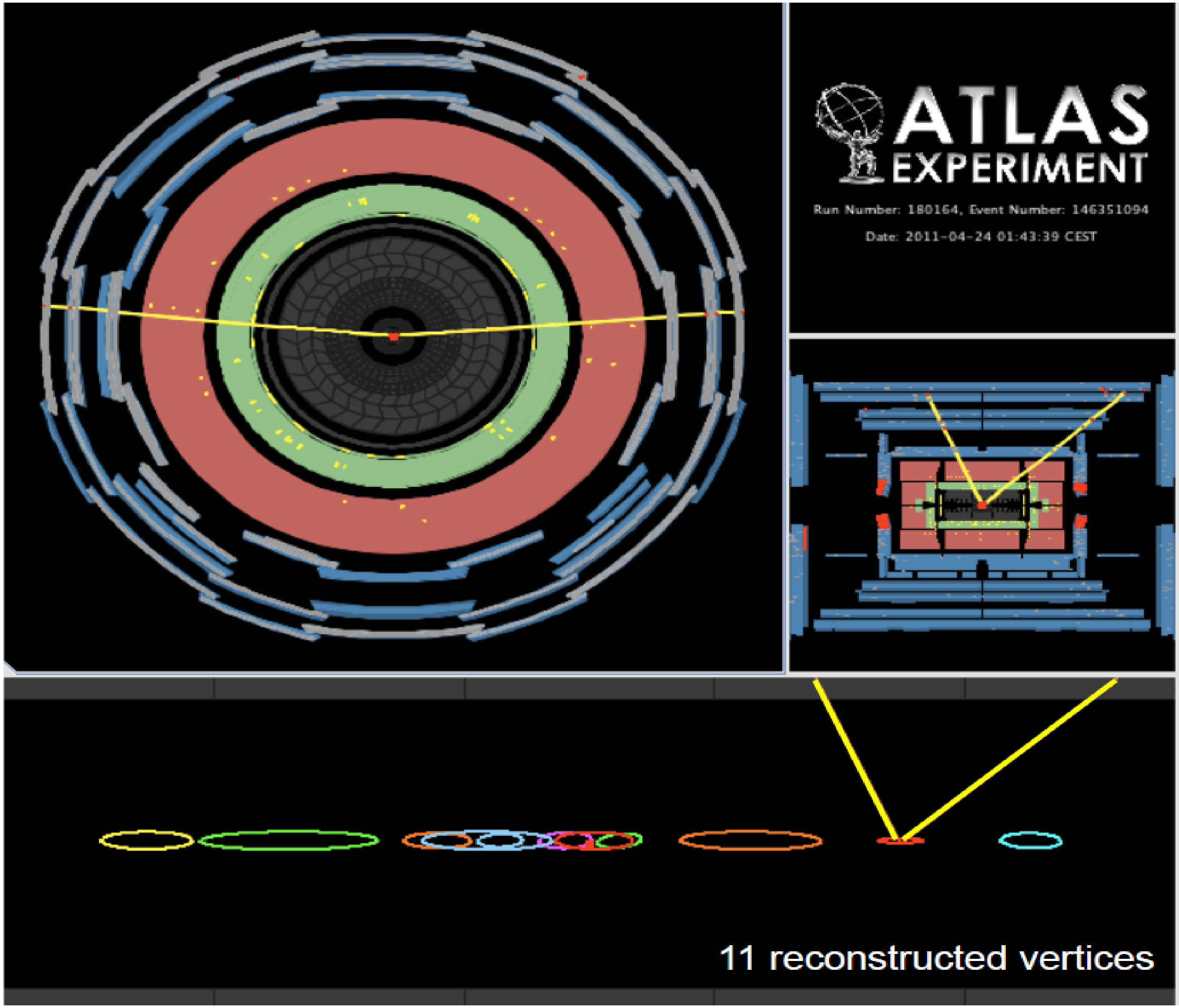
Pile-up



Z->μμ event;
2011 data.

Track pT > 0.5 GeV

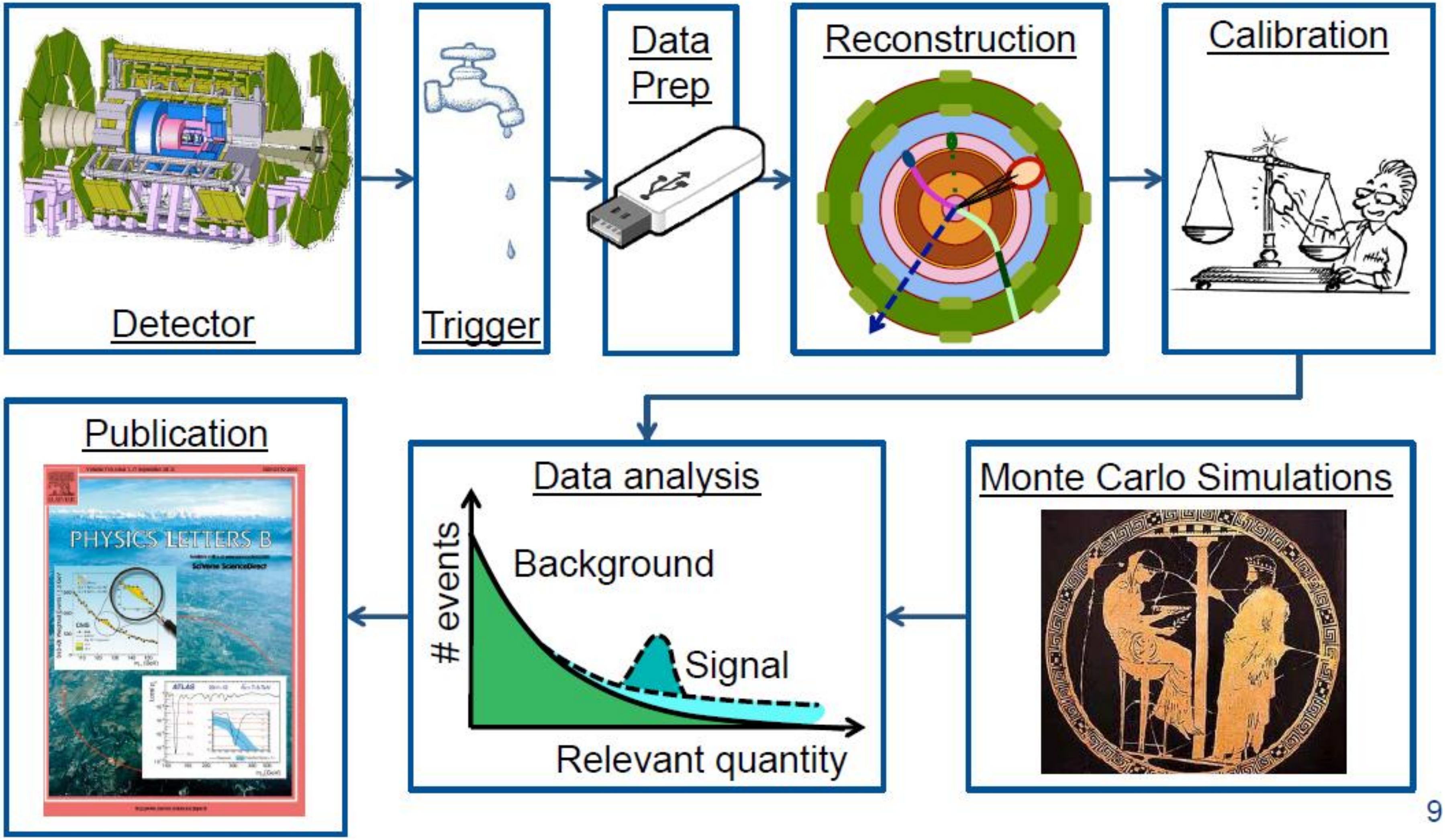
Pile-up



Z->μμ event;
2011 data.

Track pT > 10 GeV

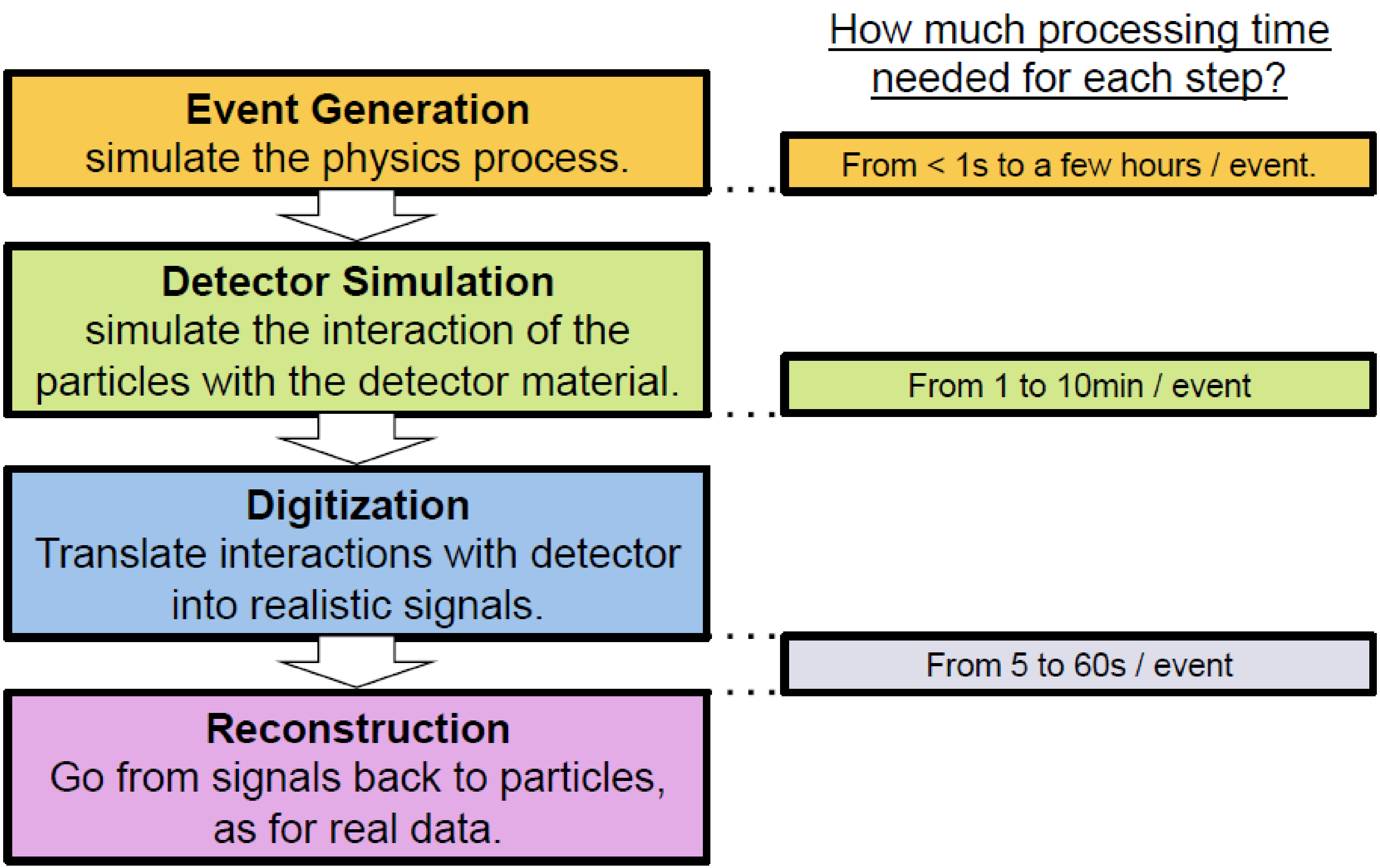
An event's lifetime



Monte Carlo simulation

- **We only build one detector.**
  - How do we compromise physics due to detector design?
  - How would a different detector design affect measurements?
  - How does the detector behave to radiation?
- **In the detectors we only measure voltages, currents, times.**
  - It's an *interpretation* to say that such-and-such particle caused such-and-such signature in the detector.
  - Simulating the detector behavior we correct for inefficiencies, inaccuracies, unknowns.
- **We need a theory to tell us what we expect and to compare our data against.**
- **A good simulation is the way to demonstrate to the world that we understand the detectors and the physics we are studying.**

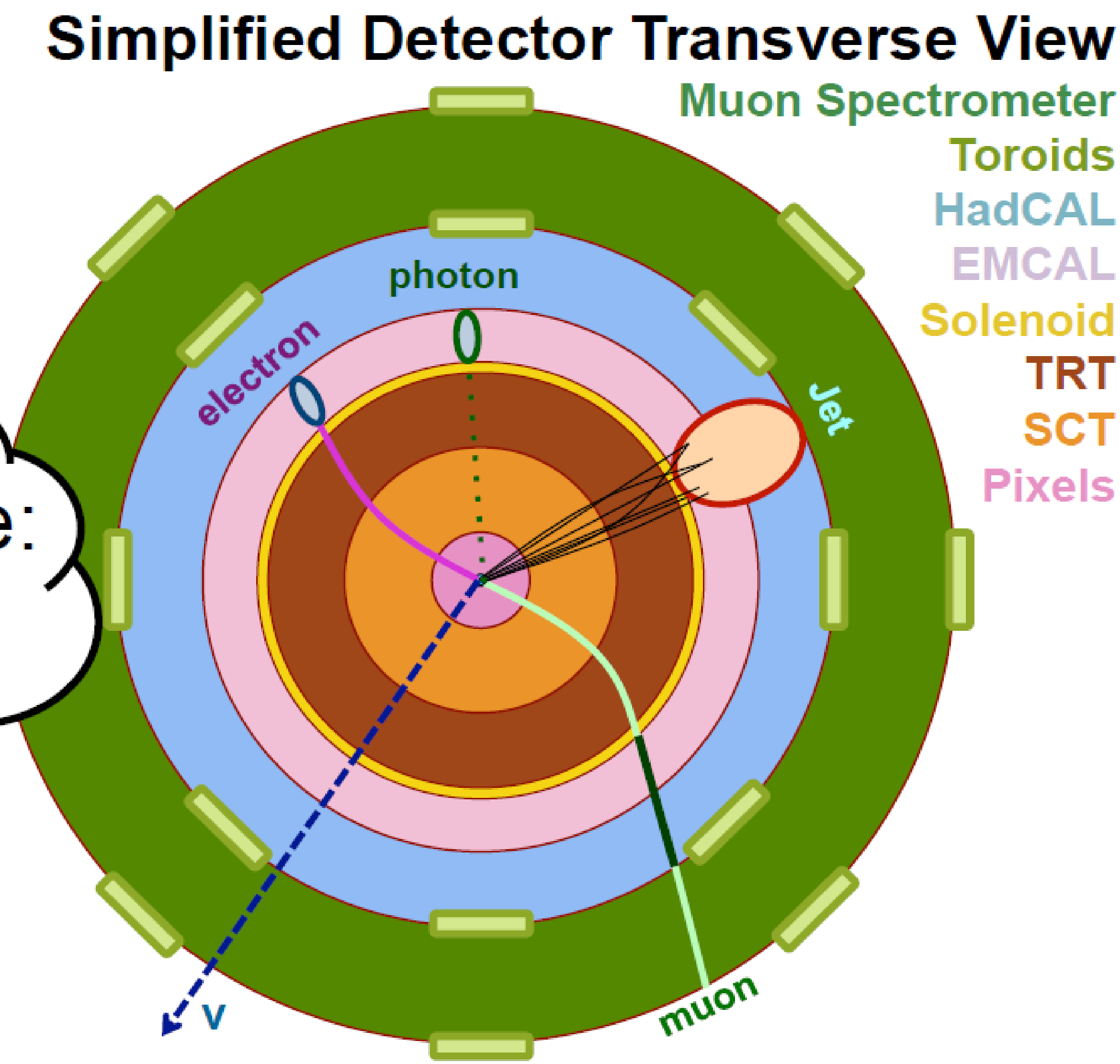
Monte Carlo production chain



What do we reconstruct?

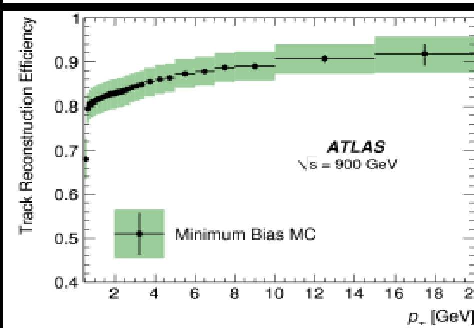
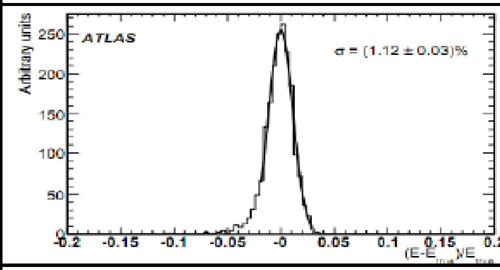
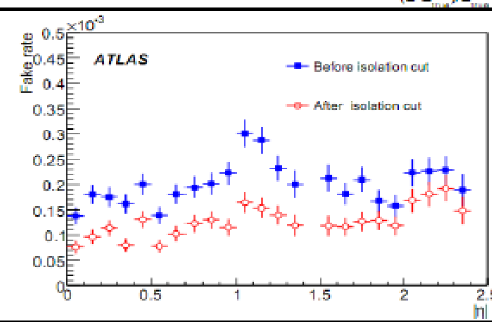
Tracks and Clusters

Combining those: "objects" ("particles")

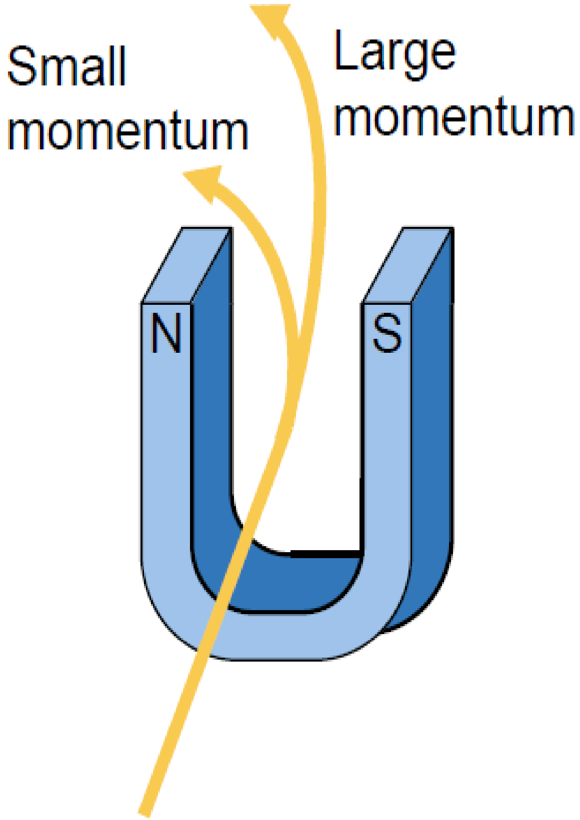


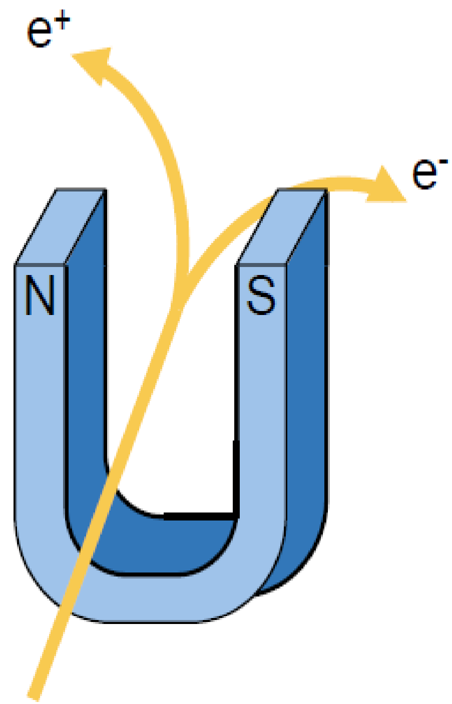
Reconstruction - figures of merit

“**true**” quantity:
quantity at MC generator level.

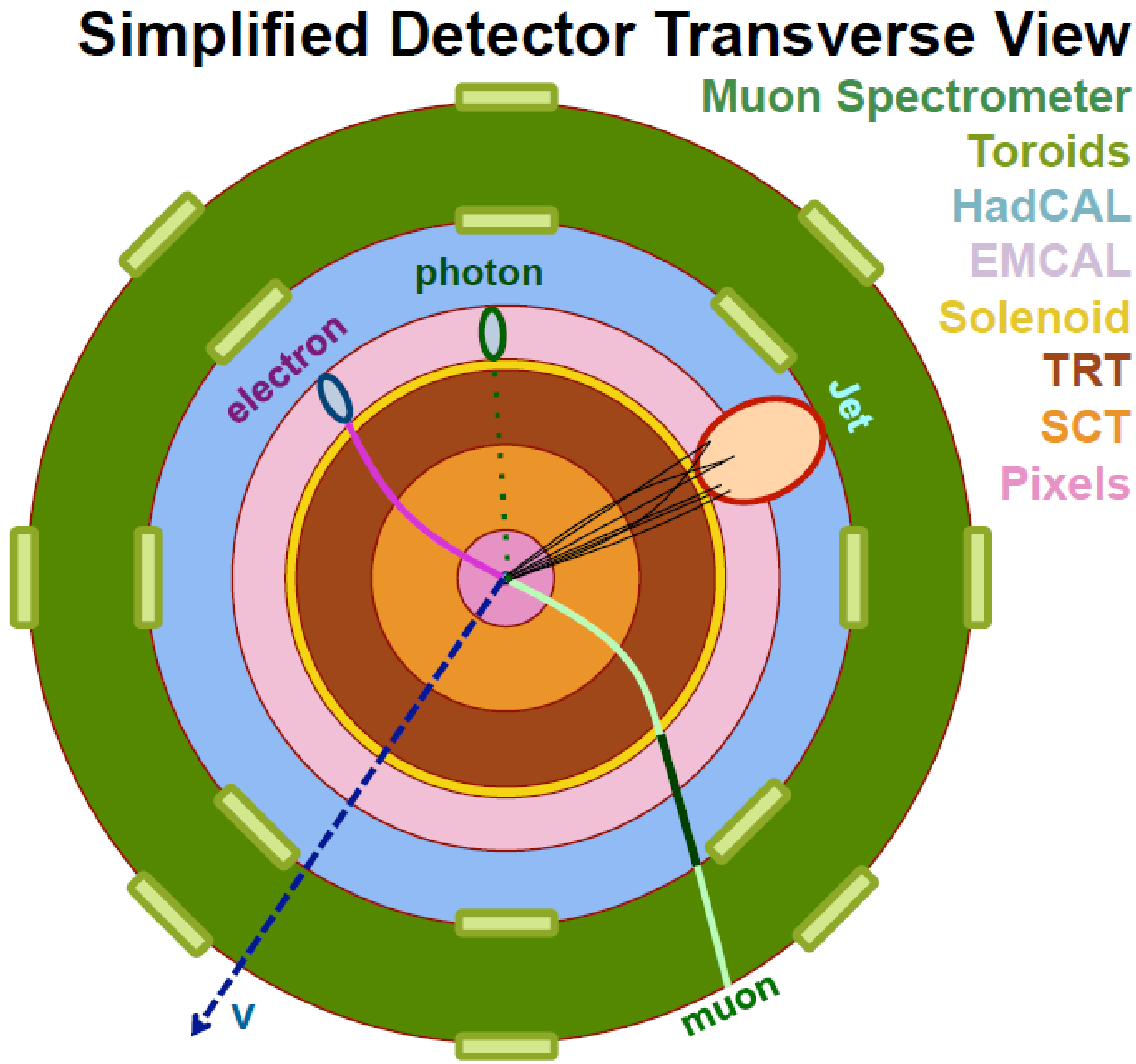
	Definition	Example		Needs be:
Efficiency	how often do we reconstruct the object	tracking efficiency = (number of reconstructed tracks) / (number of true tracks)		High
Resolution	how accurately do we reconstruct the quantity	energy resolution = (measured energy – true energy) / (true energy)		Good
Fake rate	how often we reconstruct a different object as the object we are interested in	a jet faking an electron, fake rate = (Number of jets reconstructed as an electron) / (Number of jets)		Low

Why do we need magnetic field?





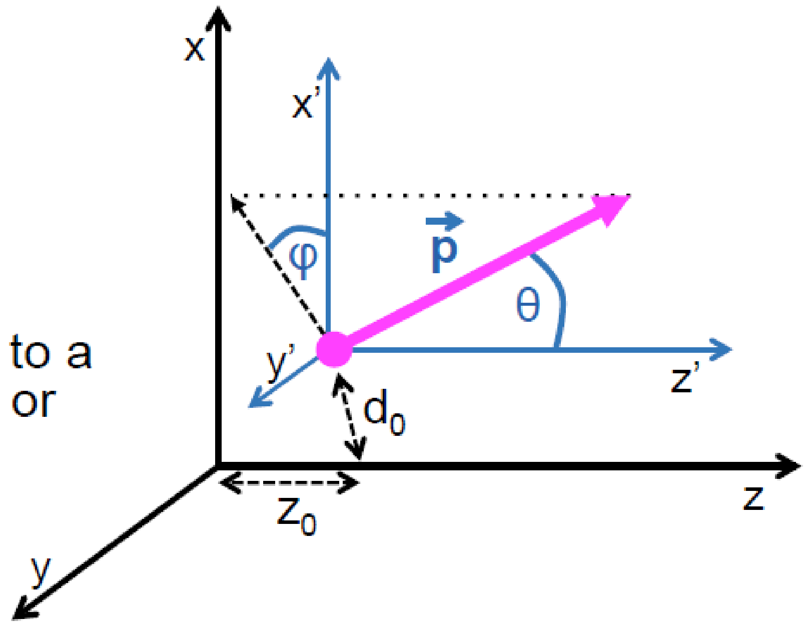
What do we reconstruct?



Tracking in a nutshell

- **A track represents a measurement of a charged particle that leaves a trajectory as it passes through the detector.**

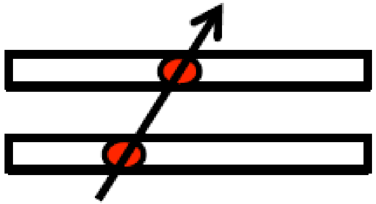
- **For a track we measure:**
  - Its momentum;
  - It's direction;
  - Its charge;
  - Its "perigee": the closest point to a reference line, transverse (d_0) or longitudinal (z_0).



- **Tracks are key ingredients of most of particle reconstruction.**

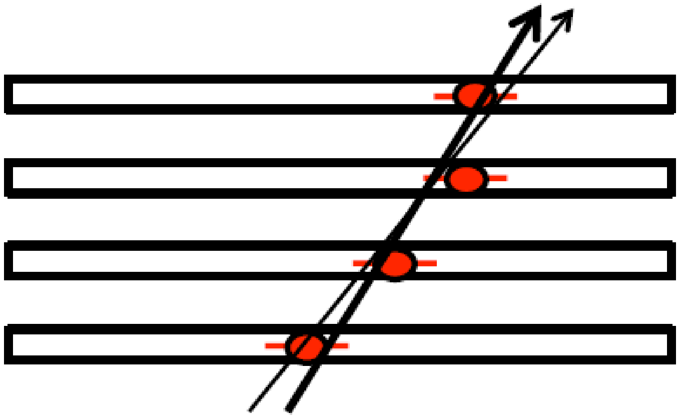
Tracking in a nutshell: track fitting

- **Perfect measurement – ideal**



- **Imperfect measurement – reality**

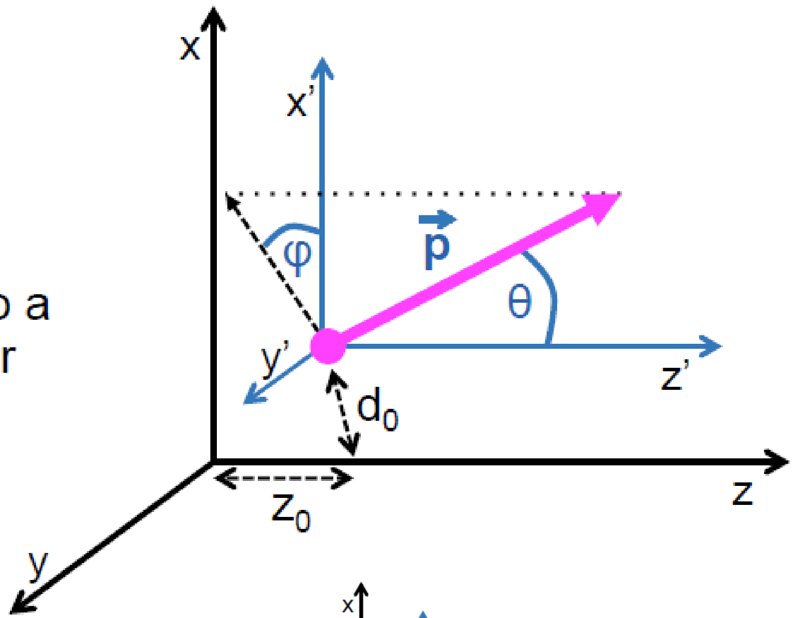
- **Small errors and more points help to constrain the possibilities**

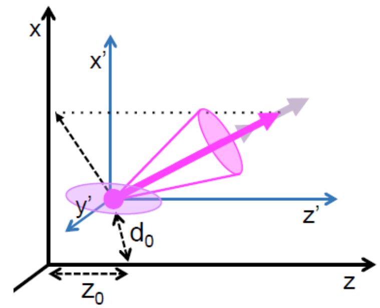


- **Quantitatively:**
  - Parameterize the track;
  - Find parameters by Least-Squares-Minimization;
  - Obtain also uncertainties on the track parameters.

Tracking in a nutshell: track fitting

- **For a track we measure:**
  - Its momentum;
  - It's direction;
  - Its charge;
  - Its "perigee": the closest point to a reference line, transverse (d_0) or longitudinal (z_0).



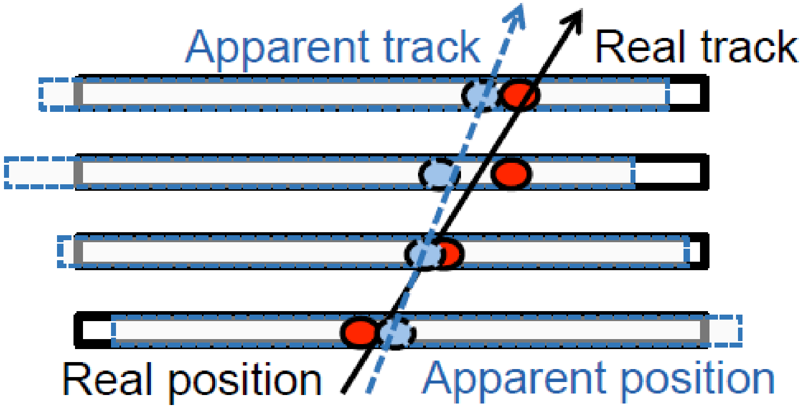


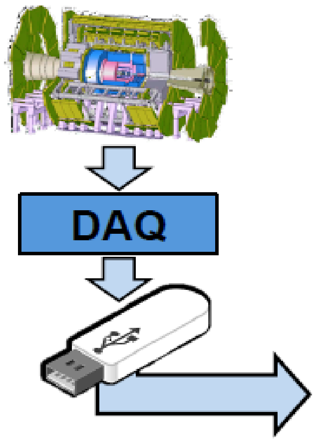
- **Small uncertainties are required.**
  - δd0 is O(10μm) and δθ O(0.1mrad).
  - Allows separation of tracks that come from different particle decays (which can be separated at the order of mm).

Tracking in a nutshell: the uncertainties

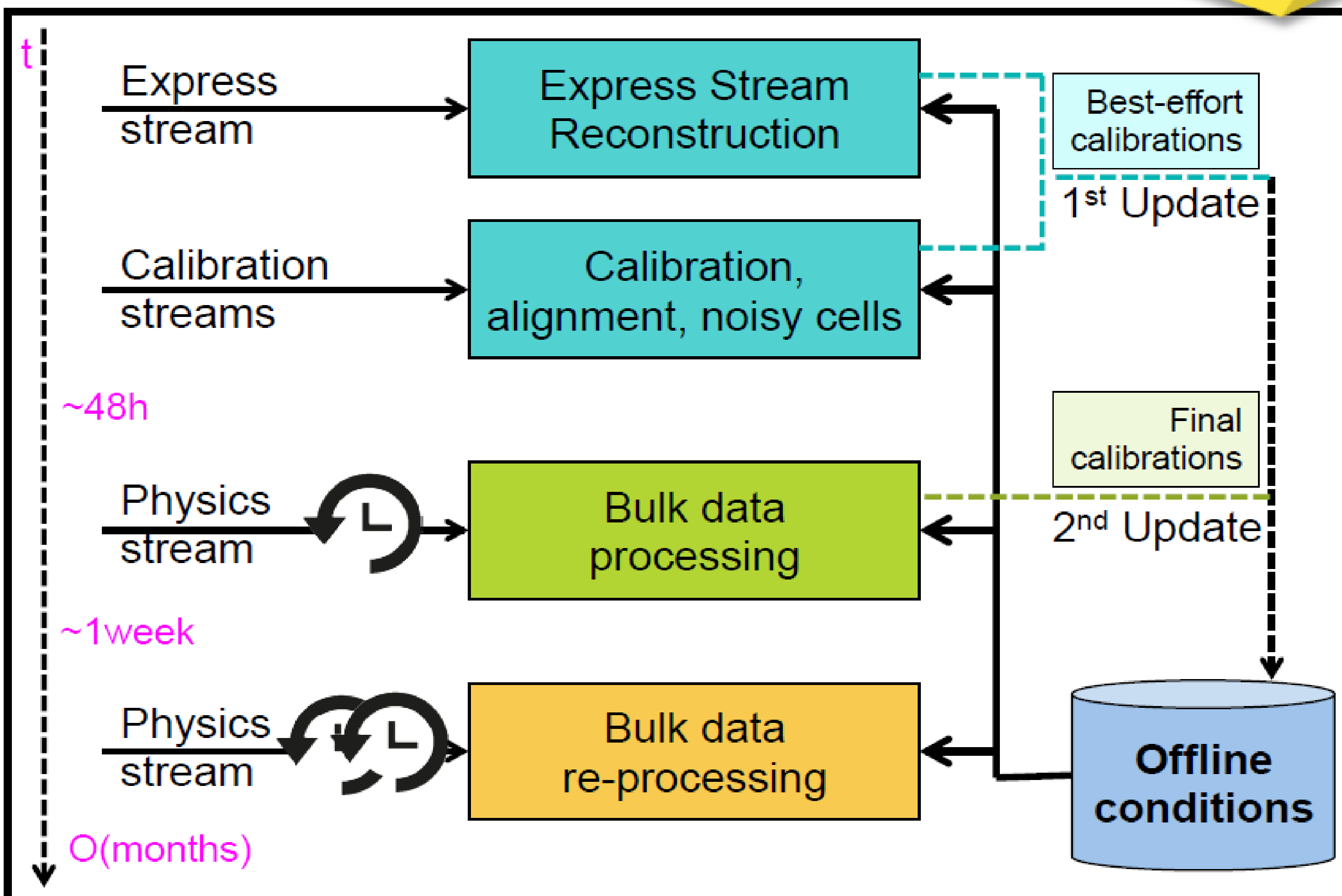
- **Presence of Material**
  - Coulomb scattering off the core of atoms
  - Energy loss due to ionization
  - Bremsstrahlung
  - Hadronic interaction

- **Misalignment**
  - Detector elements not positions in space with perfect accuracy.
  - Alignment corrections derived from data and applied in track reconstruction.

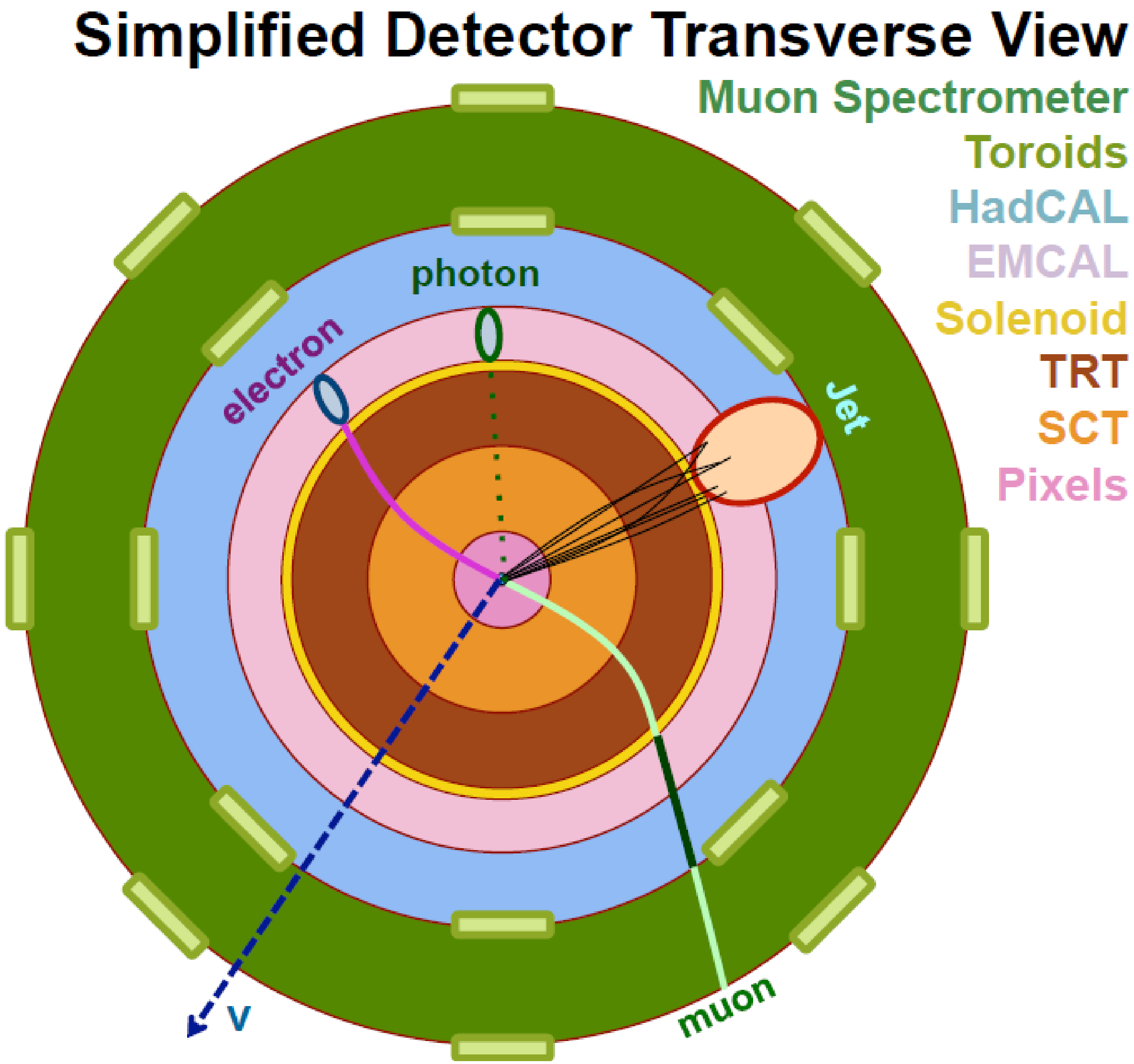




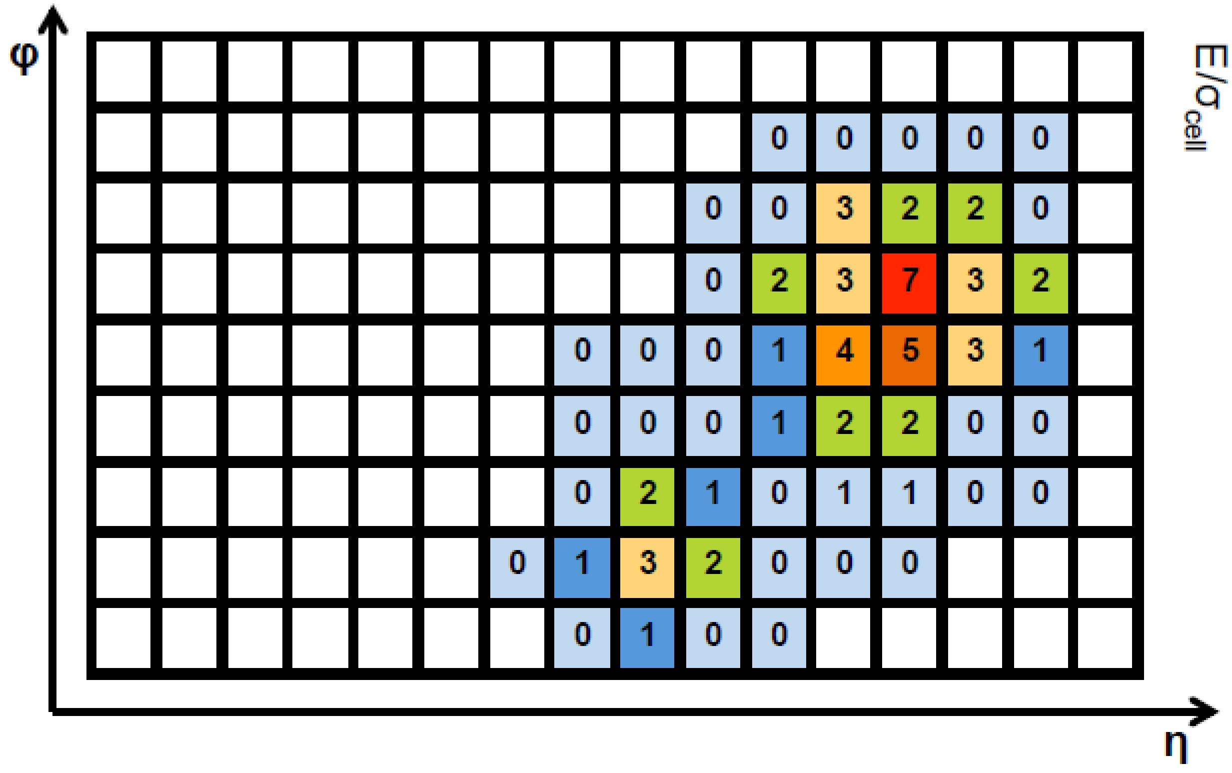
THE EVENT AT TIER0



What do we reconstruct?



A calorimeter view



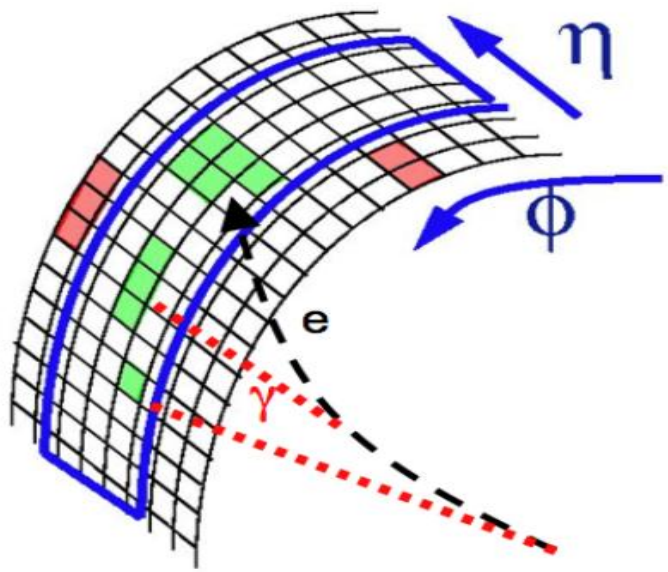
Clustering in a nutshell

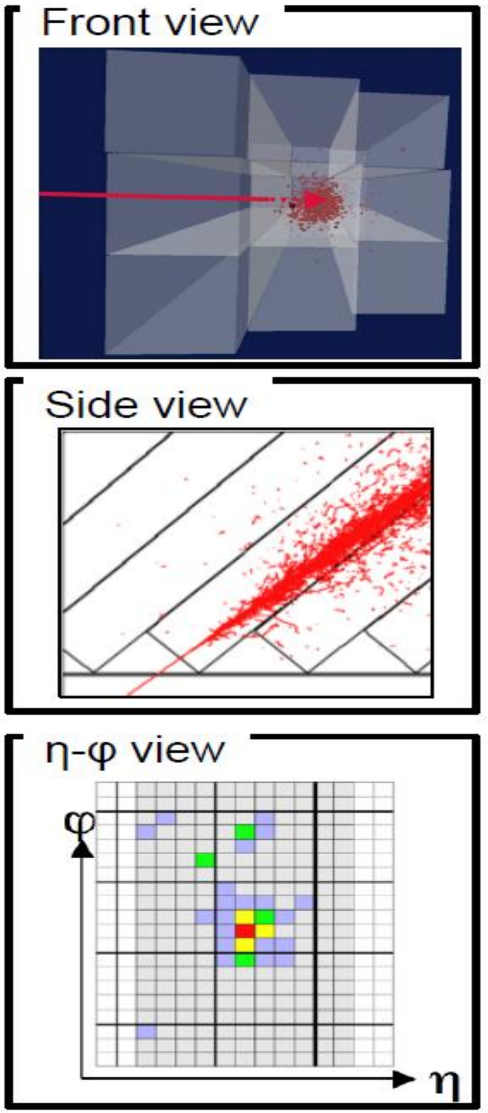
- **Reconstruct energy deposited in the calorimeter by charged or neutral particles; electrons, photons and jets.**
- **For a cluster we measure:**
  - The energy;
  - The position of the deposit;
  - The direction of the incident particles;
- **Calorimeters are segmented in cells.**
  - Typically a shower created by a particle interacting with the matter extends over several cells.
- **Various clustering algorithms, e.g.:**
  - **Sliding window**. Sum cells within a fixed-size rectangular window.
  - **Topo-clustering**. Start with a seed cell and iteratively add to the cluster the neighbor of a cell already in the cluster.

Cluster finding – an example

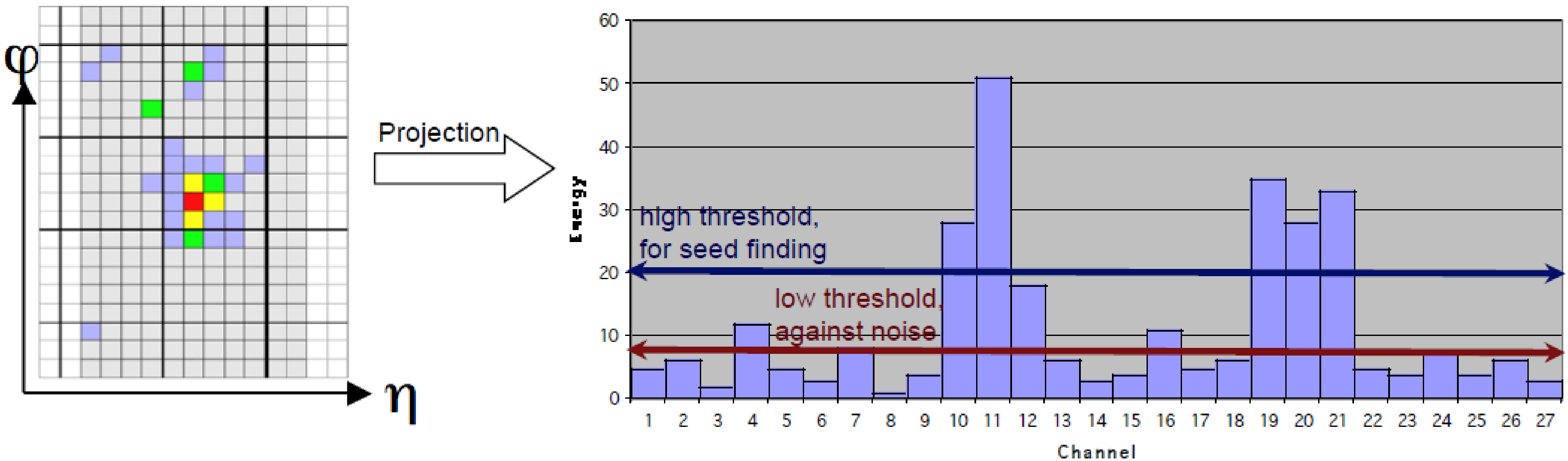
CMS crystal calorimeter – ECAL clusters

◎ electron energy in central crystal ~80%, in 5x5 matrix around it ~96%.





Cluster finding – an example



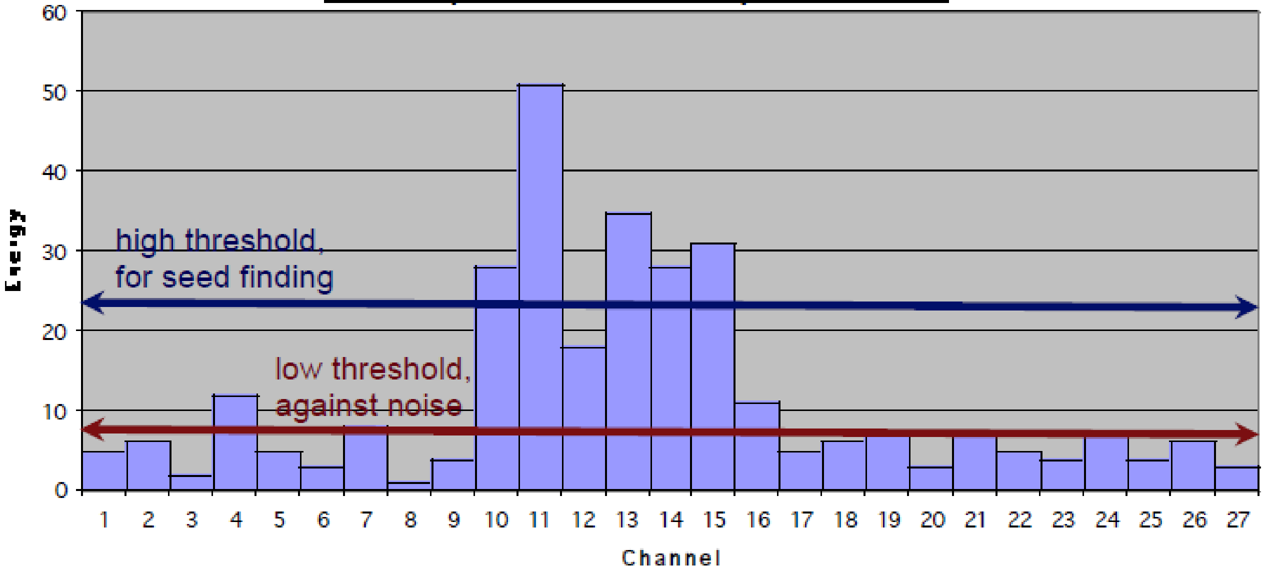
Simple example of an algorithm

- Scan for seed crystals = local energy maximum above a defined seed threshold
- Starting from the seed position, adjacent crystals are examined, scanning first in φ and then in η
- Along each scan line, crystals are added to the cluster if
  1. The crystal's energy is above the noise level (lower threshold)
  2. The crystal has not been assigned to another cluster already

Cluster finding – difficulties

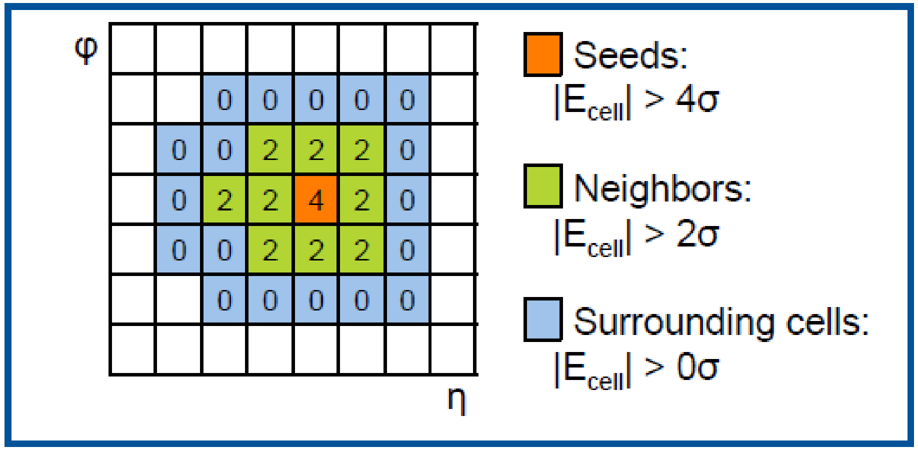
- **Careful tuning of thresholds needed.**
  - needs usually learning phase;
  - adapt to noise conditions;
  - too low : pick up too much unwanted energy;
  - too high : loose too much of “real” energy. Corrections/Calibrations will be larger.

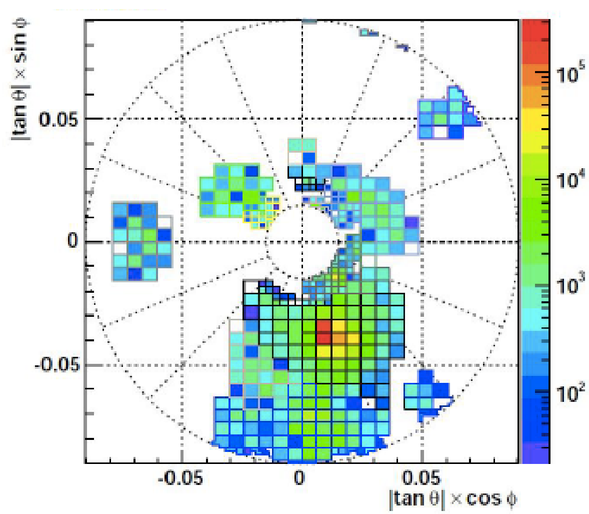
example : one lump or two?



Cluster finding – topological clustering

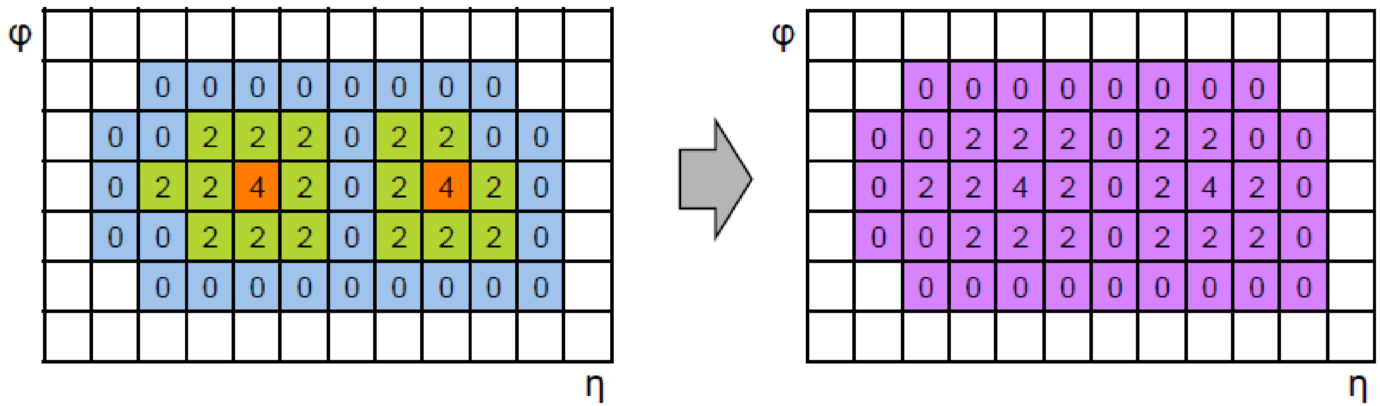
“Topological” clusters, i.e. “blobs” of energy inside the detector.



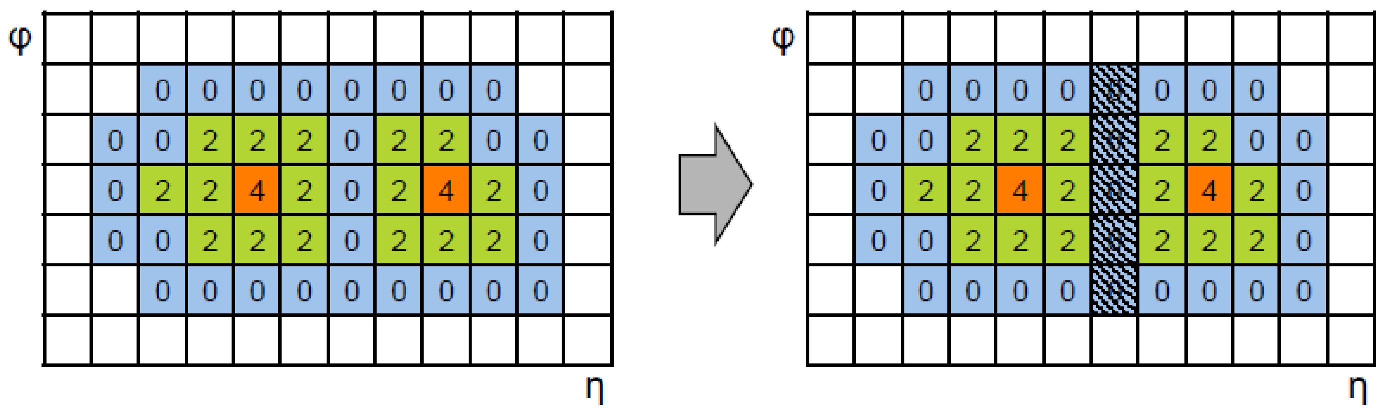


Cluster finding – merging and splitting

- If clusters have common neighboring cells, they are merged according to the basic algorithm.



- Clusters are split if more than one local maxima.



For common cells, a weight is applied to share them (shaded cells).

Cluster calibration

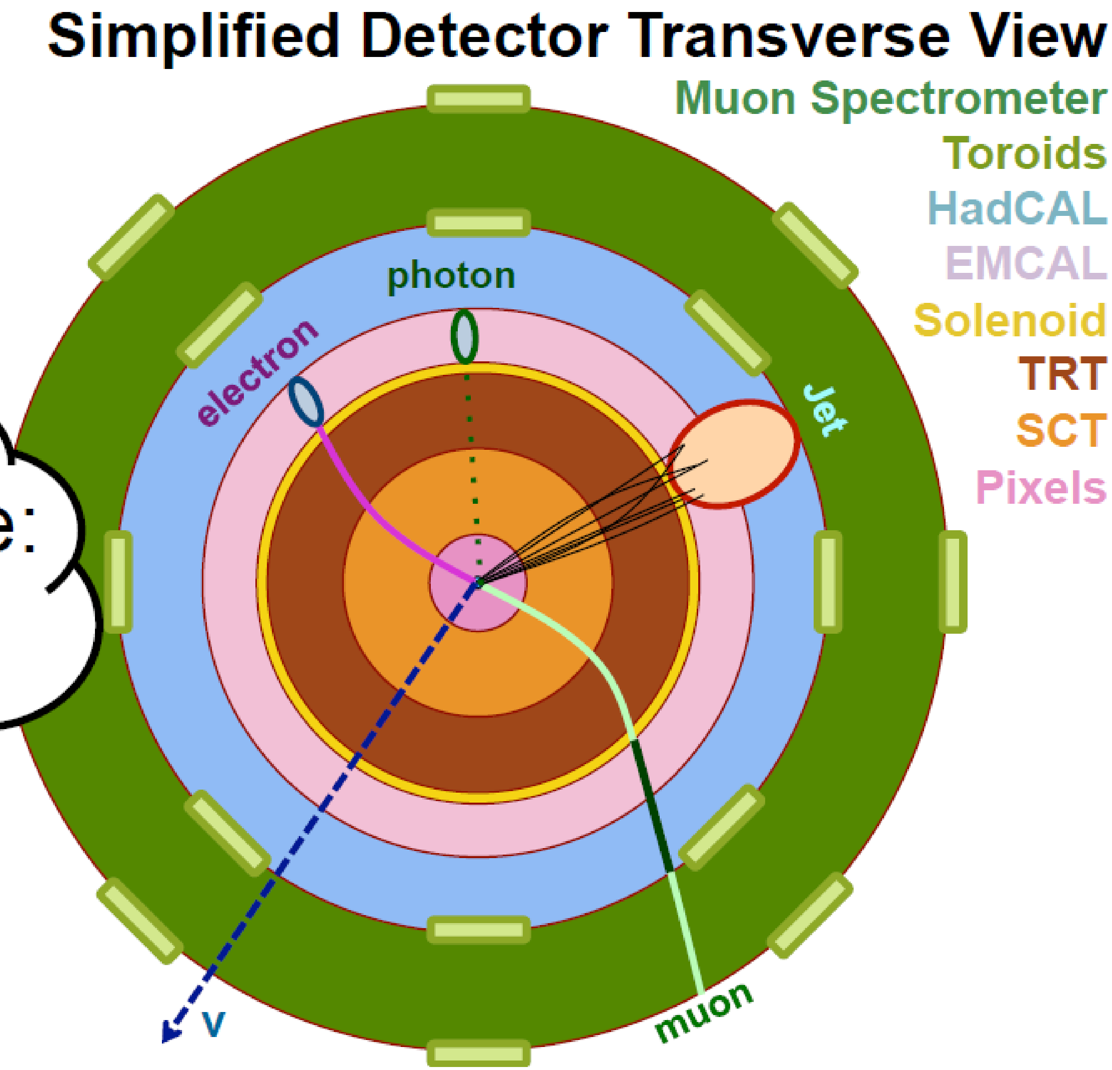
Possible energy measurements:

- **Non-calibrated clusters: sum energy using baseline cell-level detector calibration.**
  - That's NOT the true energy of the particle that originated the cluster.
- **Local calibration: apply weights to correct for:**
  - the different **calorimeter response** on an EM (e.g. π^0) or a hadronic (e.g. $\pi^\pm$) deposition.
  - the low energetic deposits, lost in the tails of the shower (**"out-of-cluster"** corrections, derived from simulation).
  - the presence of **dead material**, i.e. material without a read-out device, where energy is lost.
- **Corrections are complex functions of the energy and the position of the cluster and other parameters defining the cluster shapes.**

What do we reconstruct?

Tracks and Clusters

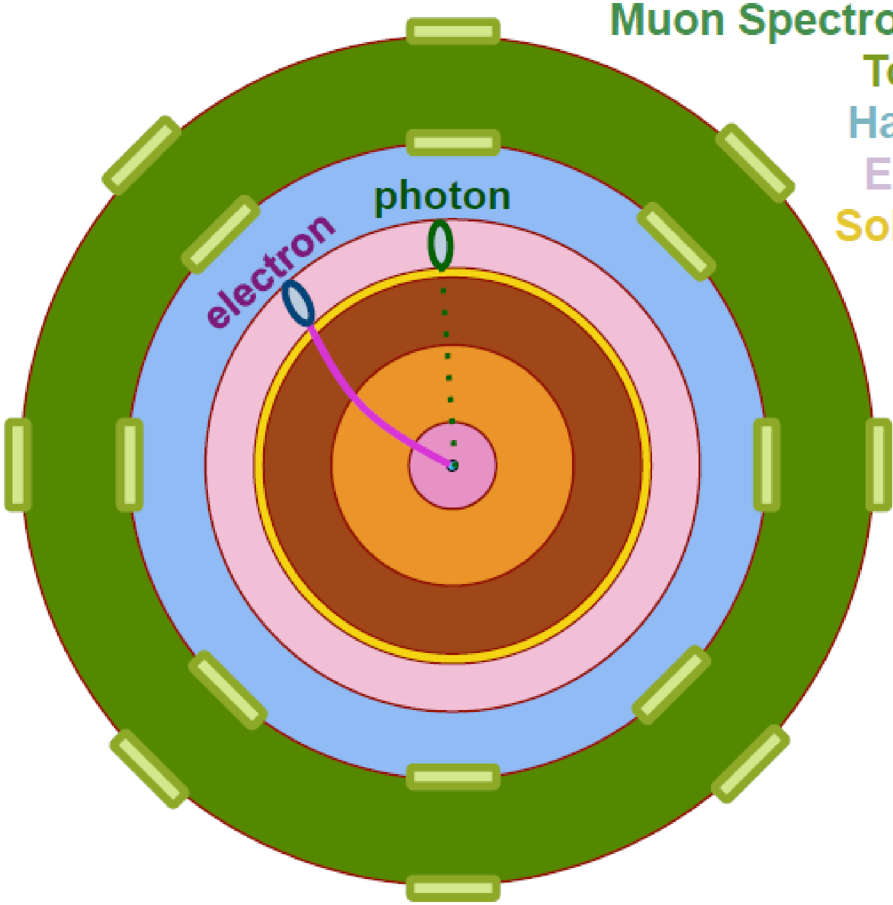
Combining those:
"**objects**"
("particles")



Electrons and photons

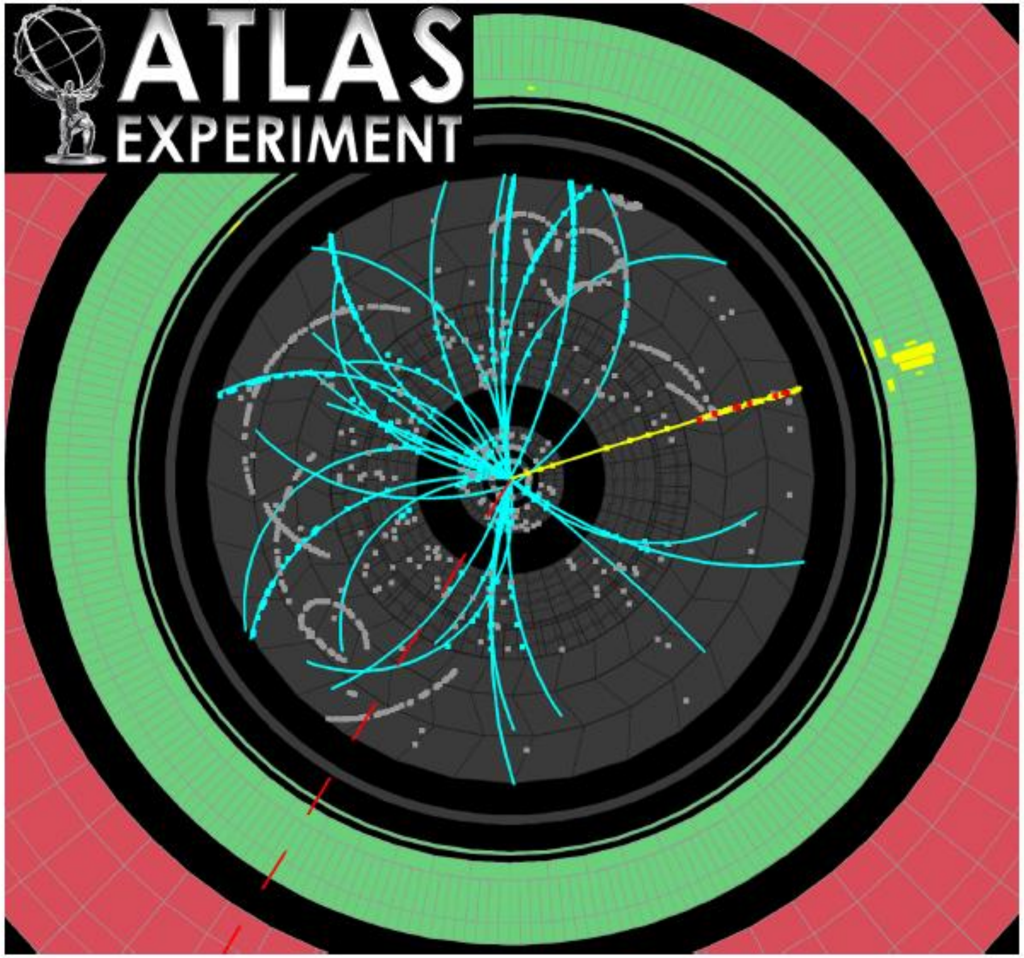
- **Final Electron momentum measurement can come from tracking or calorimeter information (or a combination of both).**
  - Often have a final calibration to give the best electron energy.
- **Often want "isolated electrons".**
  - Require little calorimeter energy or tracks in the region around the electron.

Simplified Detector Transverse View



Electrons and photons

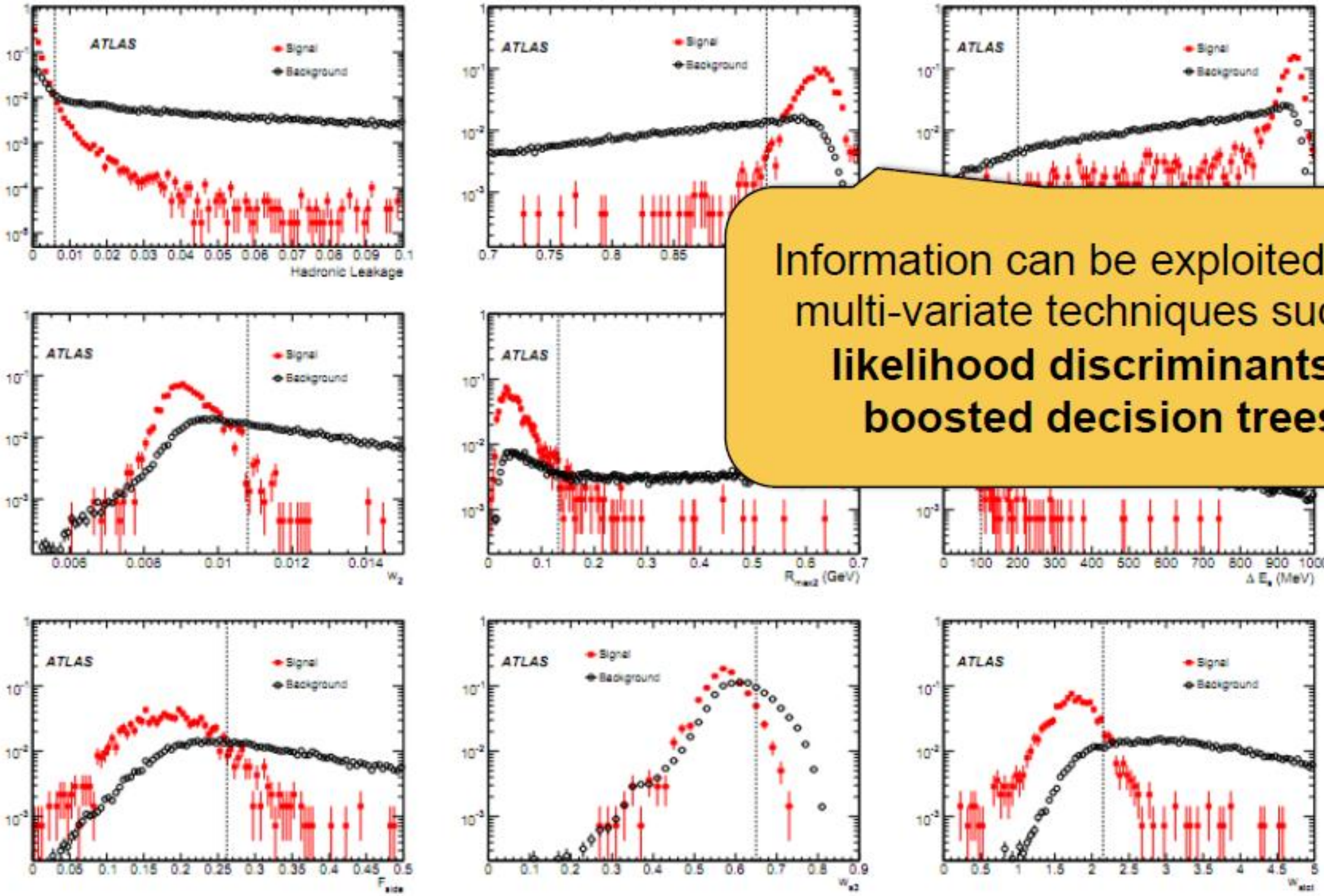
- **Final Electron momentum measurement can come from tracking or calorimeter information (or a combination of both).**
  - Often have a final calibration to give the best electron energy.
- **Often want "isolated electrons".**
  - Require little calorimeter energy or tracks in the region around the electron.



Electrons and photons (backgrounds)

- Hadronic jets leave energy in the calorimeter which can fake electrons or photons.
- Usually a Jet produces energy in the hadronic calorimeter as well as the electromagnetic calorimeter.
- Usually the calorimeter cluster is much wider for jets than for electrons/photons.
- So it should be “easy” to separate electrons from jets.
- However have many thousands more jets than electrons, so need the rate of jets faking an electron to be very small $\sim 10^{-4}$.
- Need complex identification algorithms to give the rejection whilst keeping a high efficiency.

Electrons and photons (backgrounds)

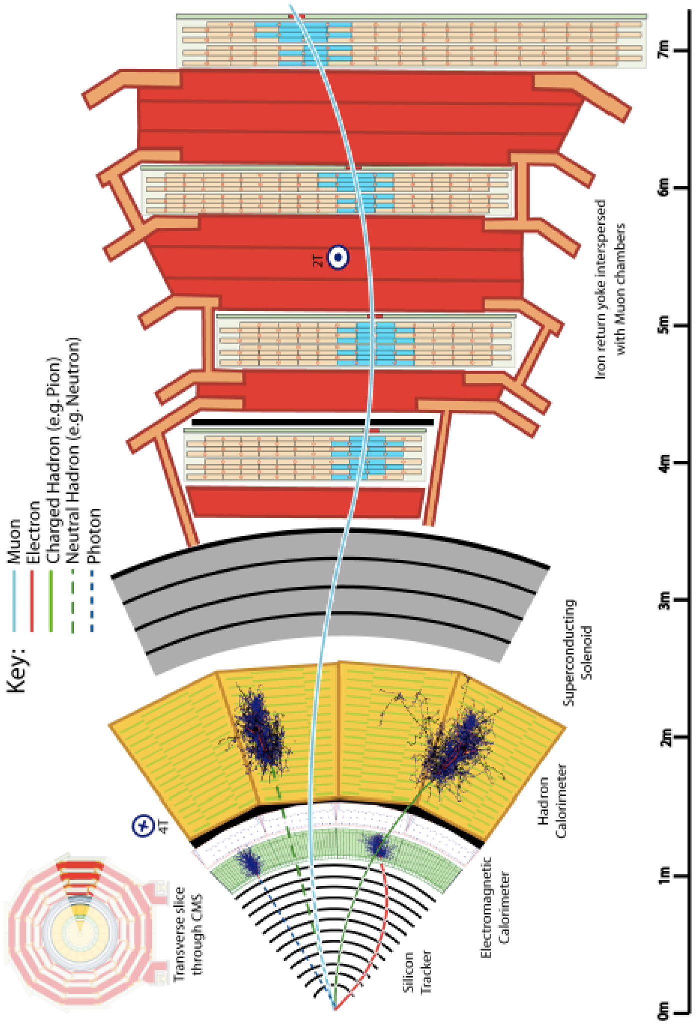


Information can be exploited using multi-variate techniques such as **likelihood discriminants** or **boosted decision trees**.

Example of different calorimeter shower shape variables used to distinguish electron showers from jets in ATLAS

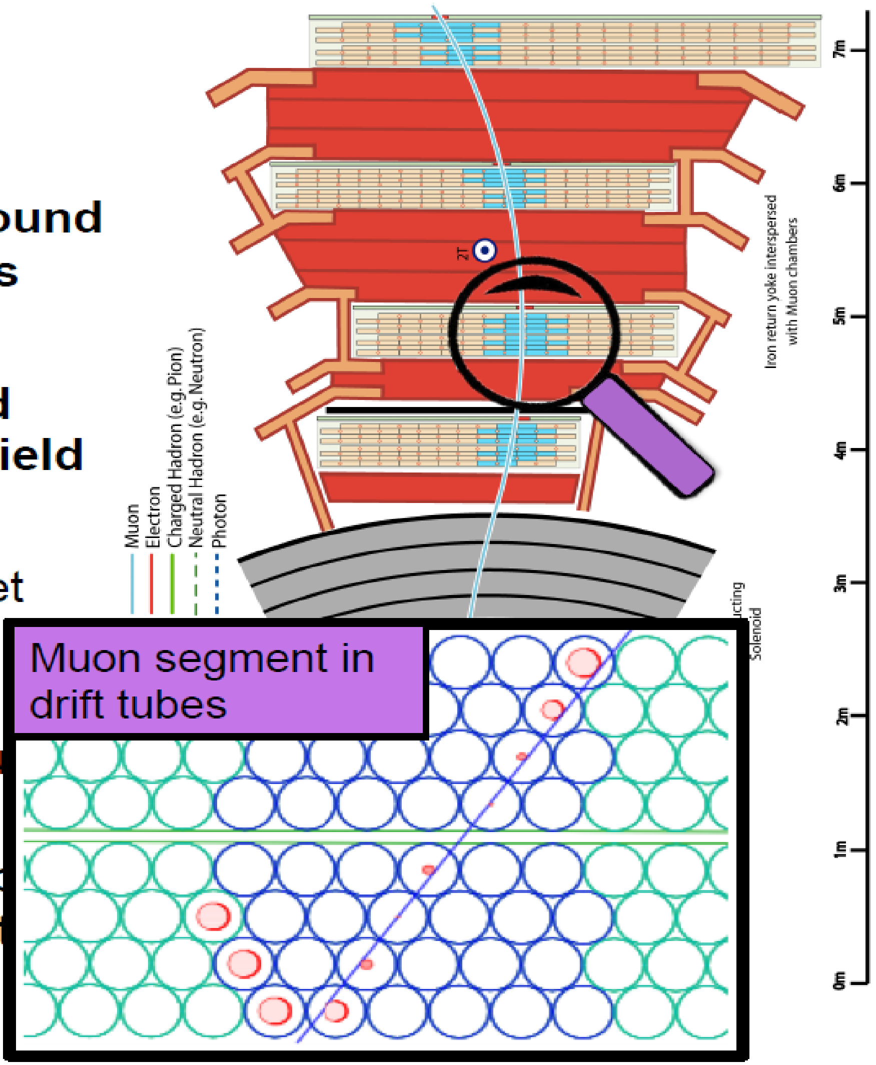
Muons

- **Combine the muon segments found in the muon detector with tracks from the tracking detector**
- **Momentum of muon determined from bending due to magnetic field in tracker and in muon system**
  - Combine measurements to get best resolution
  - Need an accurate map of the magnetic field in the reconstruction software
  - Alignment of the muon detectors also very important to get best momentum resolution



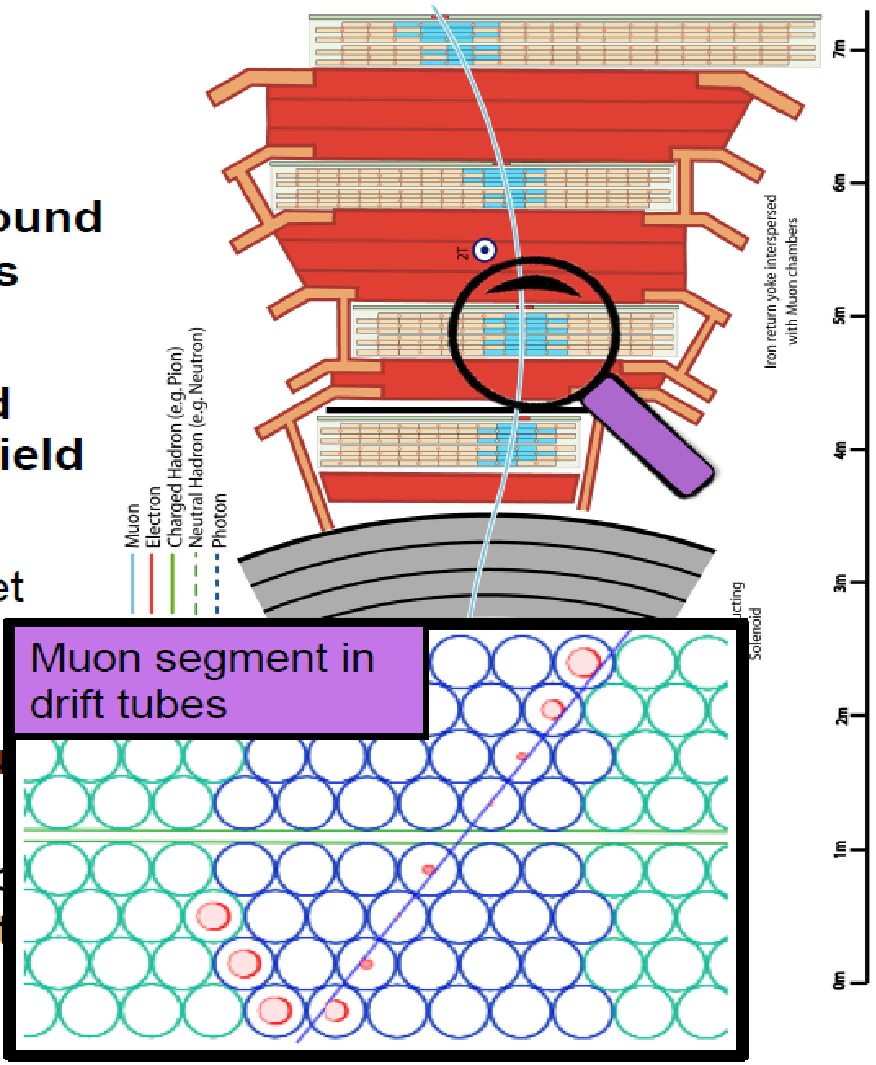
Muons

- **Combine the muon segments found in the muon detector with tracks from the tracking detector**
- **Momentum of muon determined from bending due to magnetic field in tracker and in muon system**
  - Combine measurements to get best resolution
  - Need an accurate map of the magnetic field in the reconstru software
  - Alignment of the muon detecto also very important to get best momentum resolution

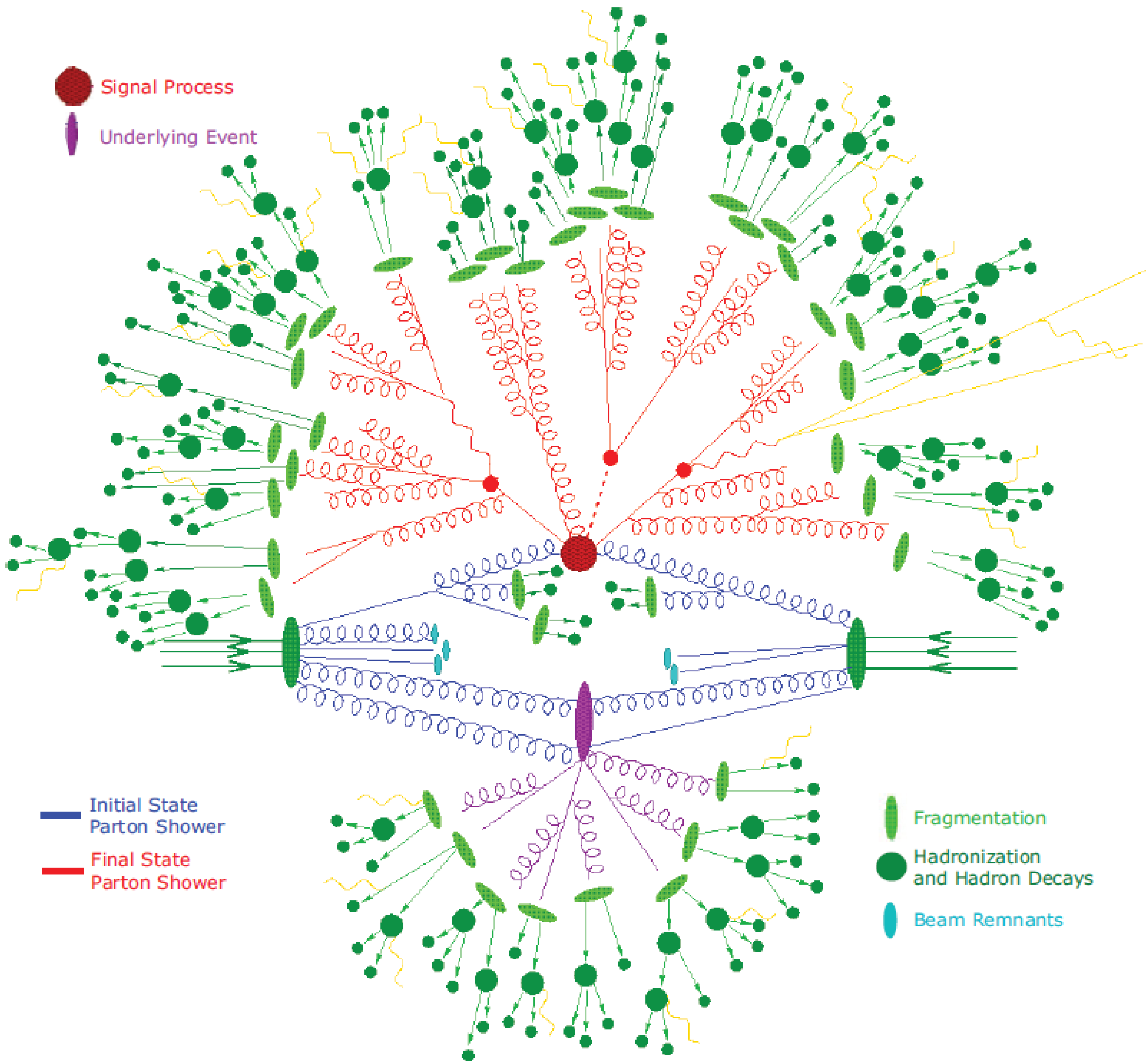


Muons

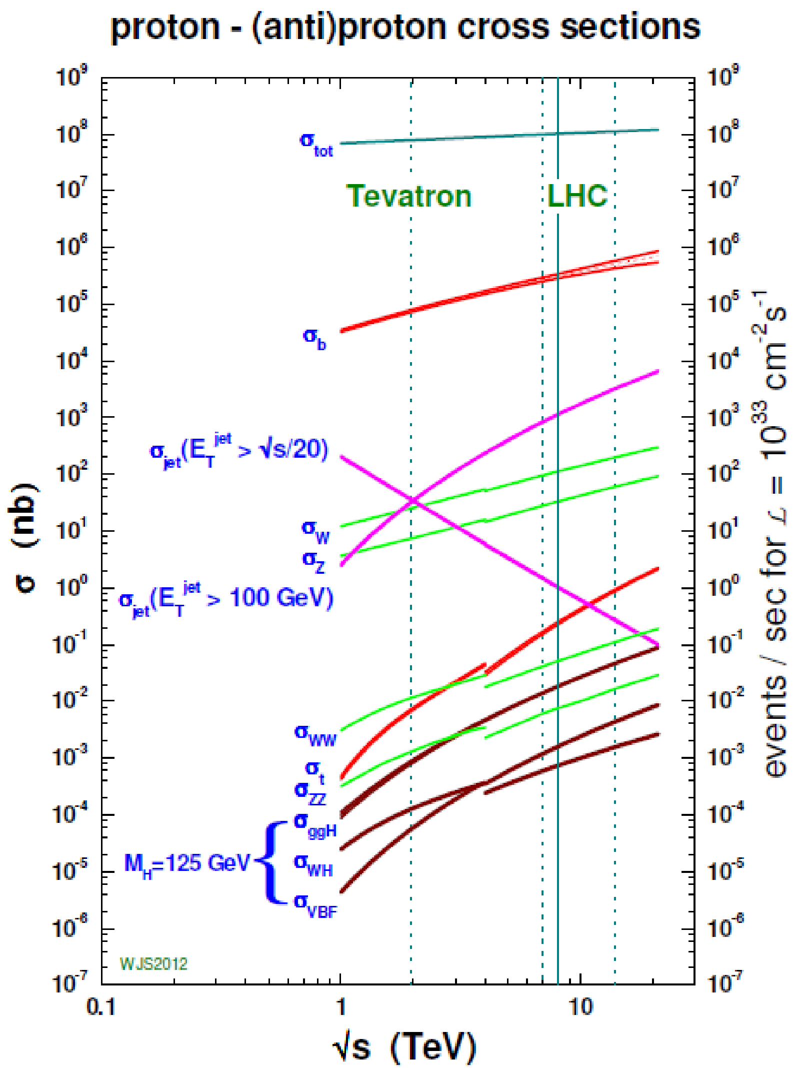
- **Combine the muon segments found in the muon detector with tracks from the tracking detector**
- **Momentum of muon determined from bending due to magnetic field in tracker and in muon system**
  - Combine measurements to get best resolution
  - Need an accurate map of the magnetic field in the reconstru
    software
  - Alignment of the muon detecto
    also very important to get best
    momentum resolution



Jets


Signal Process
Underlying Event
Initial State Parton Shower
Final State Parton Shower
Fragmentation
Hadronization and Hadron Decays
Beam Remnants

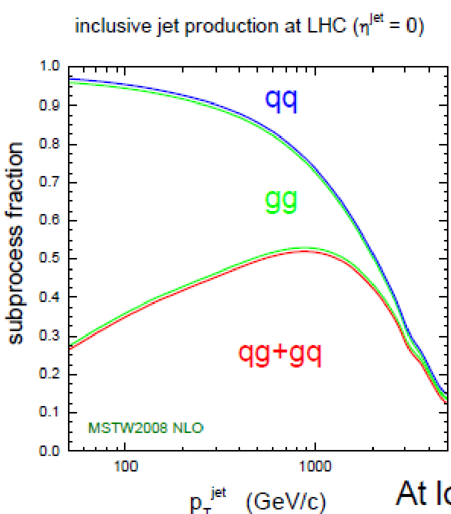
Standard Model processes



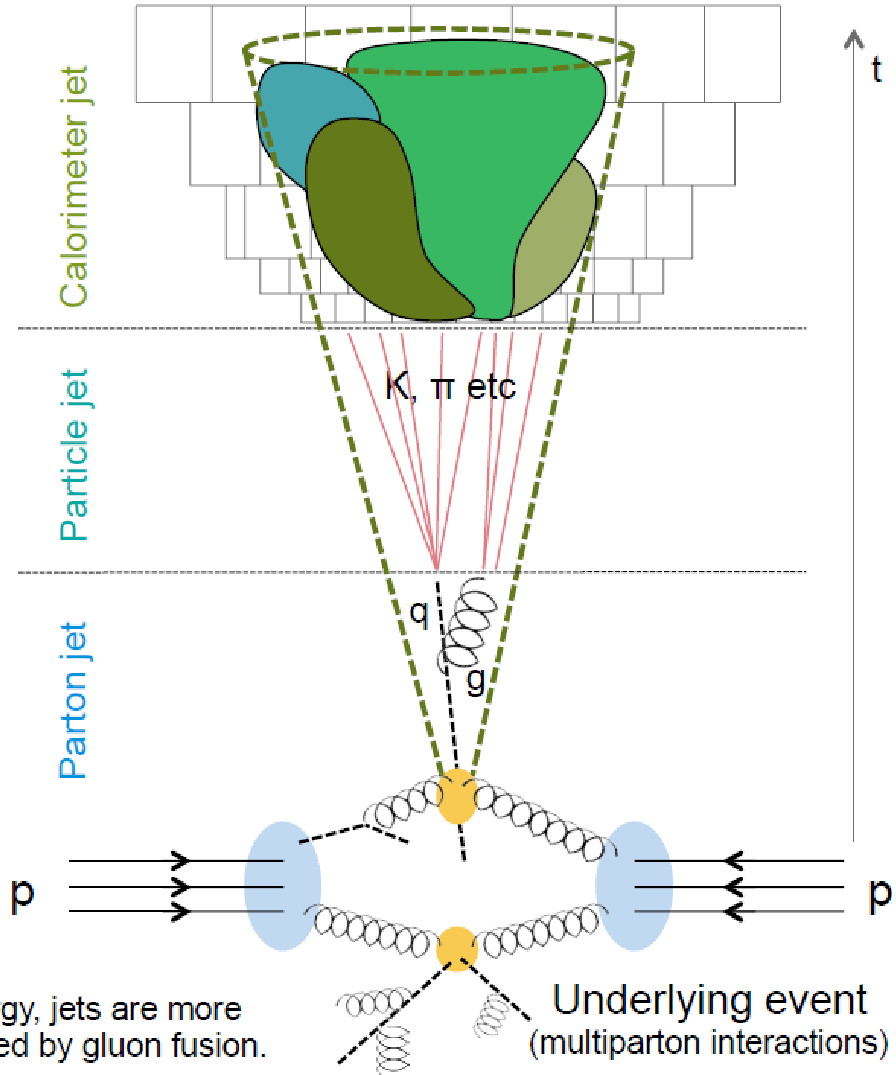
Jets are produced:

- by fragmentation of gluons and (light) quarks in QCD scattering.
- by decays of heavy Standard Model particles, e.g. W & Z.
- in association with particle production in Vector Boson Fusion, e.g. Higgs.
- in decays of beyond the Standard Model particles, e.g. in SUSY.

Jets

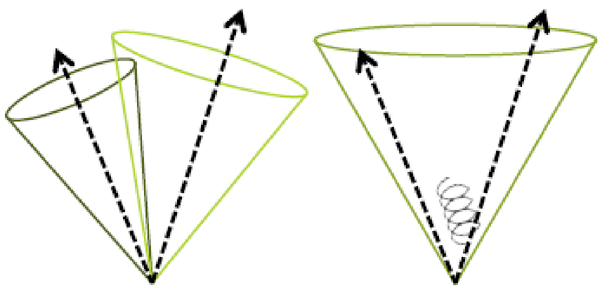


At low energy, jets are more likely produced by gluon fusion.

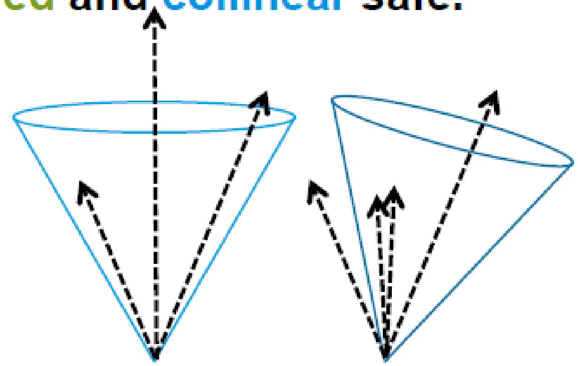


Jet algorithms

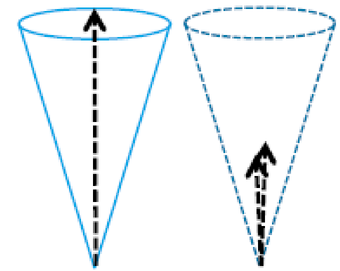
<u>Theoretical requirements</u>: infrared and collinear safe.



Soft gluon radiation should not merge jets



Final jet should not depend on the ordering of the seeds…



…and on signal split in two possibly below threshold

<u>Experimental requirements</u>: detector technology & environment independent, easily implementable.

Insignificant effects of detector
- Noise
- Dead material
- Cracks

Stability with
- Luminosity
- Pile-up
- Physics process

Fully specified
Fast

<u>Jet algorithm commonly used at the LHC</u>: 'anti-k_t'. A 'recursive recombination' algorithm. Starts from (topo-)clusters. Hard stuff clusters with nearest neighbor. Various cone sizes (standard R=0.4/0.5, "fat" R=1.0).

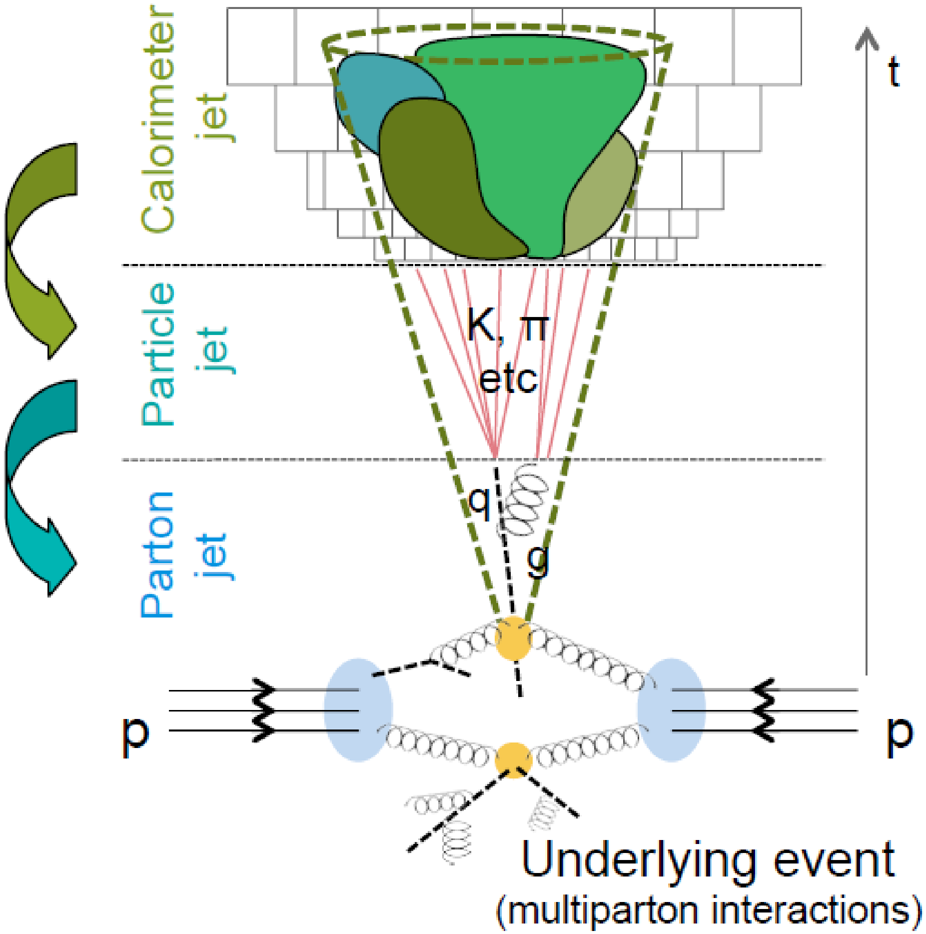
Jet calibration

Correct the energy and position measurement and the resolution.

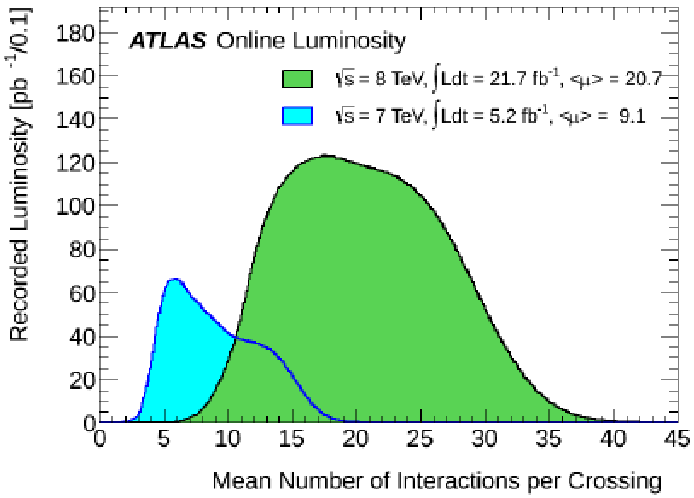
Account for:

<u>Instrumental effects</u>
Detector inefficiencies
'Pile-up'
Electronic noise
Clustering, noise suppression
Dead material losses
Detector response
Algorithm efficiency

<u>Physics effects</u>
Algorithm efficiency
'Pile-up'
'Underlying event'



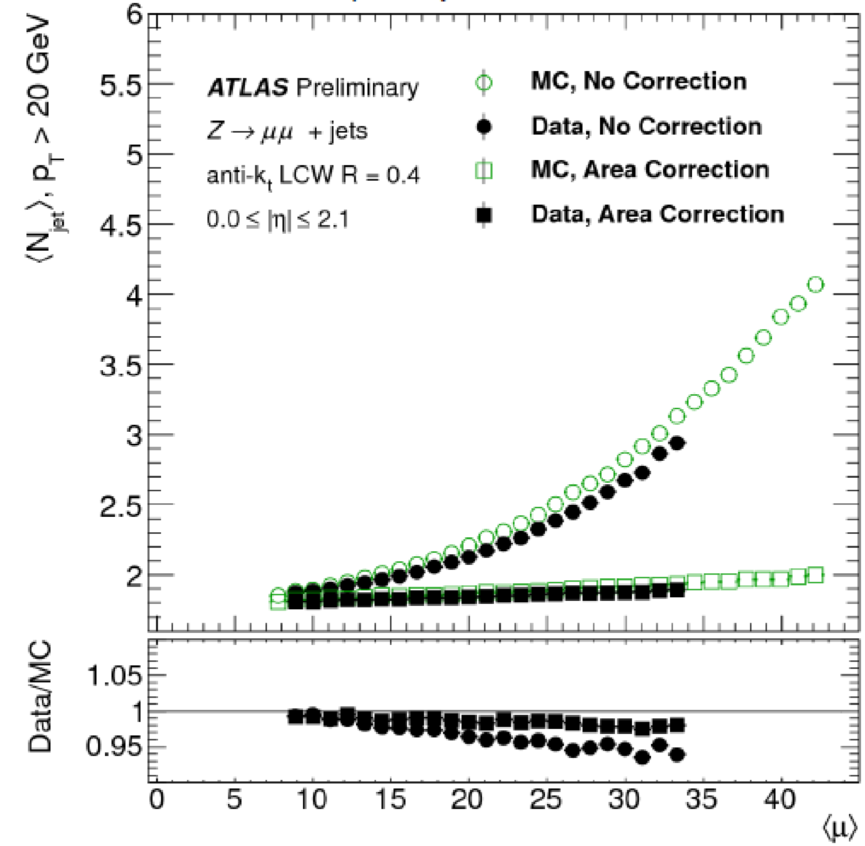
Jets & pile-up



Multiple interactions from pile-up

'Jet-areas' corrections

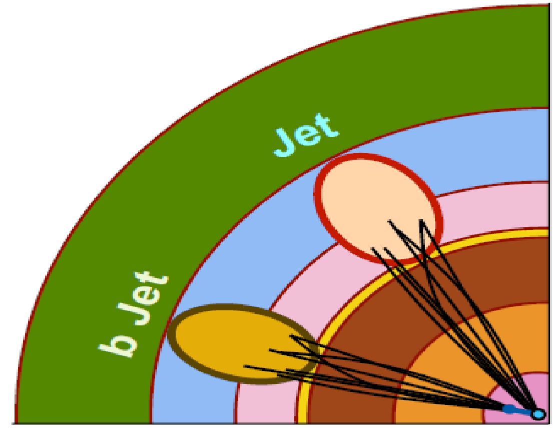
Inspired by arXiv:0707.1378

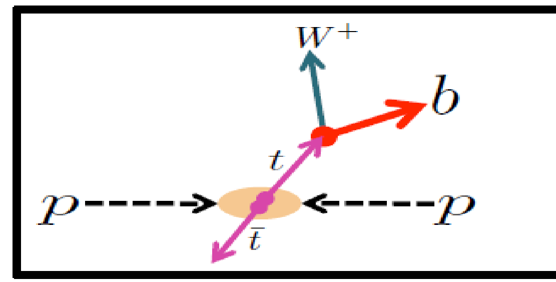


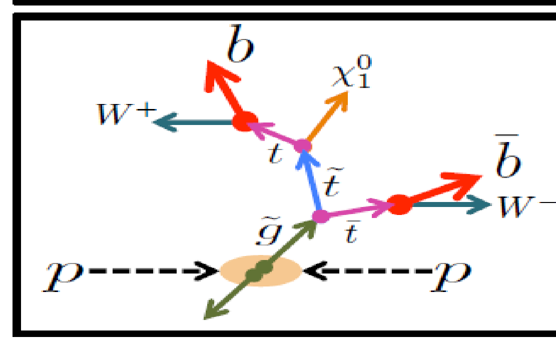
b-jets

- b-quarks have a lifetime of ~ 10^{-12} s.
- They travel a small distance (fraction of mm) before decaying.
- A "**displaced vertex**" creates a distinct jet, so b-jets can be tagged (**b-tagged**).
- b-tagging uses sophisticated algorithms, mostly **multi-variate**.

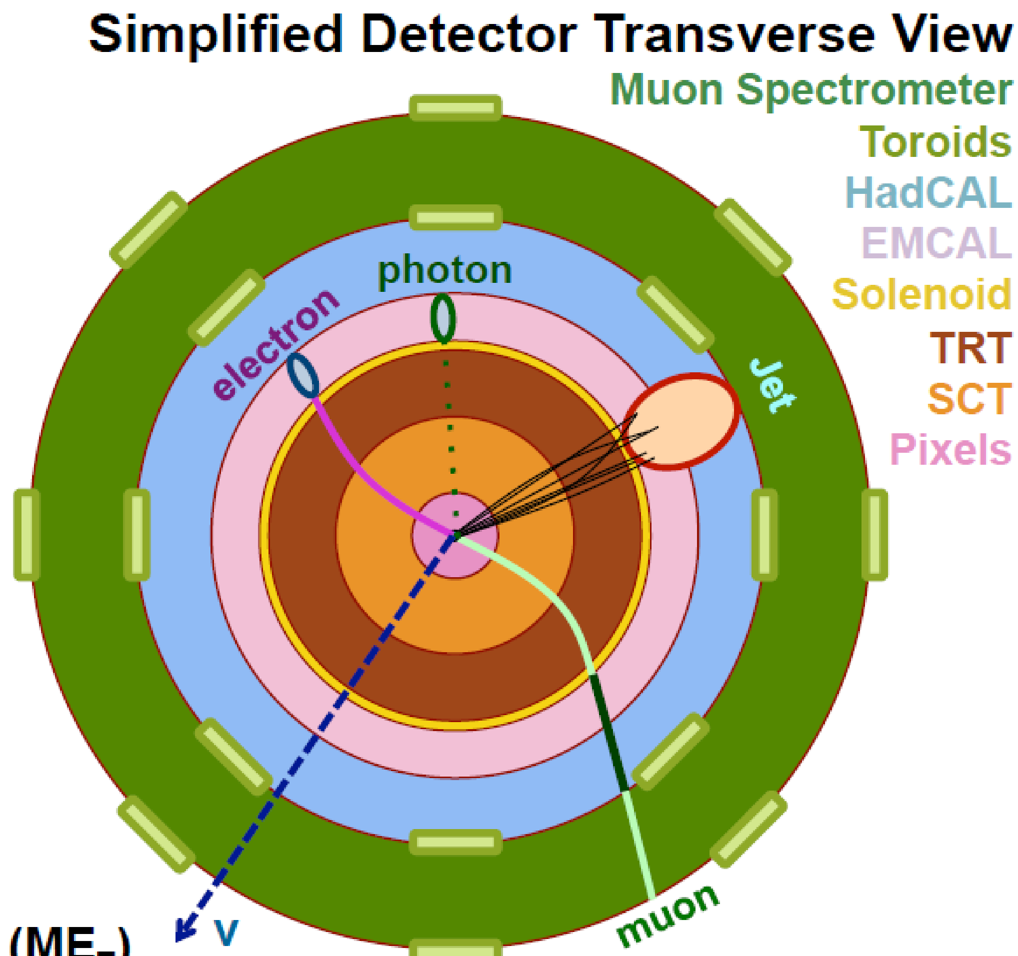
- b-jets create distinct final states, important for both **Standard Model measurements** and **searches for New Physics**.







Missing transverse momentum



In the transverse plane:

$$\sum \vec{p}_T = 0$$

Missing Transverse Momentum (ME_T)

Missing transverse momentum

Impossible to measure particles that don't interact in the detector.

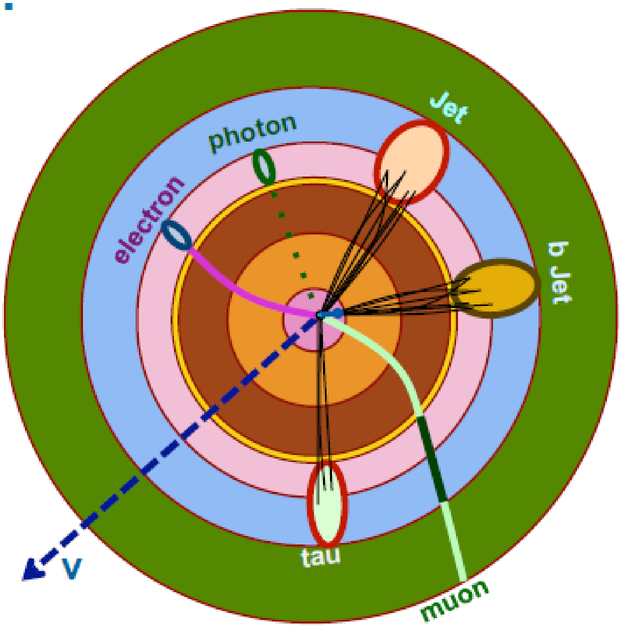
➲ Instead, measure everything else & require momentum conservation in the transverse plane.

◎ Sensitive to pile-up and detector problems.

Only as good as its inputs.

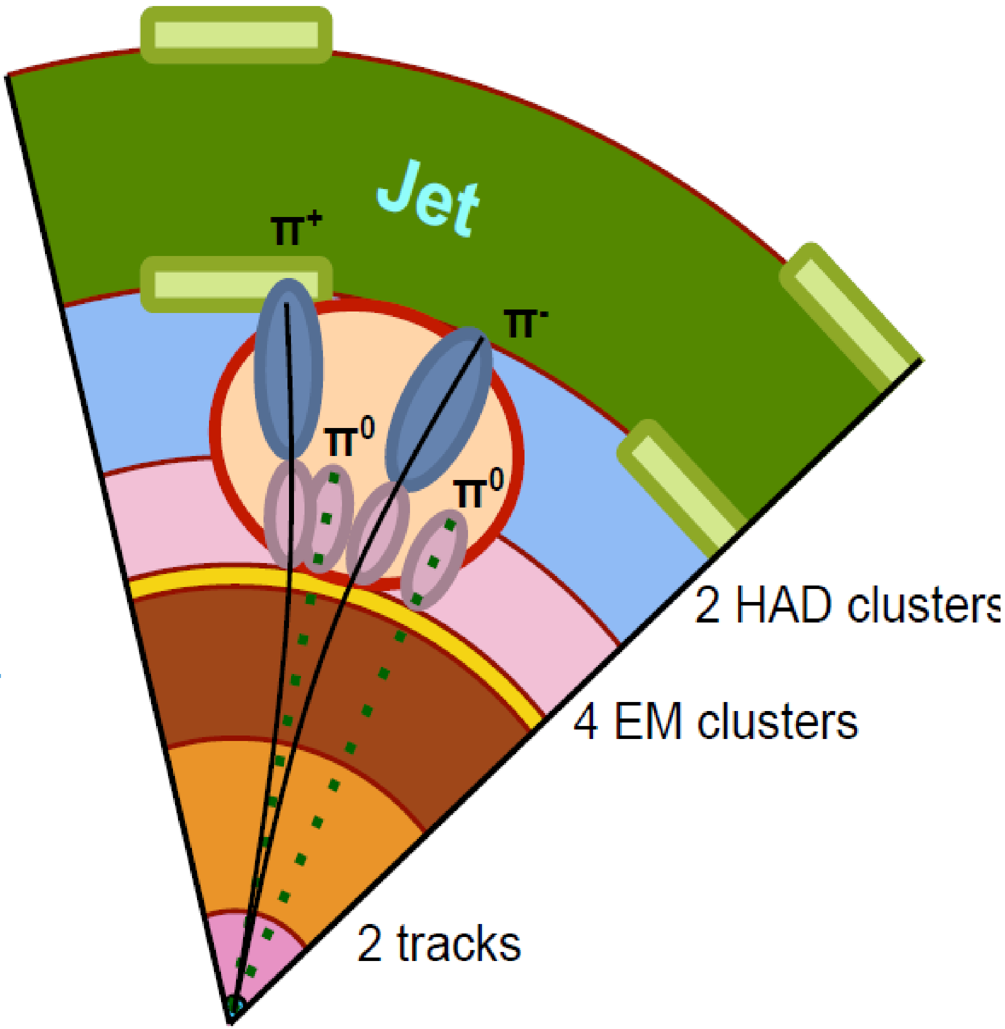
◎ Use calibrated physics objects: electrons, photons, muons, taus, jets.

◎ Add remaining soft energy.

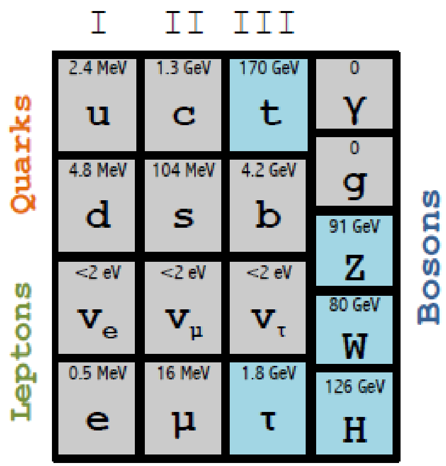


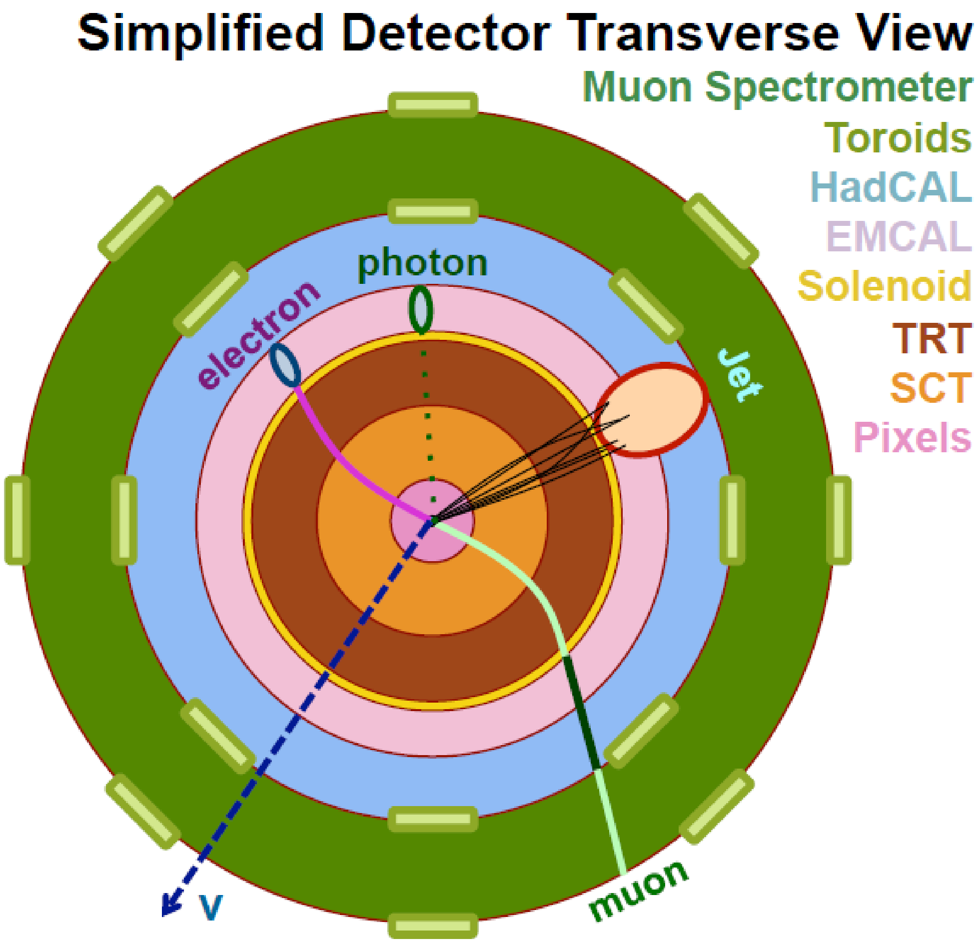
Particle flow

- "Flow of particles" through the detector.
- Reconstruct and identify all particles, photons, electrons, pions, …
- Use best combination of all sub-detectors for measuring the properties of the particles.
- First used at LEP (ALEPH) and then at the LHC (CMS).



Reconstructing particles



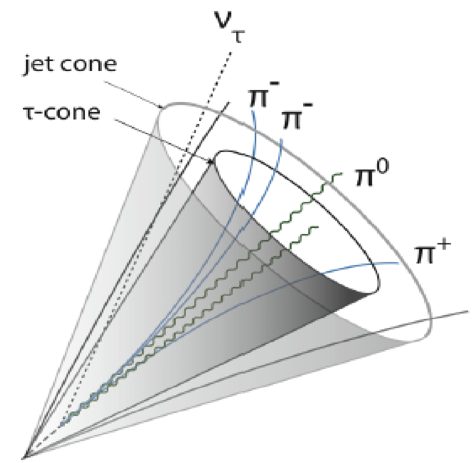


Reconstructing particles

Tau Decay Mode			B.R.
Leptonic		$\tau^{\pm} \rightarrow e^{\pm} + \nu + \nu$	17.8%
		$\tau^{\pm} \rightarrow \mu^{\pm} + \nu + \nu$	17.4%
Hadronic	1-prong	$\tau^{\pm} \rightarrow \pi^{\pm} + \nu$	11%
		$\tau^{\pm} \rightarrow \pi^{\pm} + \nu + n\pi^{0}$	35%
	3-prong	$\tau^{\pm} \rightarrow 3\pi^{\pm} + \nu$	9%
		$\tau^{\pm} \rightarrow 3\pi^{\pm} + \nu + n\pi^{0}$	5%
Other			~5%

- Hadronic tau reconstruction extremely challenging.
- Using **multi-variate** techniques based on track multiplicity and shower shapes.

A tau jet (signal)…



…vs. a QCD jet (background)