EVENT RECONSTRUCTION AND DATA ANALYSIS

LECTURE 5

CONTENT

1. Short history of discoveries
   1. Discovery of radioactivity
   2. Discovery of cosmic rays
   3. Discovery by imaging: cloud chambers
   4. Nuclear emulsion
   5. Spark chamber – discovery of the muon neutrino
   6. Bubble chamber – discovery of neutral currents
   7. Bubble and cloud chambers
   8. Electrical readout
   9. Discovery of the W and Z boson
   10. Large research centers
   11. CMS Detector
   12. ATLAS Detector
2. Event reconstruction
   1. A complex simulated event
   2. CMS overview
   3. Overview of detector signatures
   4. Principles of event reconstruction
   5. Particle flow event reconstruction
   6. Jet clustering
   7. Jet shapes for different algorithms
   8. Information on jet substructure
   9. Identification of b quarks
   10. Removal of charged pile up
   11. * Global pile up mitigaition
3. Statistical and systematical uncertainties
   1. Precision And accuracy
   2. Normal distribution (Gaussian)
   3. The central limit theorem
   4. Poisson distribution
   5. Statistical and systematic uncertainties
   6. (some) Typical Hadron collider systematics
4. Principles of data analysis
   1. Analysis workflow overview
   2. Comparing data with prediction
   3. simulation based estimation
   4. Data driven bkg estimation: ABCD method
   5. Translation factors
   6. Sideband subtraction
   7. Tag and Probe
   8. Blinding
   9. Look elsewhere effect

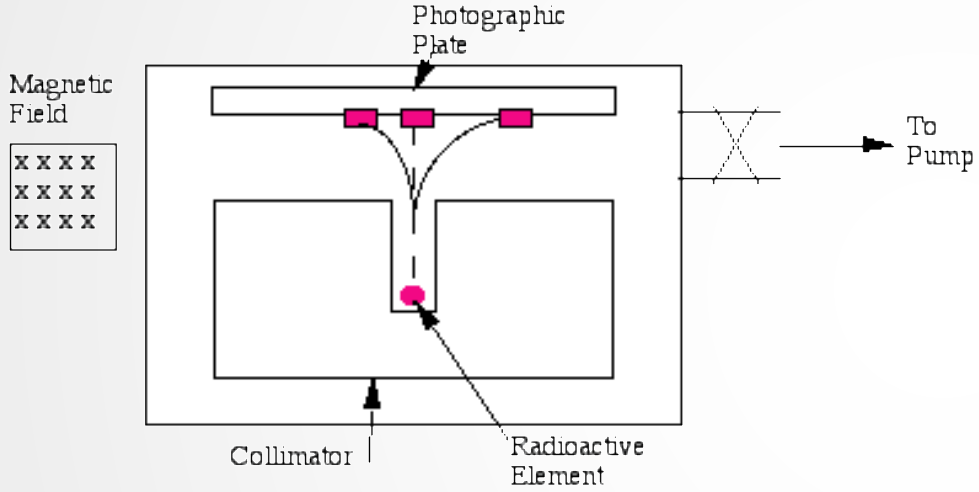
REFERENCES

- Vorlesung M. Krammer "Teilchendetektoren" TU Wien, SS 2015
- Erika Garutti, Lecture notes, "The Physics of Particle Detectors", DESY, SS 2012
- For the statistics part:
  Klaus Reygers, Heidelberg "Statistical Methods in Particle Physics", WS 2017/18

SHORT HISTORY OF DISCOVERIES

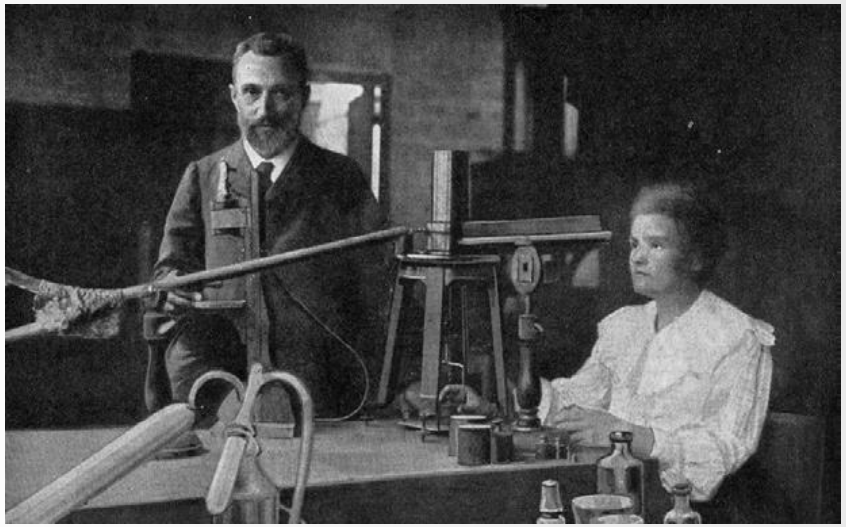
1.1 DISCOVERY OF RADIOACTIVITY

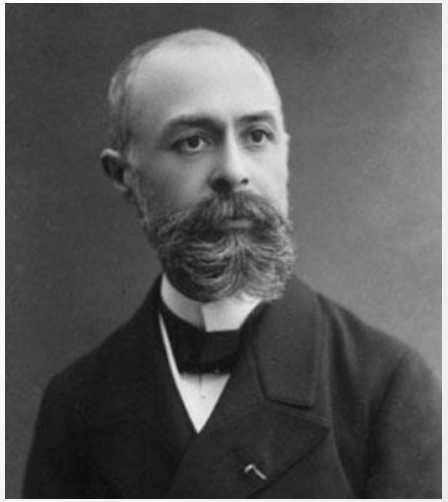
- 1896: Radioactivity ("Uranstrahlen") discovered by Henri Becquerel while studying Röntgen rays



  - He showed that there are several types of radioactivity by deflecting the radiation with a magnetic field

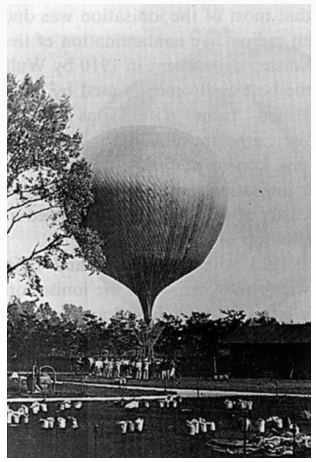
- 1898: Radium discovered by Marie & Pierre Curie
- 1900: Radon (noble gas, decay product of Radium) discovered by Friedrich Ernst Dorn and Rutherford

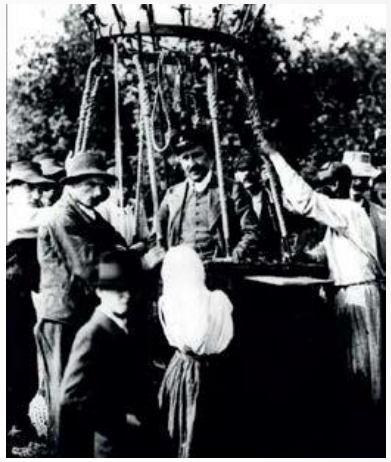


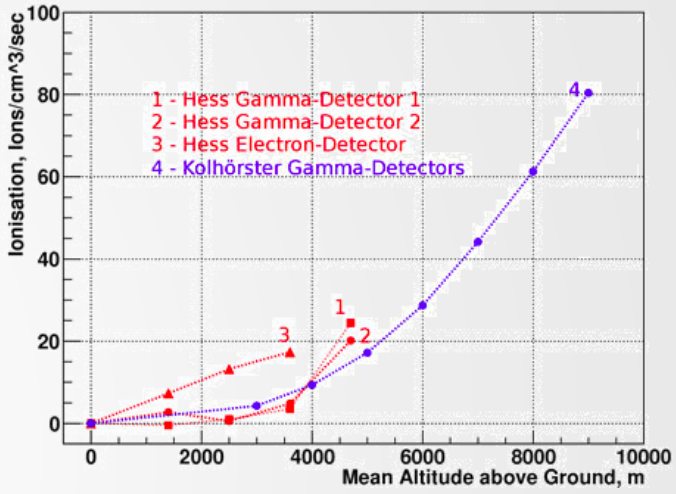


1.2 DISCOVERY OF COSMIC RAYS

- Discovery of cosmic radiation by Victor Hess in 1911/1912
- Seven balloon flights funded by „Imperial Academy of Sciences“
- First six flights around Vienna (starting in Prater) with „Leuchtgas“ (H_2, CH_4, N_2 CO_2 mixture) limited to 1000m altitude
- Last flight 7. August 1912 from Aussig/Elbe to Berlin with pure H_2
  - ⟶ altitude 5000 m
  - ⟶ discovery
    - Professor @ KFU Graz

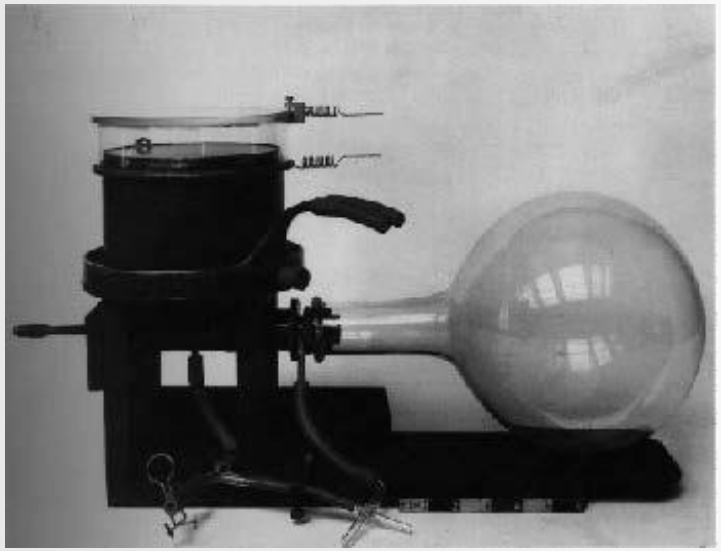




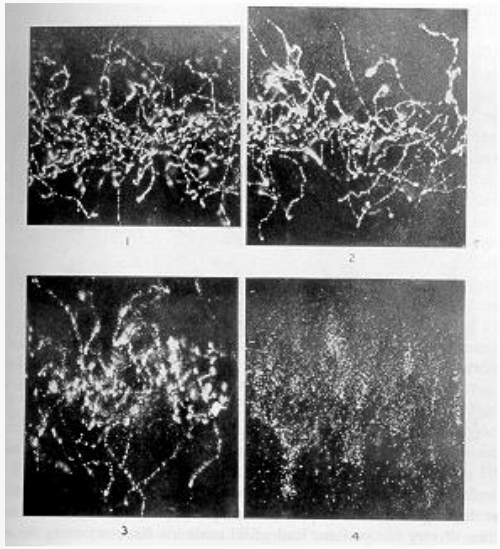


1.3 DISCOVERY BY IMAGING: CLOUD CHAMBERS

Wilson Cloud Chamber 1911



X-rays, Wilson 1912



α-rays, Phillip 1926

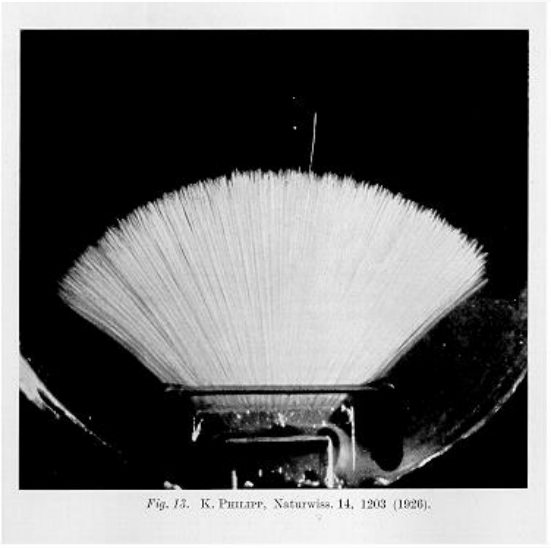
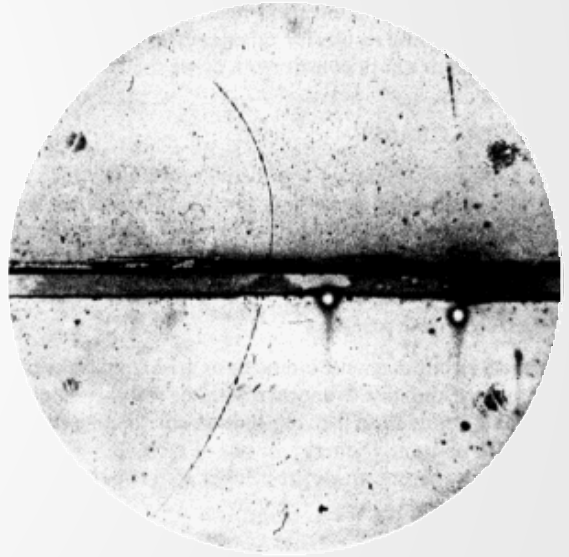


Fig. 13. K. Phillipp, Naturwiss. 14, 1203 (1926).

Ionization tracks from He_4 nuclei emitted from a radioactive source (Compare: energy loss by absorption, lecture 3)

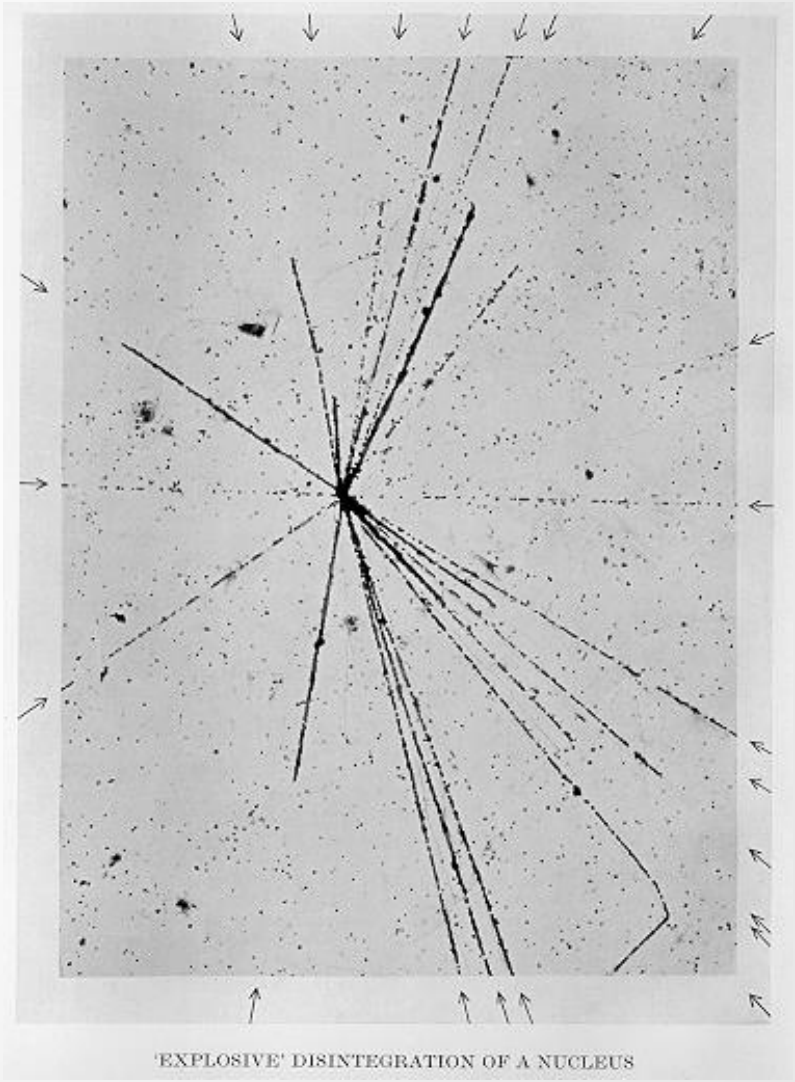
Positron discovery, Carl Andersen 1933



B=1.5T, chamber diameter 15cm.
A 63 MeV positron passes through a 6mm lead plate, loosing 23 MeV. All particle properties (ionization, etc.) except for the charge agree with the electron

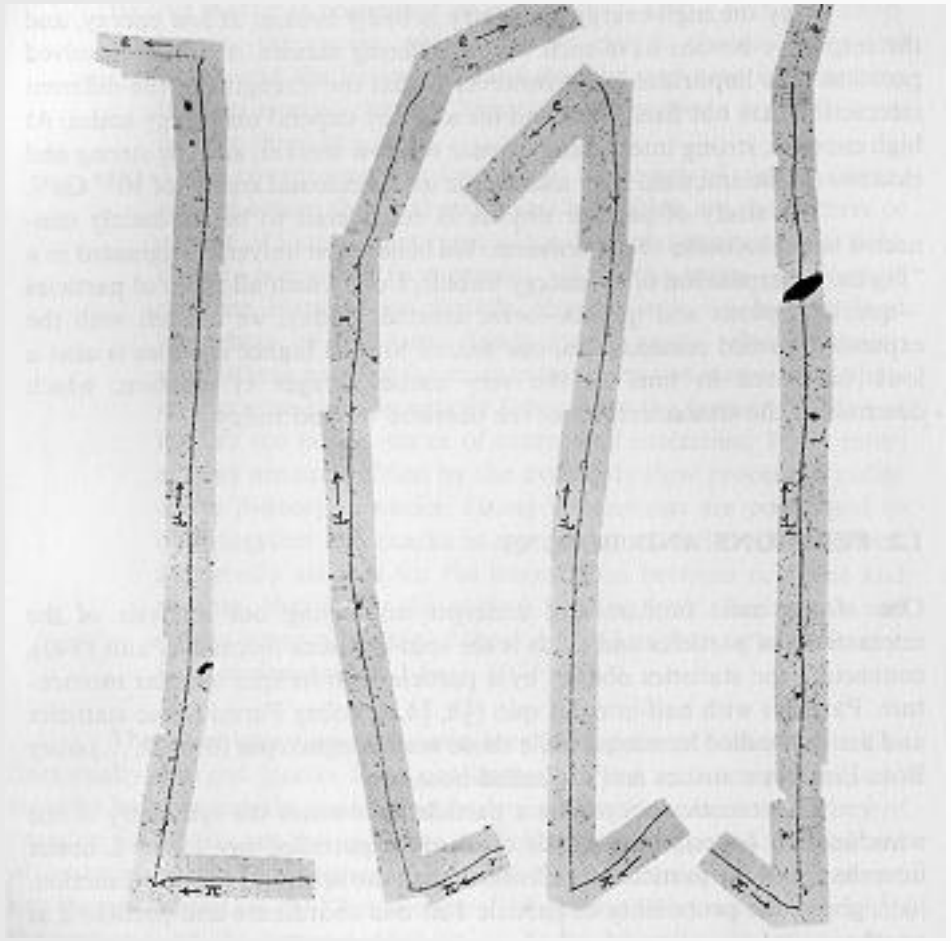
1.4 NUCLEAR EMULSION

- Film played an important role in the discovery of radioactivity but was first seen as a means of studying radioactivity rather than photographing individual particles.
- Between 1923 and 1938 Marietta Blau pioneered the nuclear emulsion technique.
- E.g. Emulsions were exposed to cosmic rays at high altitude for a long time (months) and then analyzed under the microscope.
- In 1937, nuclear disintegrations from cosmic rays were observed in emulsions. The high density of film (high stopping power dE/dx) compared to the cloud chamber gas made it easier to see energy loss and disintegrations.



'EXPLOSIVE' DISINTEGRATION OF A NUCLEUS

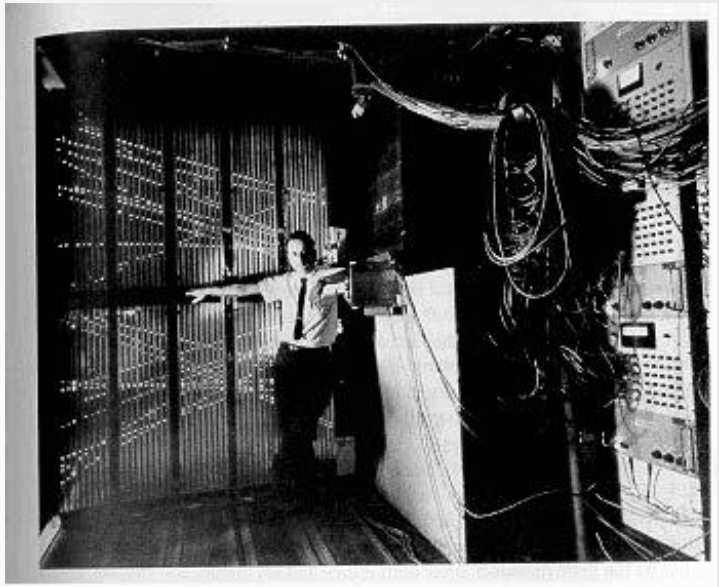
1.4 NUCLEAR EMULSION

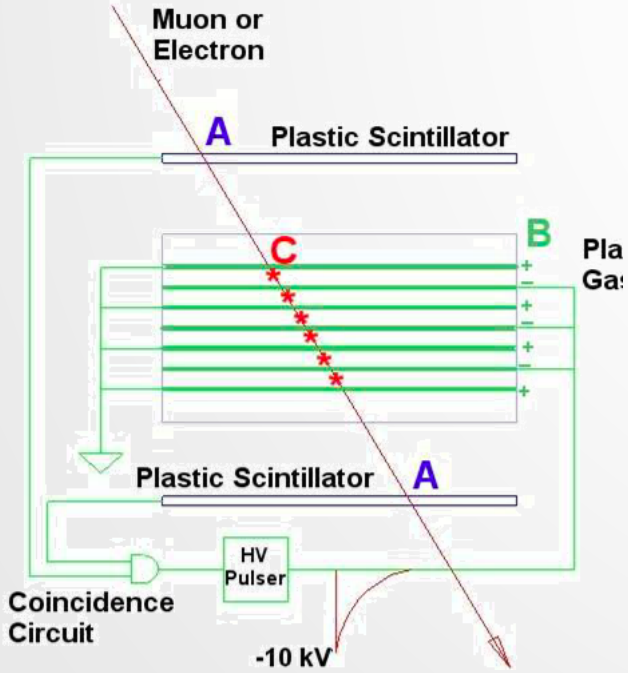


- Discovery of the pion
  - Muon discovery in the 1930s, first believed to be Yukawa's meson that mediates the strong force.
  - Problem with long range of the muon
  - In 1947, Powell et. al. discovered the pion in Nuclear emulsions exposed to cosmic rays, and they showed that it decays to a muon and an unseen partner.
    - $\pi^+ \rightarrow \mu^+ \rightarrow e^+$ from cosmic rays
  - The constant range of the decay muon indicated a two body decay of the pion.

1.5 SPARK CHAMBER – DISCOVERY OF THE MUON NEUTRINO

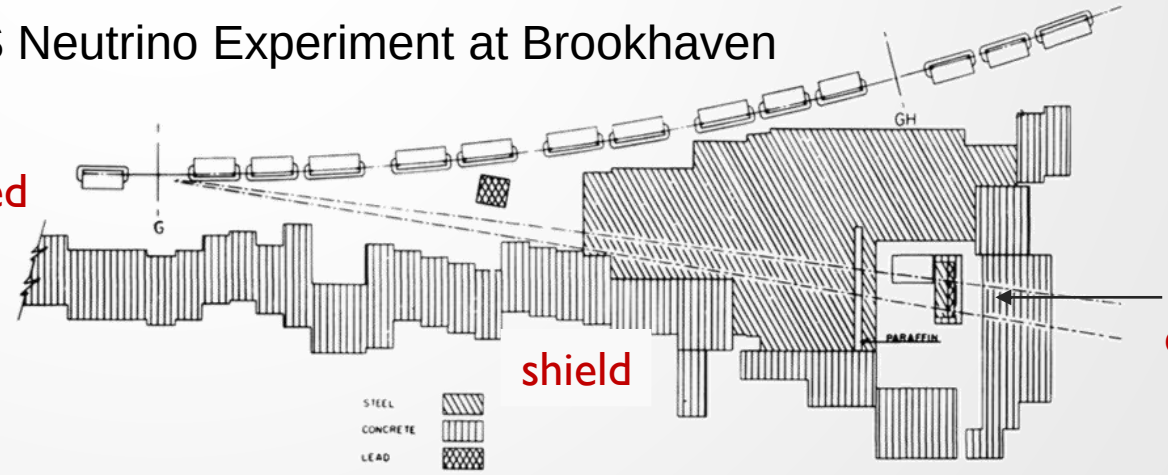
- The Spark Chamber was developed in the early 1960s.
- Schwartz, Steinberger and Lederman used it in discovery of the muon neutrino in 1962
  - A charged particle traverses the detector and leaves an ionization trail.
  - The scintillators trigger an HV pulse between the metal plates and sparks form in the place where the ionization took place.





AGS Neutrino Experiment at Brookhaven

pions produced



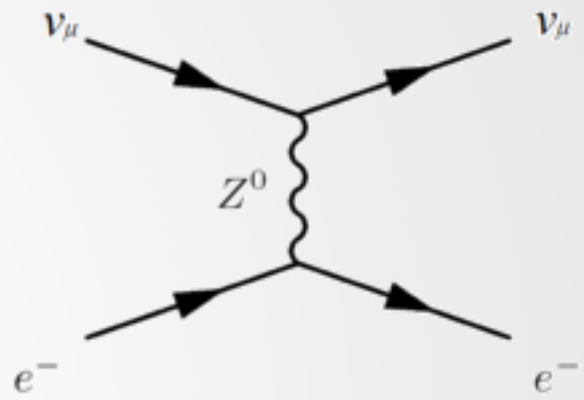
shield

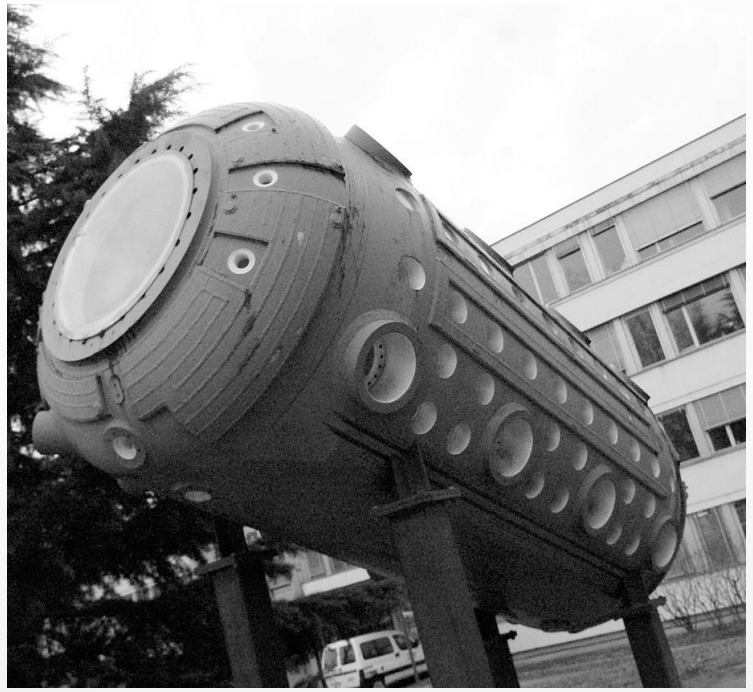
spark chambers

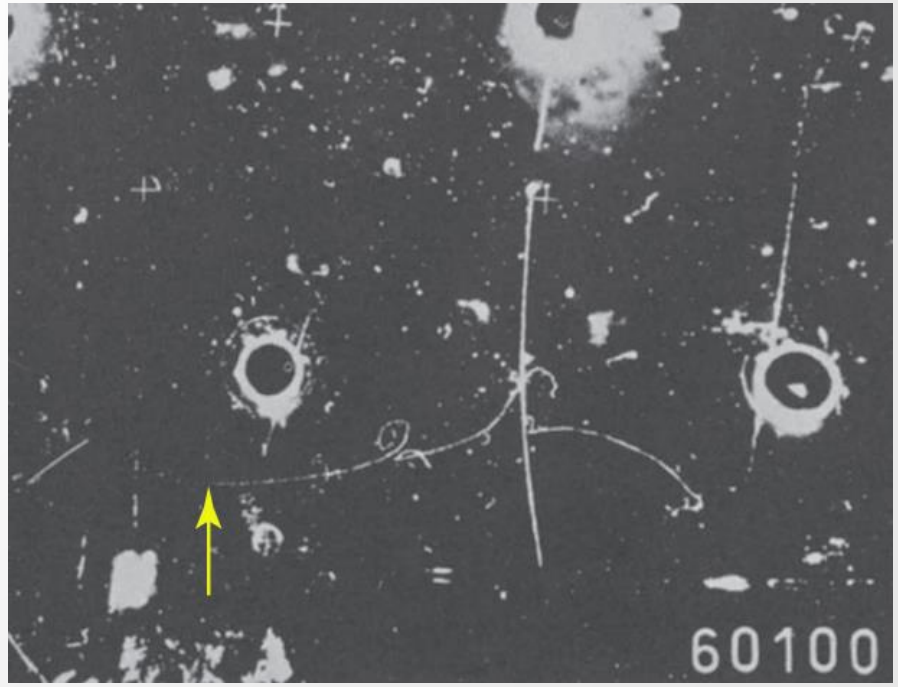
FIG. 1. Plan view of AGS neutrino experiment.

1.6 BUBBLE CHAMBER – DISCOVERY OF NEUTRAL CURRENTS

- Gargamelle, a very large heavy-liquid chamber, came to CERN in 1970.
  - 2m in diameter, 4 m long, Freon at 20 atm, conventional 2T magnet
  - Discovery of neutral currents in 1973 in liquid Freon
- Bottom right picture: muon neutrino from the left, interacts with an electron (yellow arrow) causing a subsequent EM "shower"

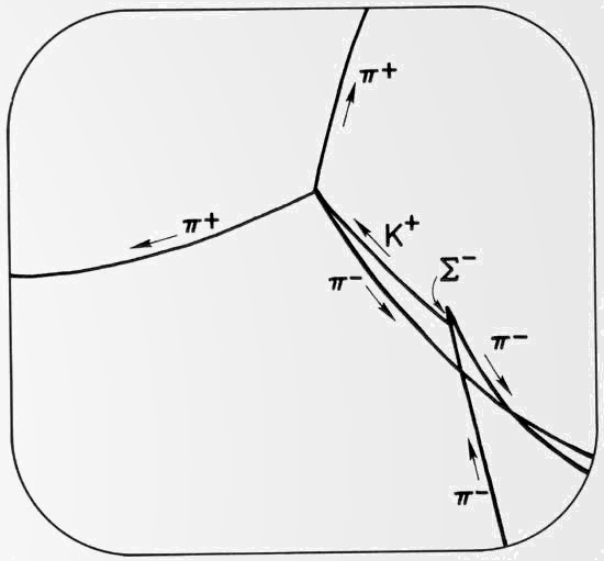


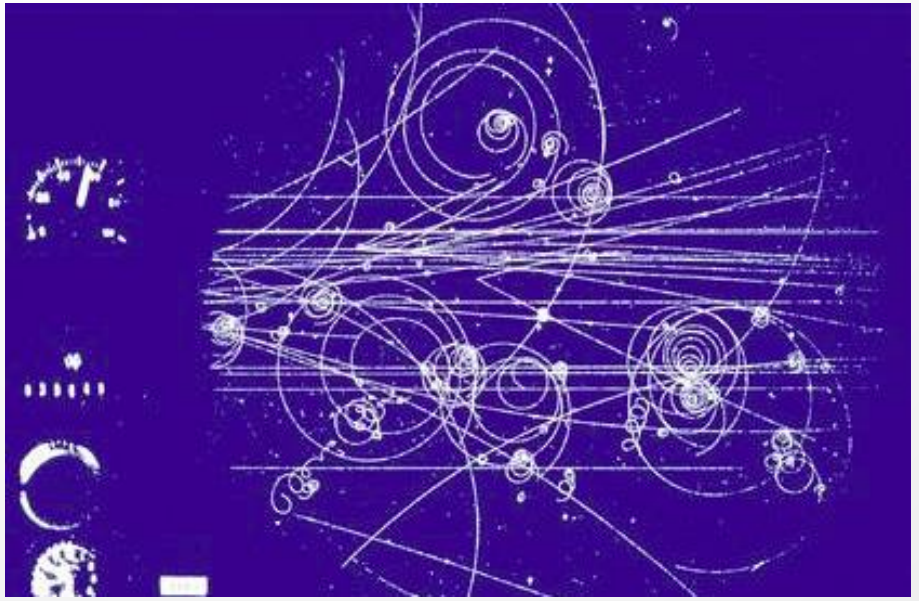


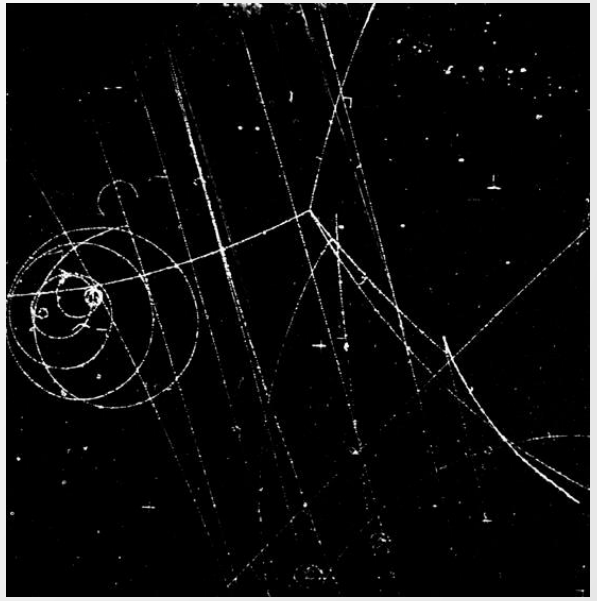


1.7 BUBBLE AND CLOUD CHAMBERS

- Bubble Chambers, Cloud Chambers have 4π coverage
  - Data acquisition system (DAQ) was a stereo photograph!
  - Effectively no trigger:
    - Each expansion was photographed based on accelerator cycle
    - High level trigger was human (scanners).
    - Slow repetition rate: Only most common processes were observed
  - Some of the high repetition experiments (>40 Hz) had some attempt at triggering.
- Emulsions still used in ν exp. (eg. CHORUS ν_μ/ν_τ oscillations)
  - Events selected with electronical readout detectors
  - scanning of emulsion seeded by external tracks

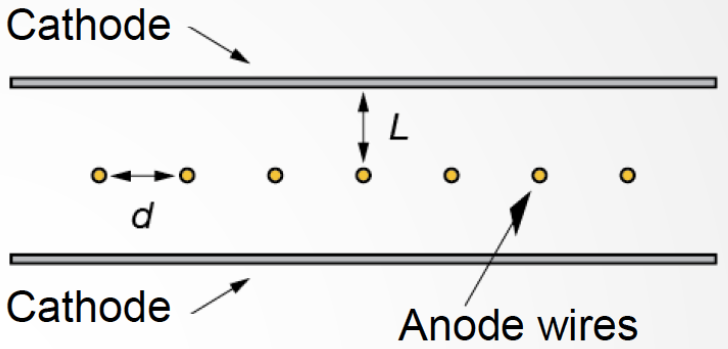


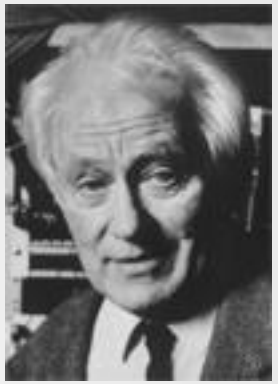




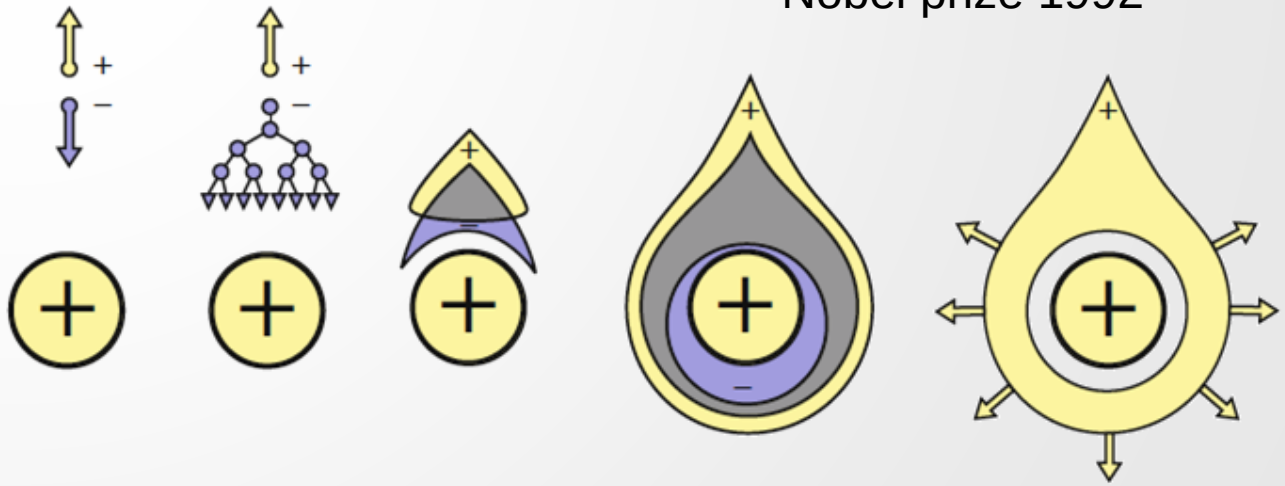
1.8 ELECTRICAL READOUT

- Move from photographs to electrical readout
- MWPC developed 1968 by Georges Charpak and others (R. Bouclier, F. Sauli, …).
- MWPC was the first full electronic detector! Every anode wire is connected to an amplifier and the signals can be processed electronically.
- The Nobel Prize in Physics 1992 was awarded to Georges Charpak "for his invention and development of particle detectors, in particular the multiwire proportional chamber".

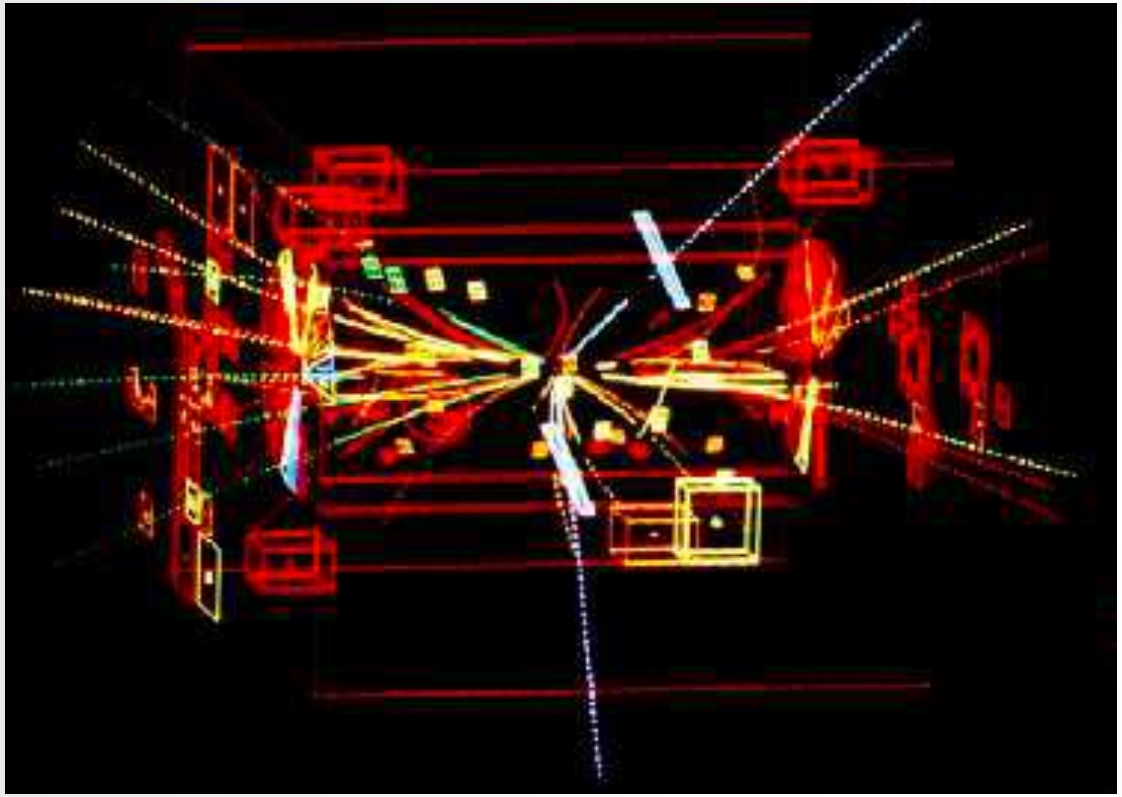




Georges Charpak
1924-2010
Nobel prize 1992

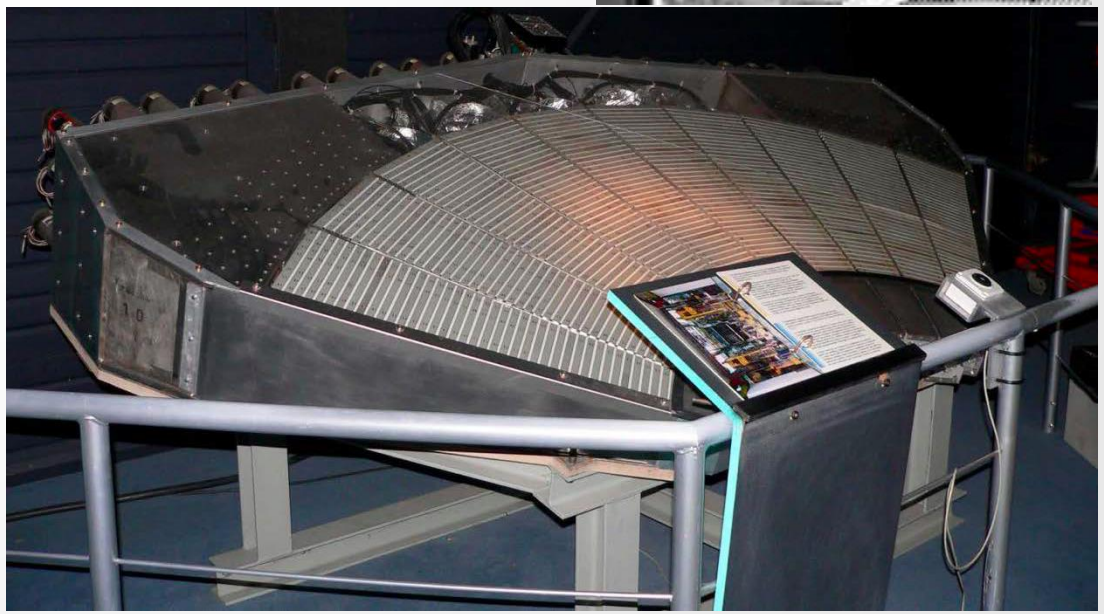


1.9 DISCOVERY OF THE W AND Z BOSON (83/84)



UA1 experiment at the SppS. Shown are tracks of charged particles from the proton-antiproton collision. The two white tracks reveal the Z⟶ee decay.

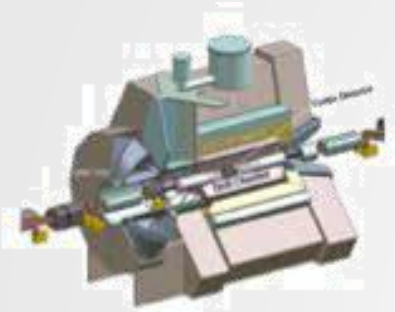
Carlo Rubbia
Simon van der Meer
(Stochastic cooling is particularly relevant for the Spps anti-protons)

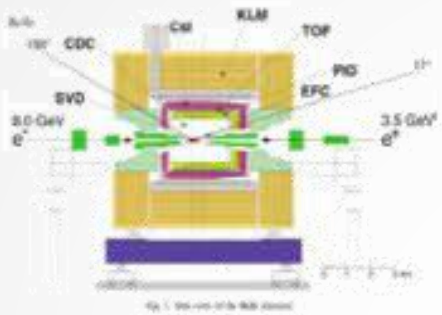


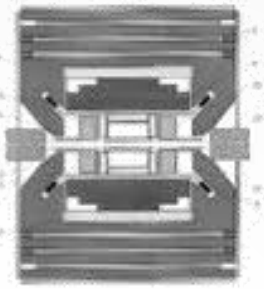
UA2 calorimetry
ECAL: lead/scintillator, HCAL: iron/scintillator sandwich
Both read out with photomultipliers

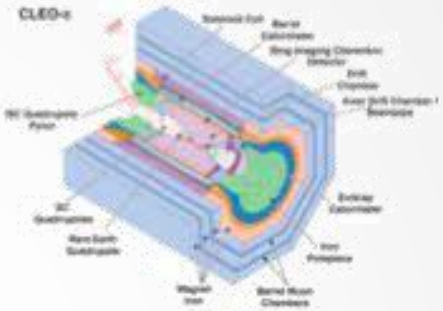
1.10 LARGE RESEARCH CENTERS

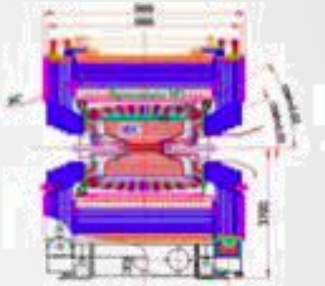
A. Korytov, Presentation at INSTR'08, Novosibirsk, 2008

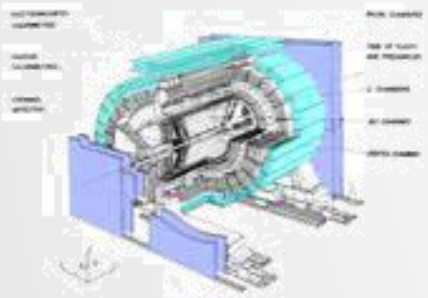


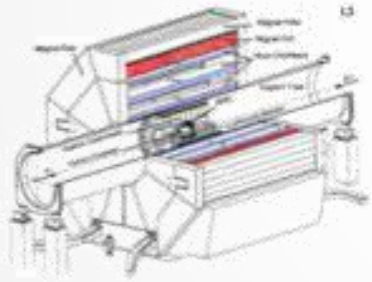


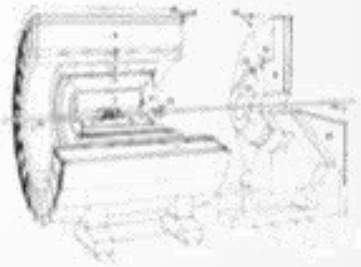


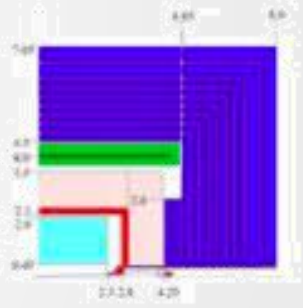


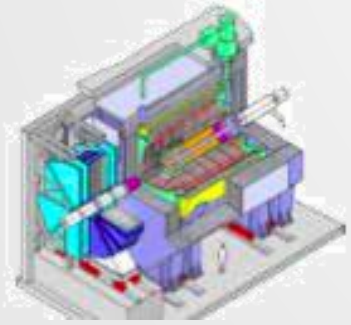


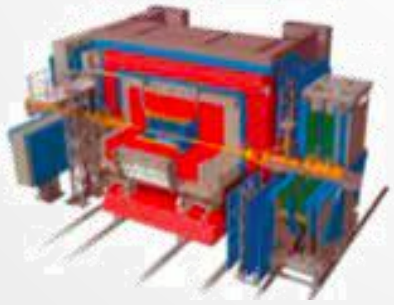


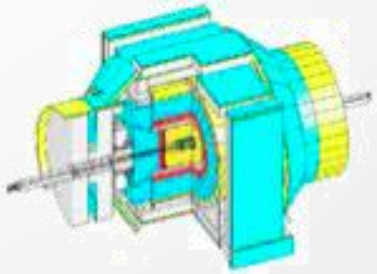


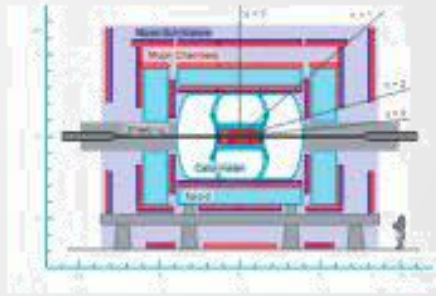




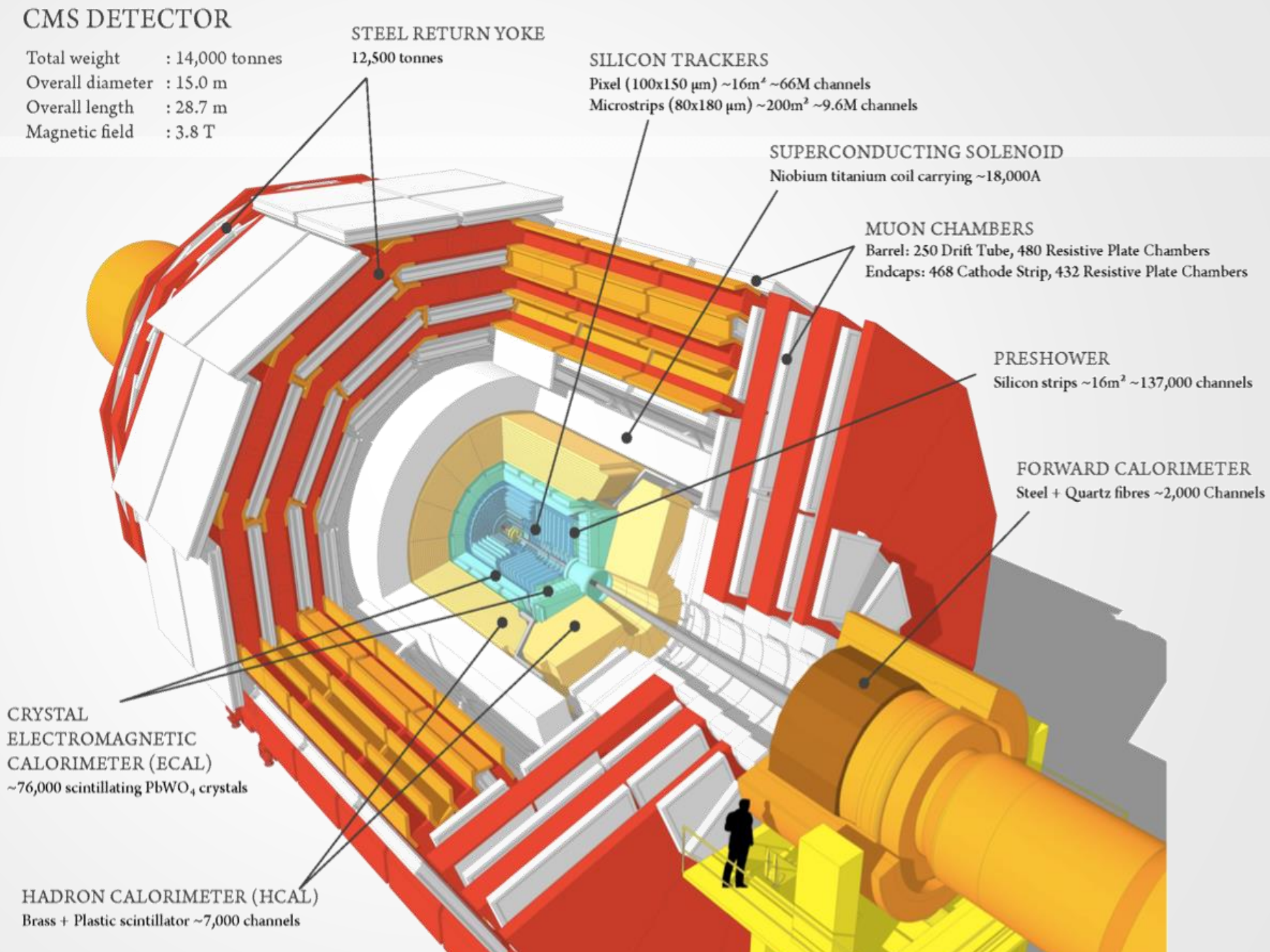




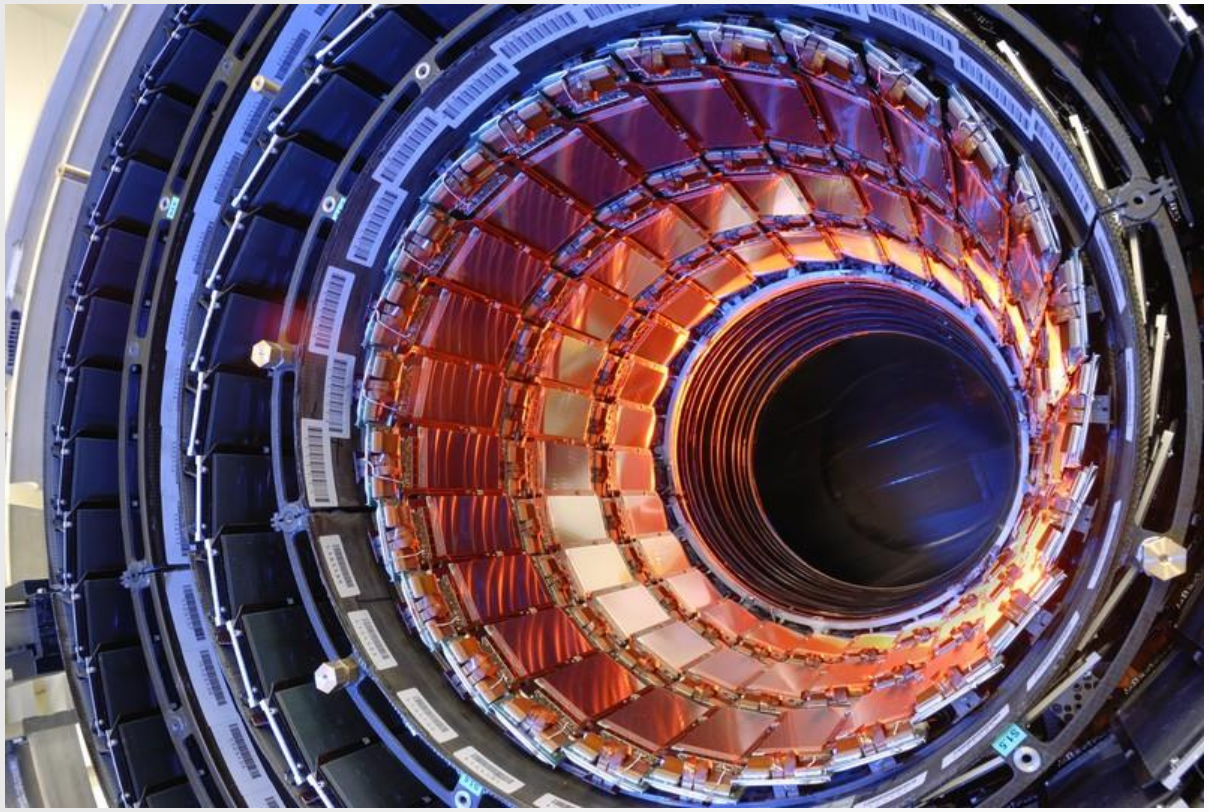


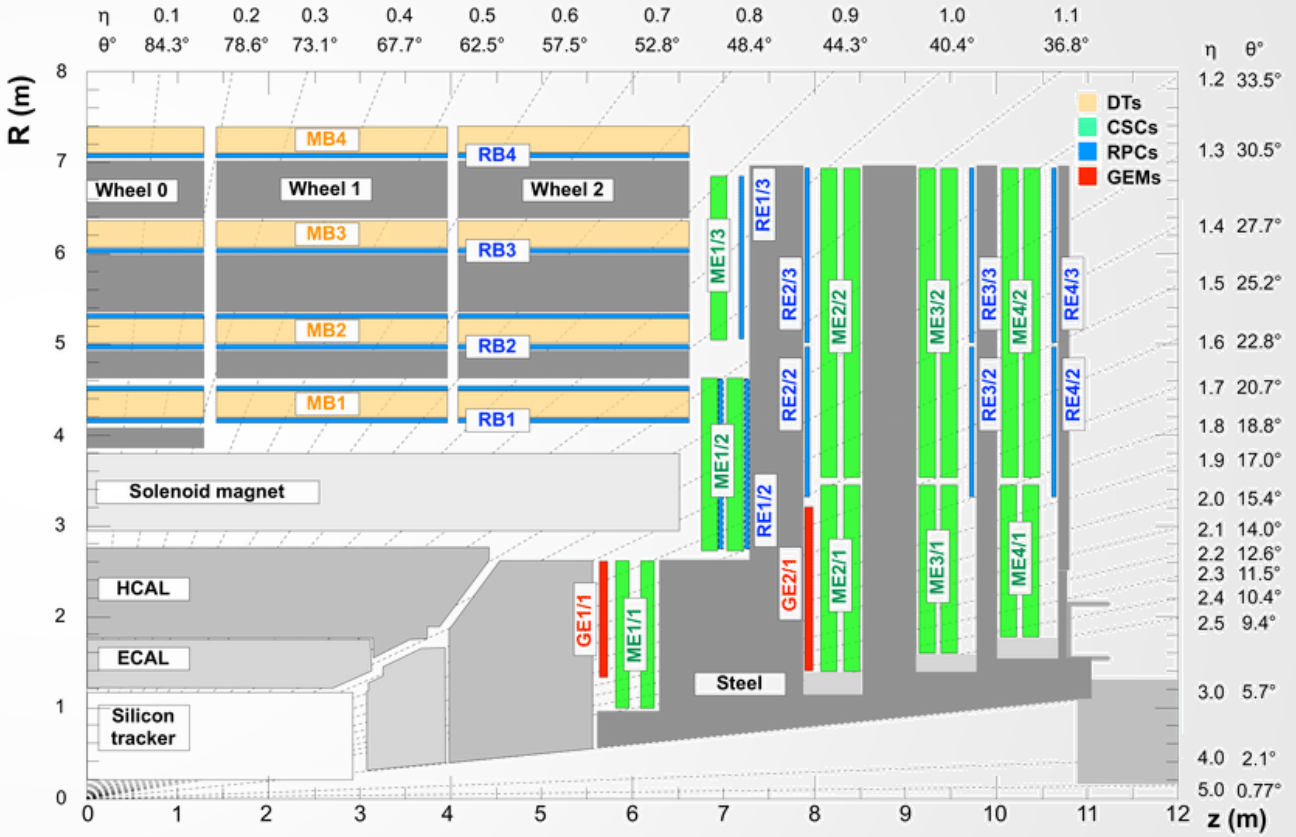


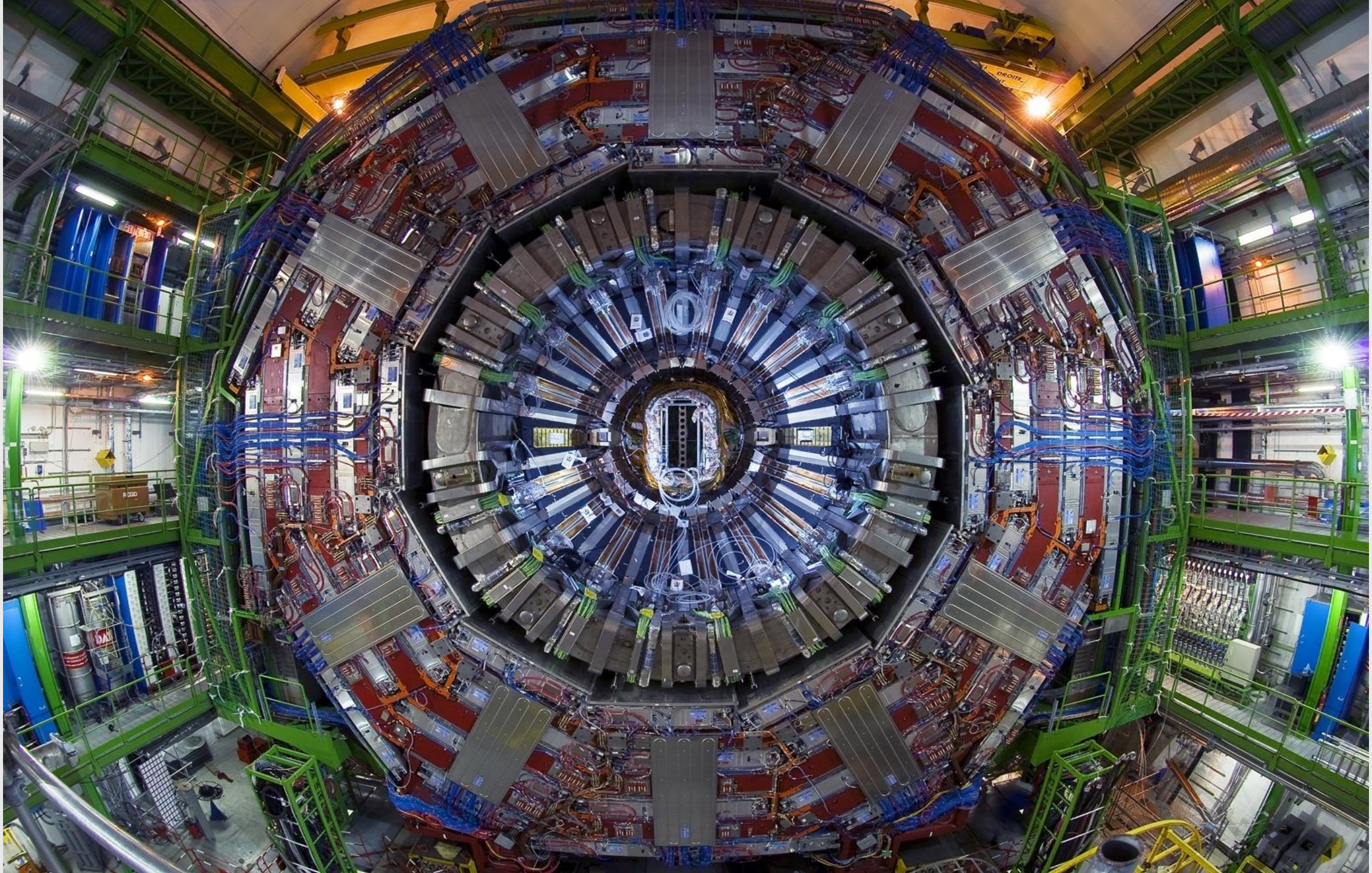
1.11 CMS DETECTOR



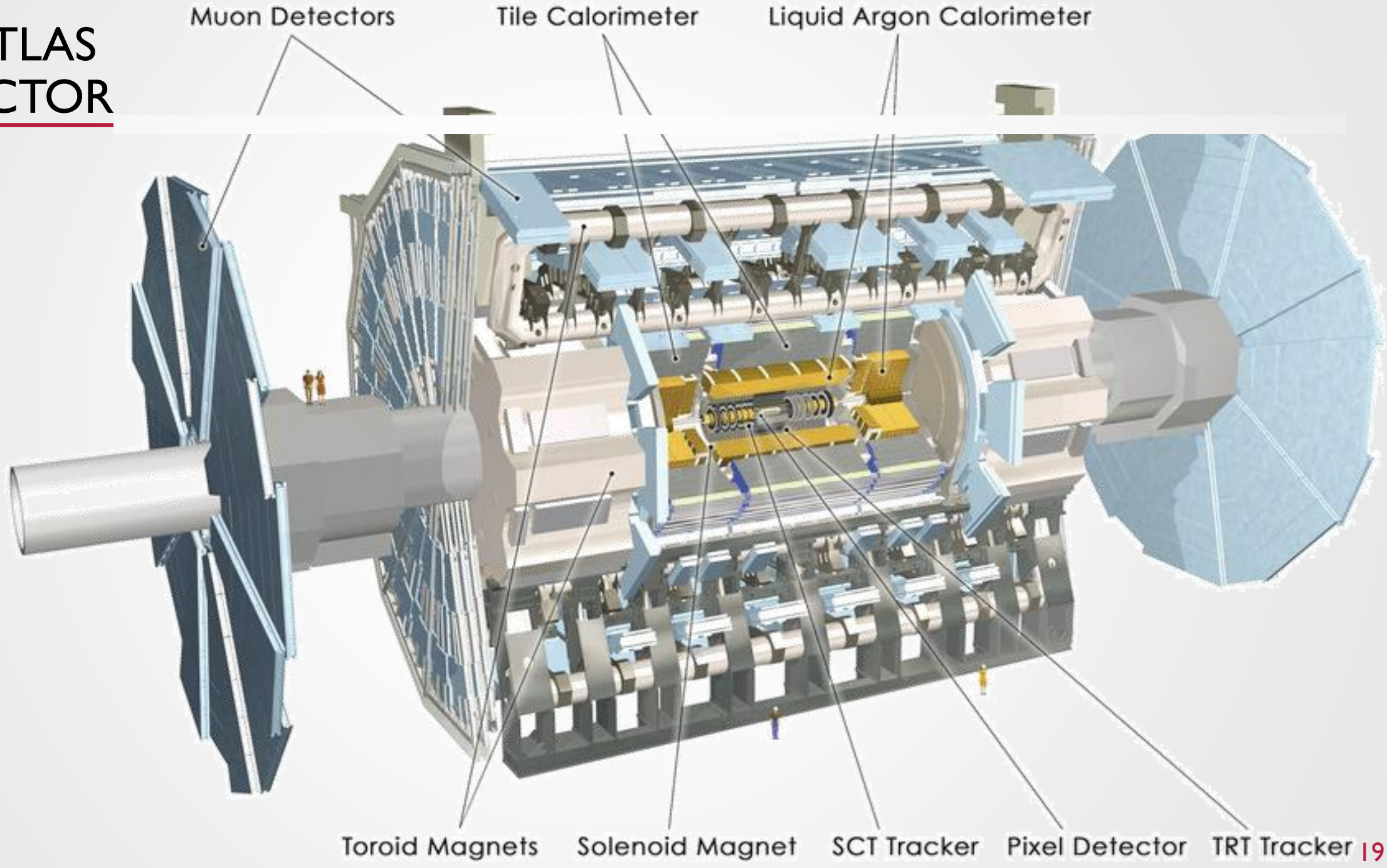
1.11 CMS DETECTOR



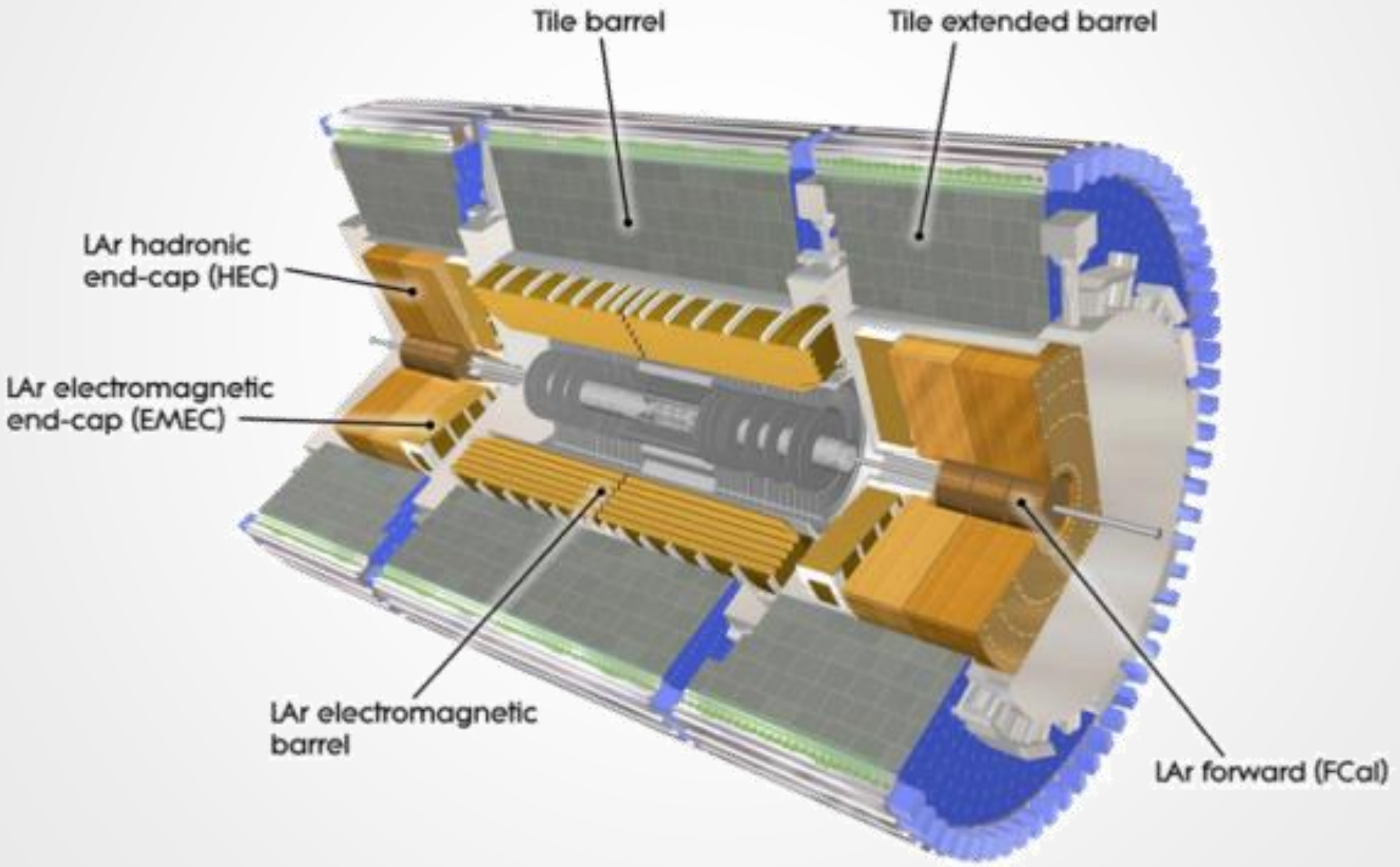




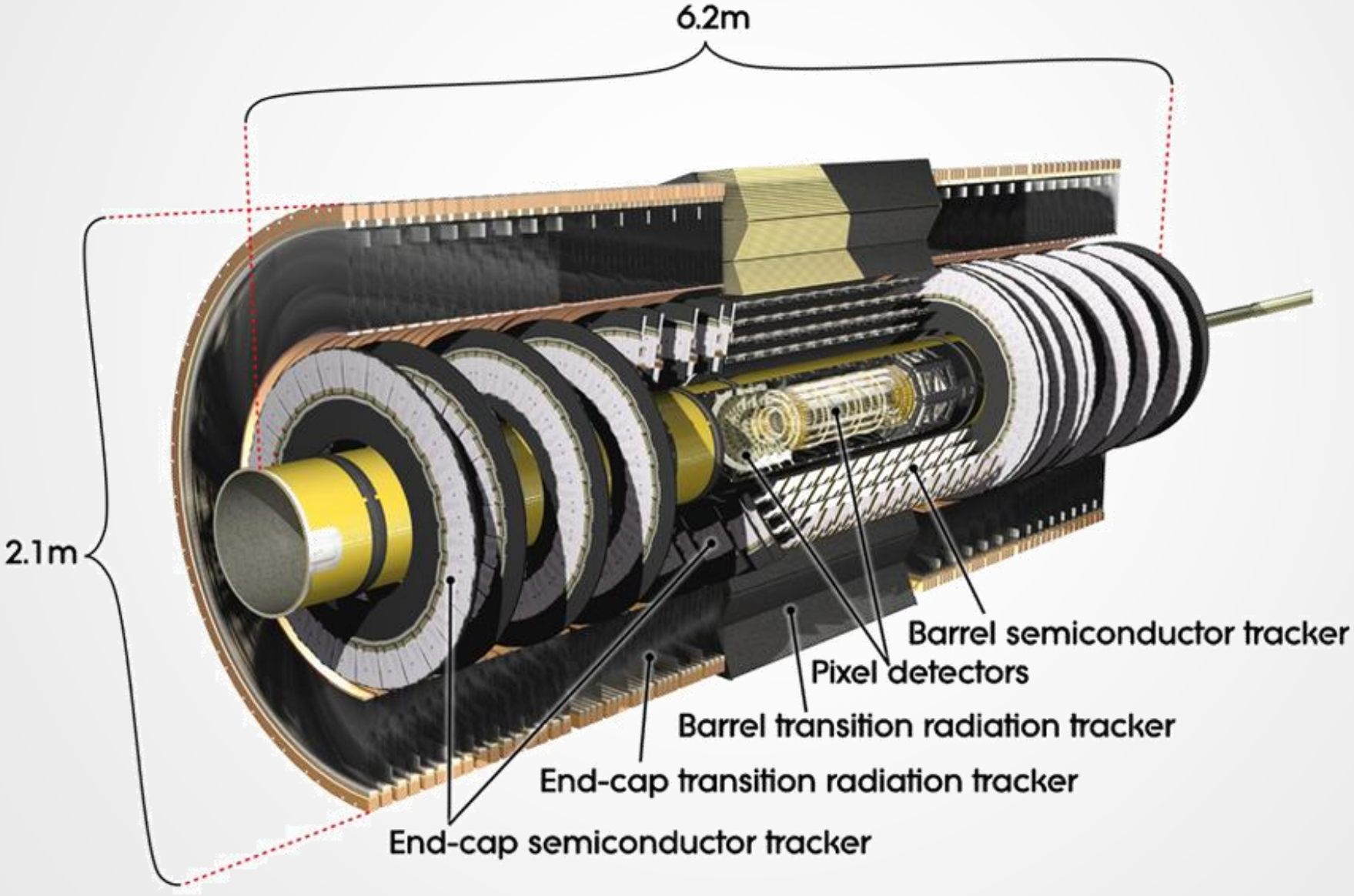
1.12 ATLAS DETECTOR



1.12 ATLAS DETECTOR (CALORIMETRY)



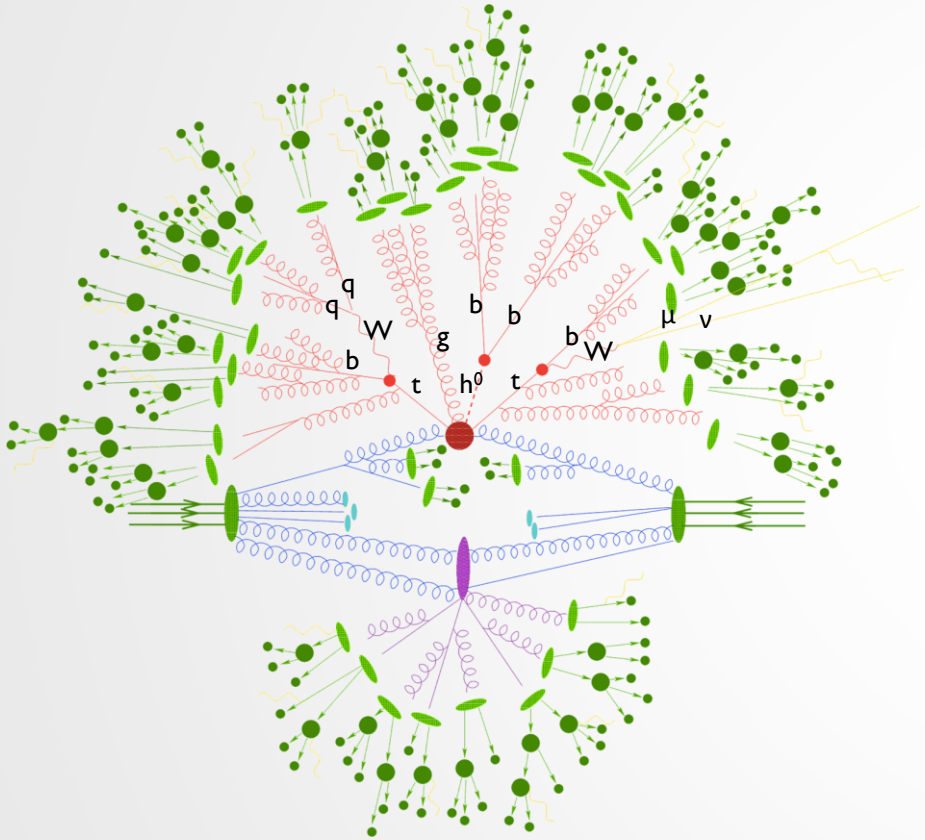
1.12 ATLAS DETECTOR (INNER DETECTOR)



2. EVENT RECONSTRUCTION

2.1 REMINDER: A COMPLEX SIMULATED EVENT

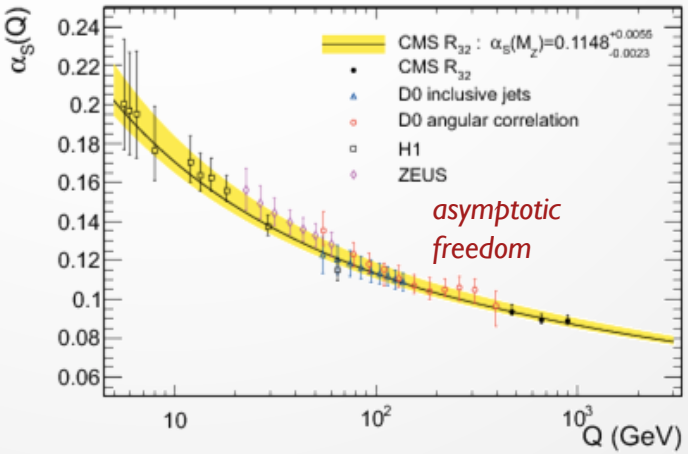
parton = quarks or gluons in the proton



- A ttH event has decayed as ttH→bW bW bb→ bbbbqq$\mu\nu$
- reconstruct hard parton collision even for collimated decay products?

Sketch of a hadron-hadron collision (simulated).

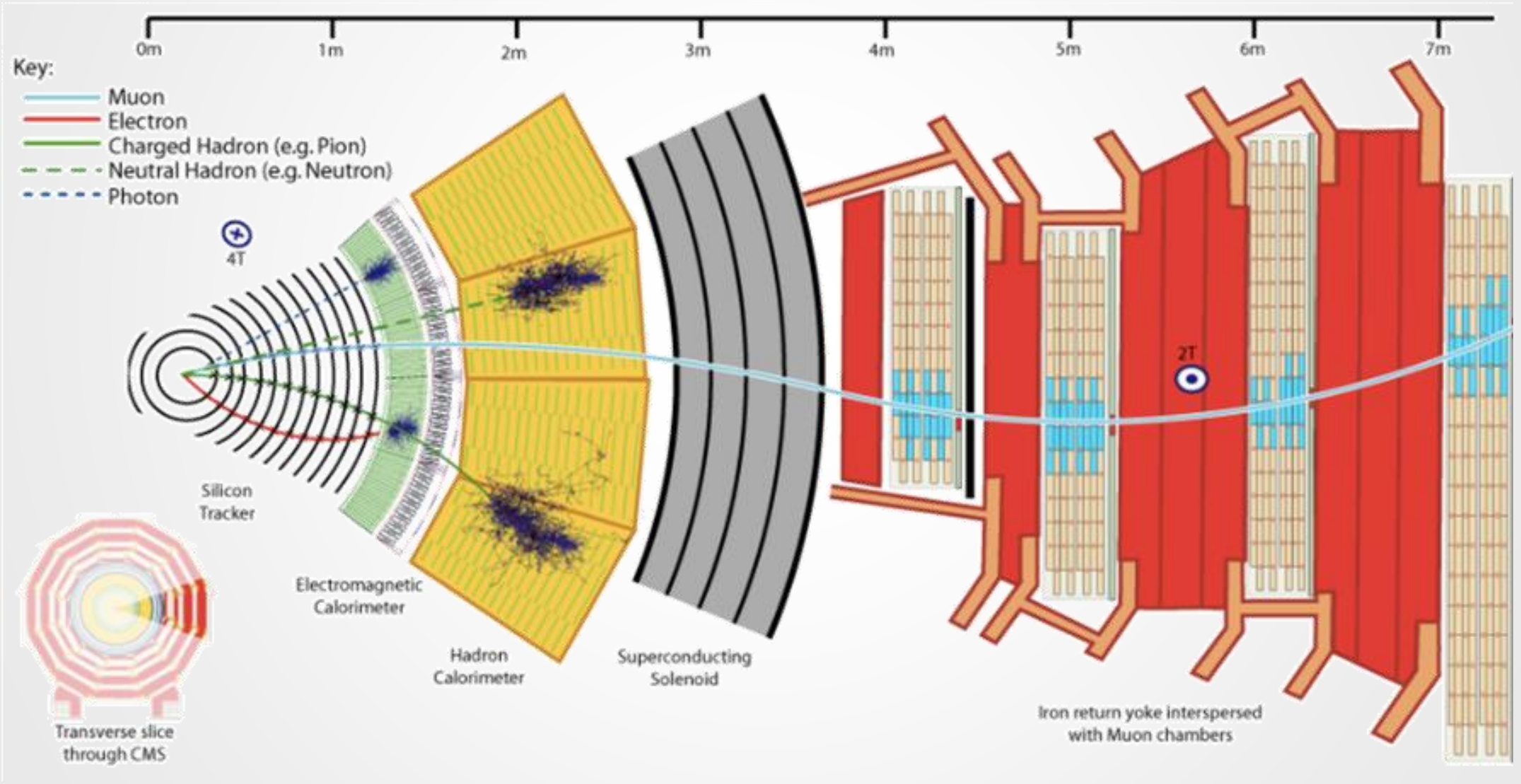
- hard parton collision (calculable)
- decays (calculable) and bremsstrahlung ("parton showers", effective model)
- secondary hard scattering
- Parton-to-hadron transitions
- hadron decays, while yellow lines signal soft photon radiation.



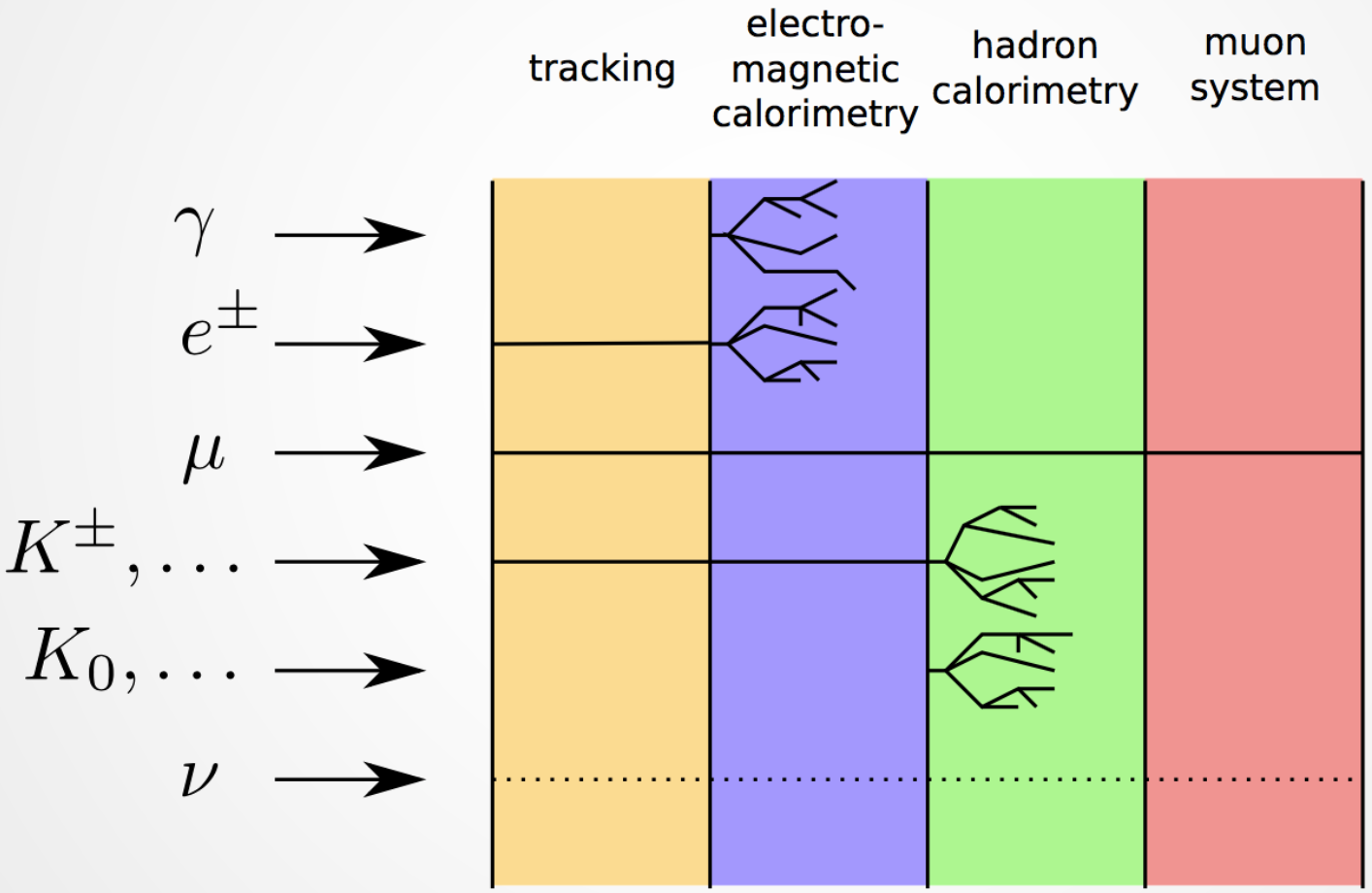
Only at high energies, can we perturbatively calculate QCD!

Gross/Politzer/Wilczek Nobel prize 2004

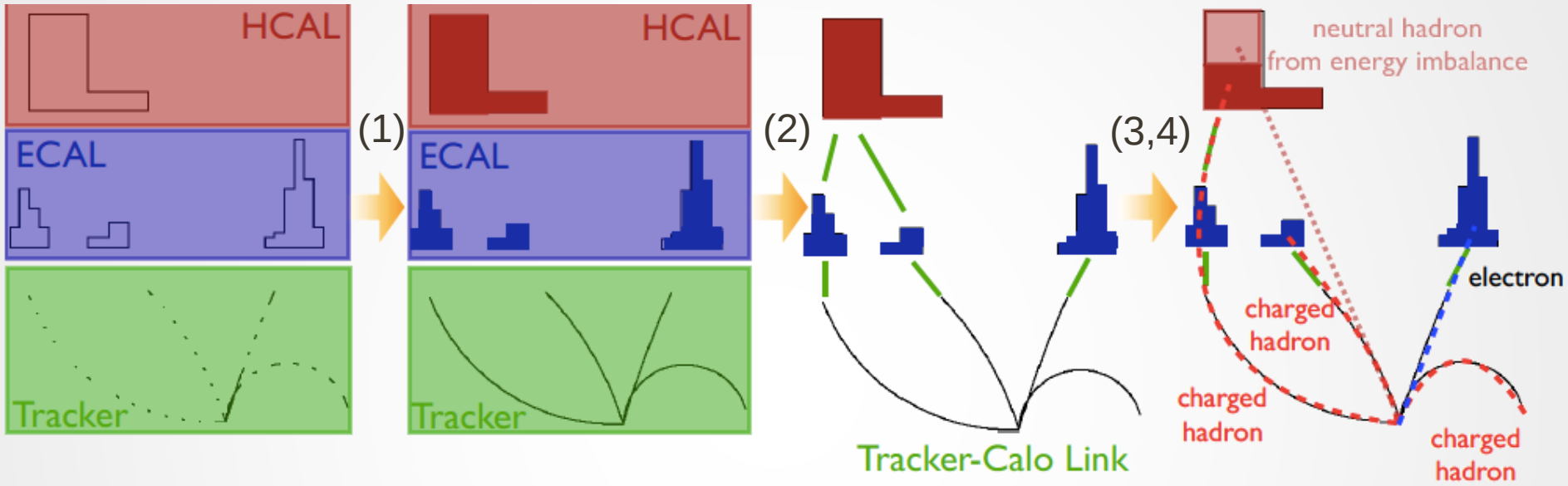
2.2 REMINDER: CMS OVERVIEW



2.3 REMINDER: OVERVIEW OF DETECTOR SIGNATURES



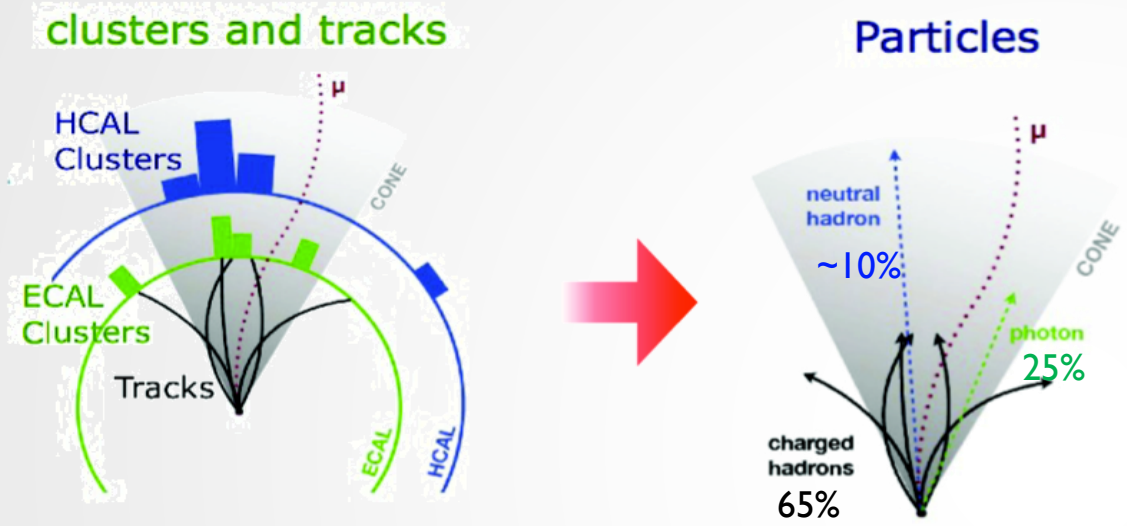
2.4 PRINCIPLES OF EVENT RECONSTRUCTION

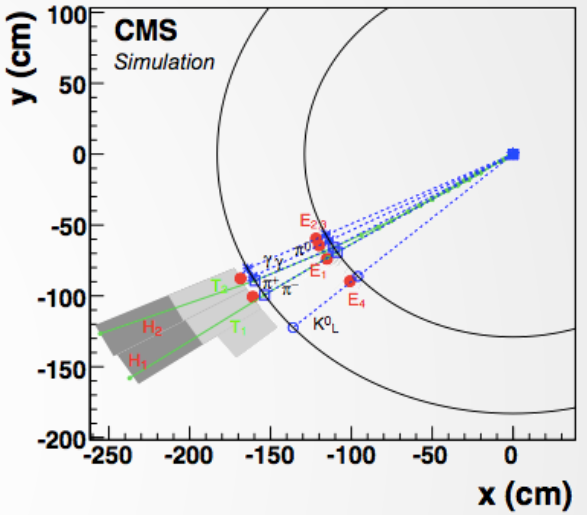


When an event is recorded, the hits in the detector cells are stored. Main algorithmic steps are:

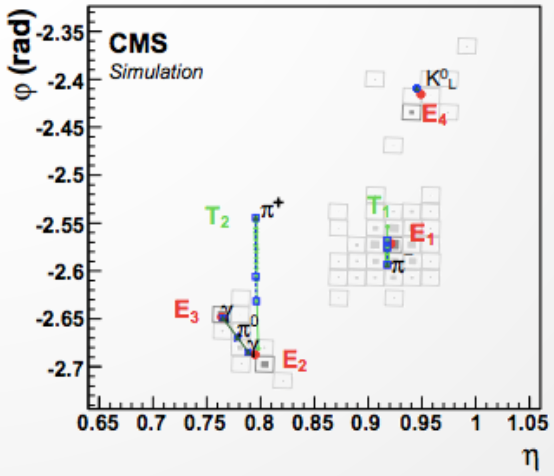
1. Build muon candidates (not shown), tracks, and calorimetry clusters
2. Link tracks and the calorimetry clusters based on spatial proximity
3. Identify 'charged hadron candidates' among the links by associating calorimetric energy to track momenta, when tracks are close
4. 'photon' and 'neutral hadron' candidates from excess energy

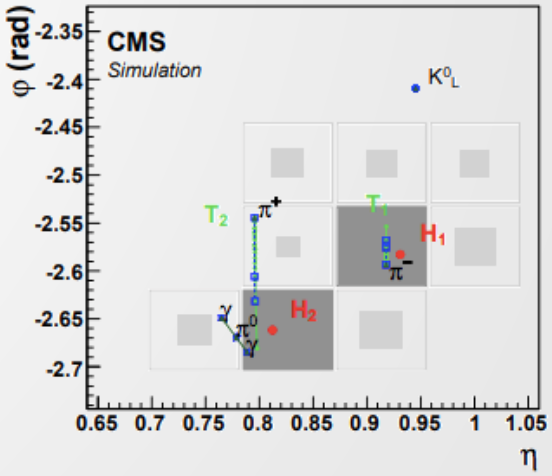
2.5 PARTICLE FLOW EVENT RECONSTRUCTION





- reconstruct charged/neutral hadrons, γ,$e^{\pm}$,$\mu^{\pm}$
- fully exploit tracking resolution and fine ECAL granularity
- optimal combination of track- and calorimetry resolution
  - resolution: track: <1% at low p_T, photons: ≈ 3%/sqrt(E), neutral hadron: ≈ 100%/sqrt(E)
  - Note: E(jet) > 1 TeV: calorimeter resolution dominates because $\sigma(p)/p \propto p$ and thus the tracker measurement has large uncertainty





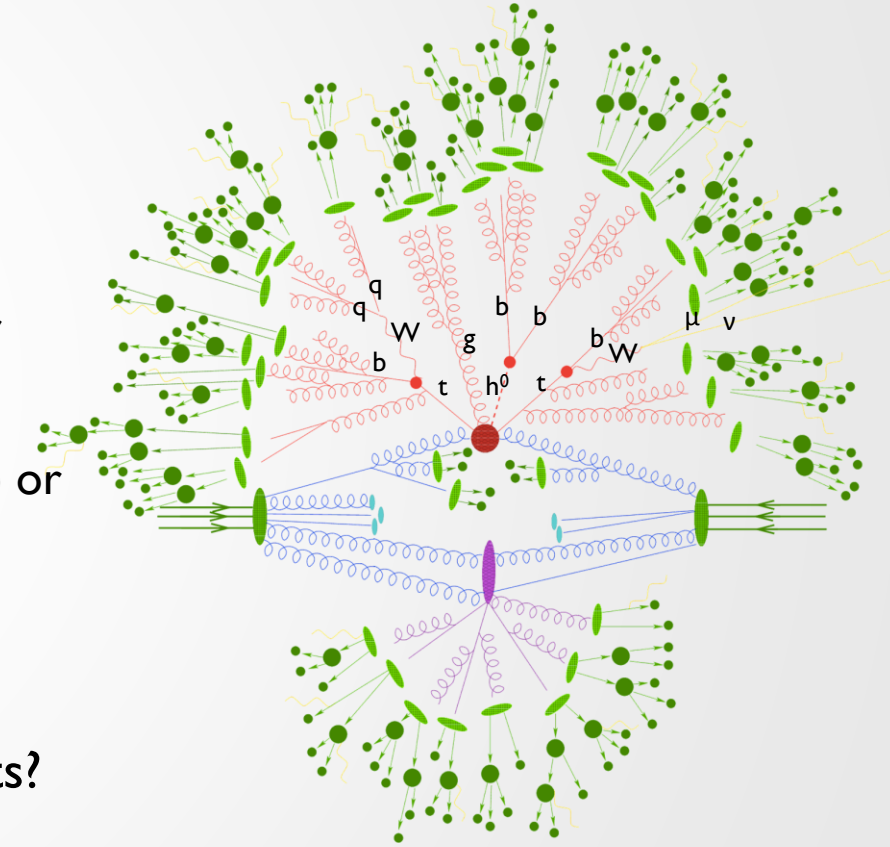
2.5 PARTICLE FLOW EVENT RECONSTRUCTION

	detector	Tracker	ECAL	HCAL	Muon	
	quantity	p_T(Trk)	E_{ECAL}	E_{HCAL}	$p_T(\mu)$	p_T and E reconstructed by
	algorithm	fit	cluster	cluster	fit	
	object	**reconstruction**				
1.	muon	✓	✗	✗ (except ~1GeV MIP)	✓	'global' fit of tracker- and muon-track
2.	electron	✓	✓	✗	✕	fit, dominated by p_T(Trk) at low energy, ECAL at high energy
3.	charged hadron	✓	✓	✓	✕	track-p_T and E_{ECAL} , $E_{HCAL,}$ linked to track
4.	photon	✗	✓	✗	✕	E_{ECAL} (not linked to track or HCAL)
5.	neutral hadron	✗	✓	✓	✕	E_{ECAL} + E_{HCAL} energy that was not linked to a track

✓ required ✓ allowed ✕not expected/ignored at this step ✗ vetoed

2.6 JET CLUSTERING

- Event reconstruction provides e.g. list of particle candidates
  - Need to correlate 'sprays' of particles from hadronization (jets) with the initial partons of the hard scatter event
  - This is done by "sequential" jet clustering algorithms:
    - Identify seeds and devise a recursive procedure to associate other particles until the whole event is clustered
    - Controlled by an angular distance measure e.g. ΔR=0.4 ('slim' jets) or 0.8/1.0/1.2 ('fat' jets) that defines the angular size of the jets
      - Remember Lecture 2: $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ is a boost invariant angular distance measure
  - What other criteria are relevant when clustering particles to jets?



2.6 JET CLUSTERING

JHEP 0804:063,2008

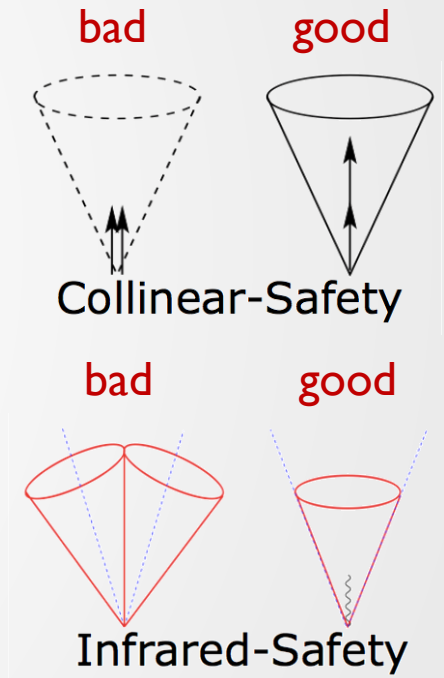
- The parton shower is governed by QCD and contributes many soft and collinear particles
  - Collinear splitting & soft ('infrared') particles shouldn't change jets
- Jet 'catchment' area should be a disk of radius ΔR in $\Delta\eta, \Delta\varphi$ coordinates, even in dense environment of many pile-up particles
- Anti-k_T algorithm satisfies all criteria!
  1. Select a cone size R (e.g. R=0.4)
  2. For particle i, compute all distances d_{ij} and d_{iB}.

$$d_{iB} = \frac{1}{p_{Ti}^2}, \qquad d_{ij} = \min\left(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2}\right)\frac{\Delta R_{ij}^2}{R^2}$$

p_T^{-2} prefers early merge of close & energetic particles

  3. If a pair (ij) has smallest distance in d_{ij}, merge & add momenta. Repeat step 2.
  4. Otherwise label jet, remove from list, start again with 2. until fully clustered.

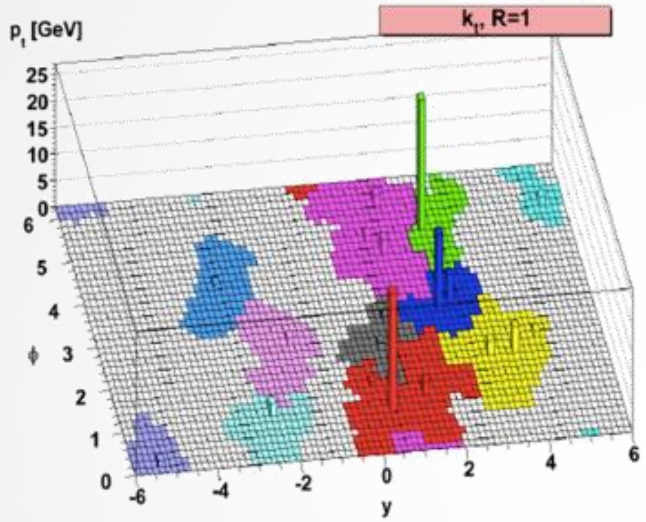
Desired properties of jet clustering algorithms:

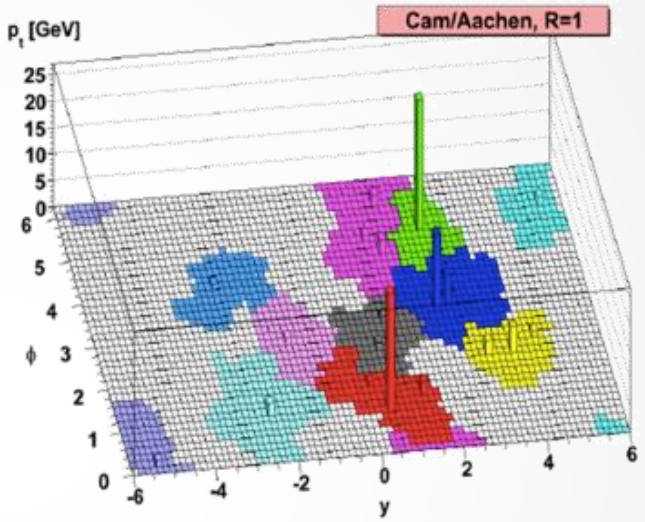


2.7 JET SHAPES FOR DIFFERENT ALGORITHMS

JHEP 0804:063,2008

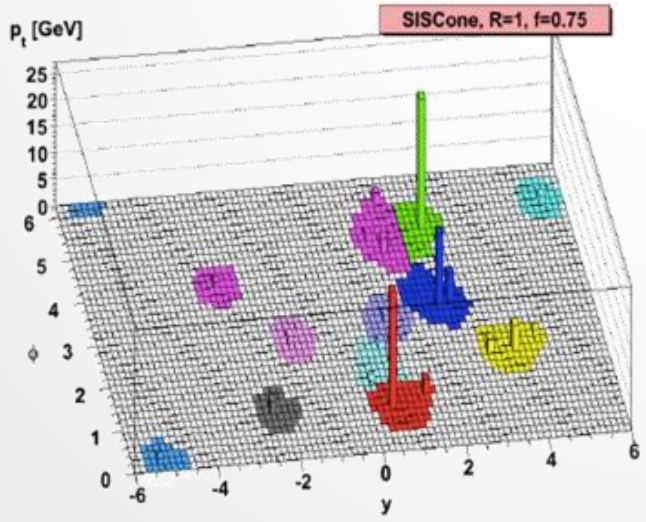
k_T algorithm
use p_T^2 instead of p_T^{-2}
late merge of high p_T objects (interesting for jet substructure information)

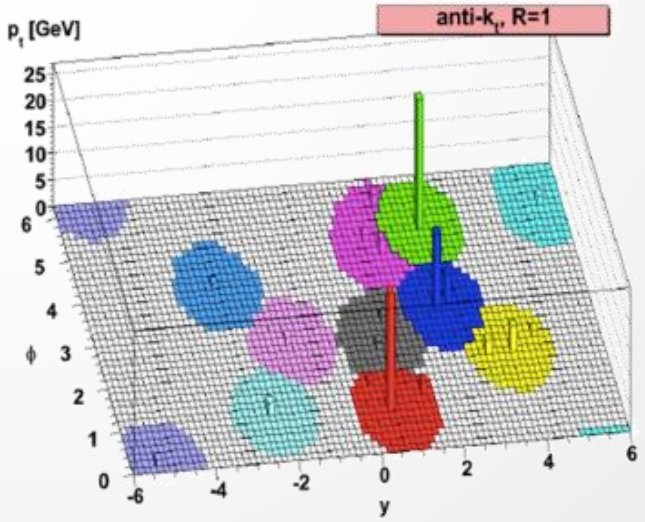




Cambridge-Aachen (CA)
Same as kT but only using ΔR

SisCone algorithm
Tries to find 'stable cones' when iterating an association procedure





Anti-k_T
Standard algorithm

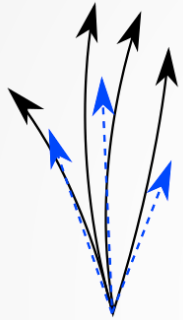
2.8 INFORMATION ON JET SUBSTRUCTURE

- secondary vertex

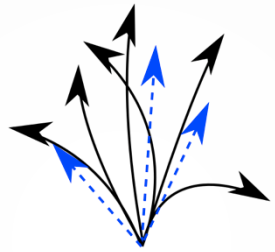
neutral particles

charged particles

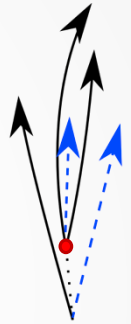
light quarks



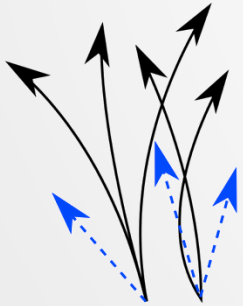
gluons



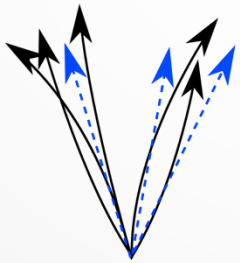
b/c



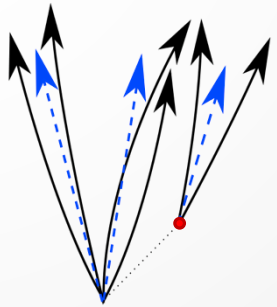
boosted heavy resonances clustered into a jet



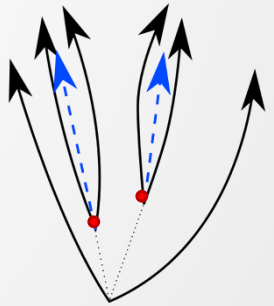
pile-up



W/Z



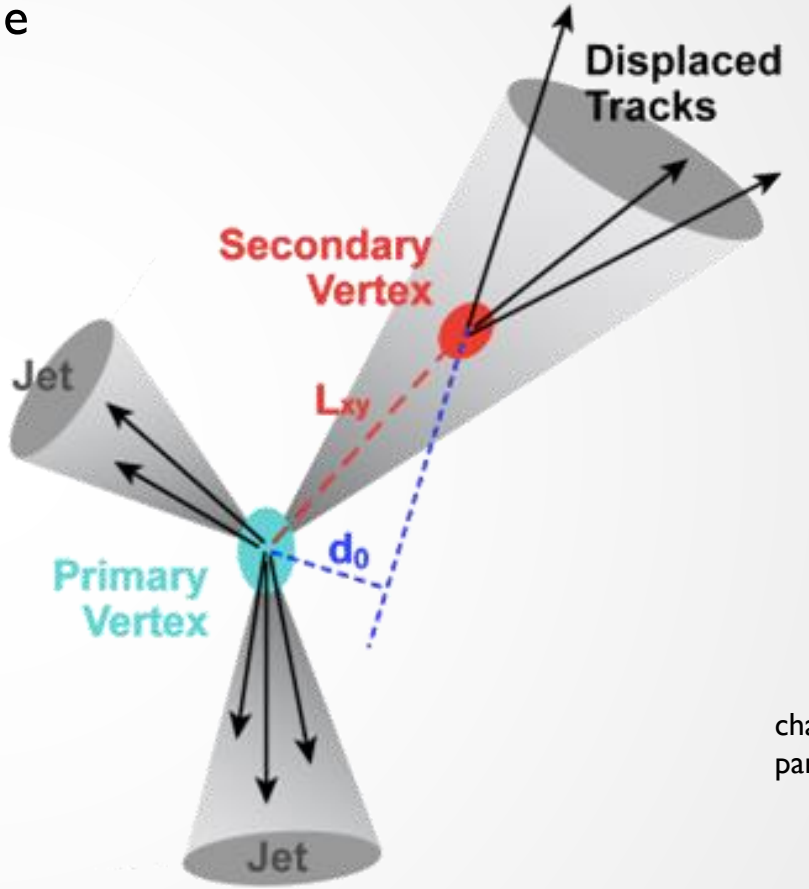
top



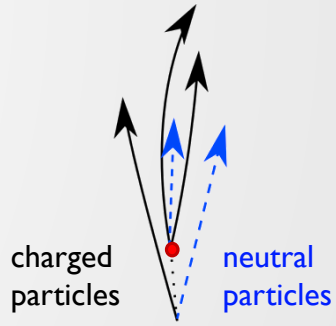
Higgs

2.9 IDENTIFICATION OF B QUARKS (B-TAGGING)

- in many physical processes, b-quarks play a crucial role
  - e.g. t⟶bW with BR≈100%
- b-quarks hadronize into B-hadrons
- B-hadrons have a finite life time
  - displaced particles are clustered in jet
- Identify b-jets by the properties of the decay products
  - B-hadron with ~5 GeV mass
  - large life-time $c\tau \approx 450\mu m$, at E=70 GeV: $\beta\gamma c\tau \approx 5mm$
    - displaced vertex identified by finding tracks that cross at large impact parameter L_{xy}, d_0
  - ~3 tracks at the displaced vertex
  - potentially a lepton at the displaced vertex
  - potentially tertiary vertex (B-meson decay to charmed hadron $c\tau \approx 120\text{-}310\mu m$)
- all information is used in MVA classifier: typically find 60% at 1% mistag

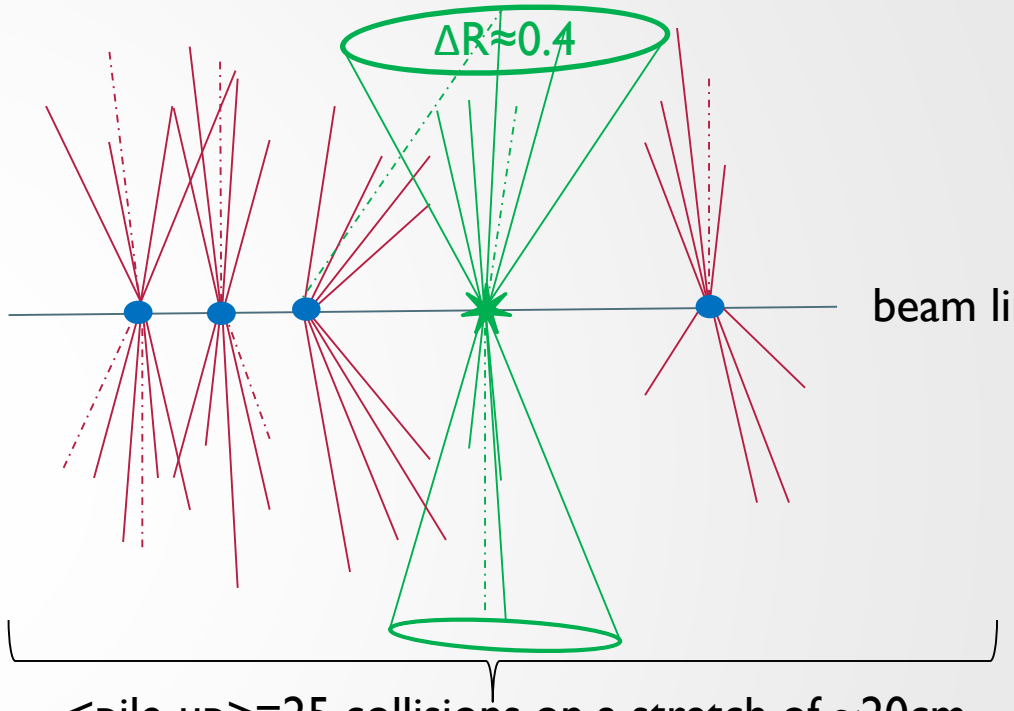


b (or c)

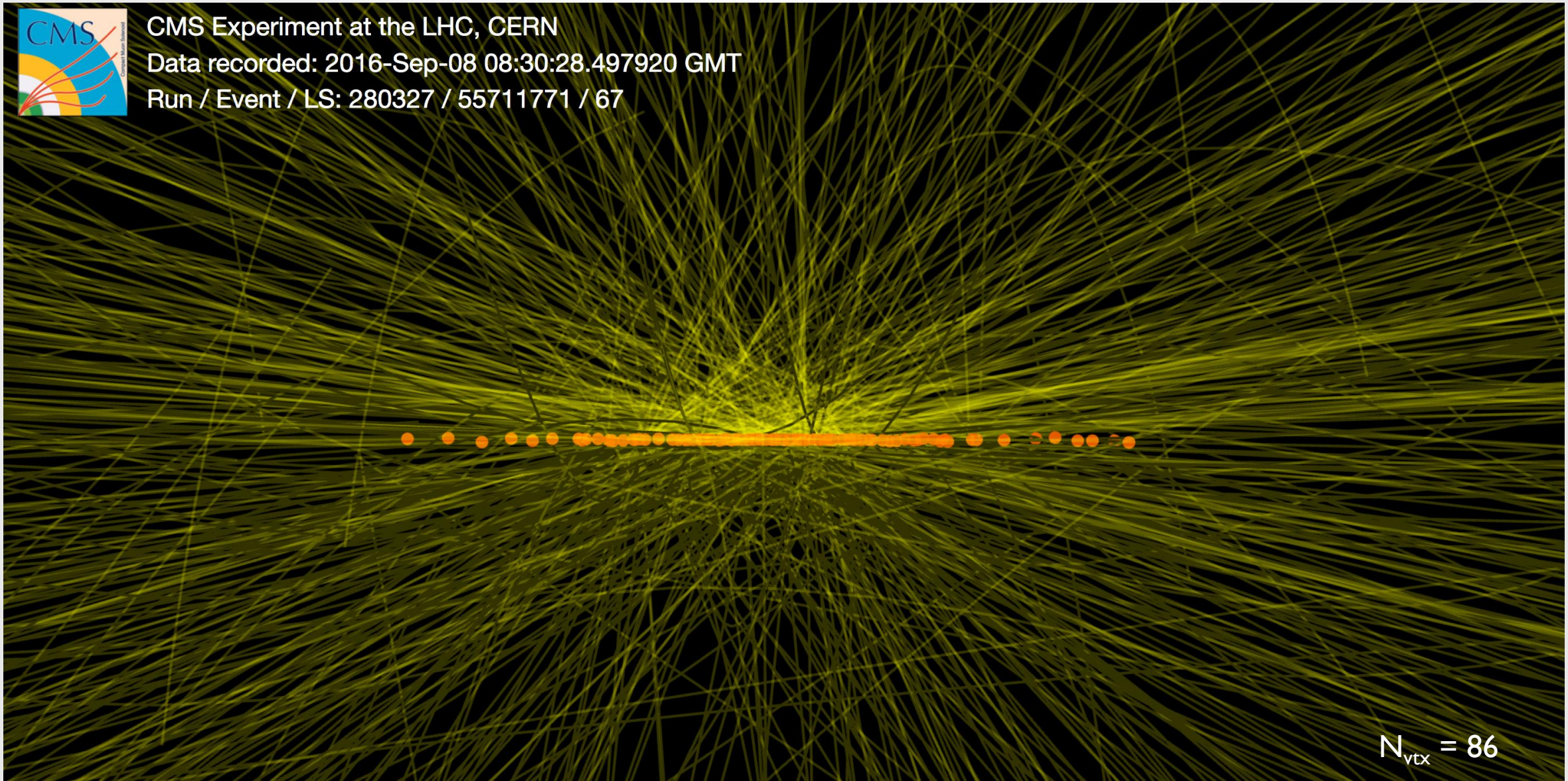


2.10 REMOVAL OF CHARGED PILE UP

- There are 10-40 pile-up (PU) collisions at each bunch crossing
  - Need to remove particles from PU collisions
- Vertices with charged particles are identified, and the most energetic one (leading vertex) is associated with a 'hard scatter'
- Remove all charged particles without association to the leading vertex
  - 'Charged hadron subtraction'
  - Neutral particles have no vertex association).
- Now proceed with jetclustering
  - neutral hadrons and photons from pile-up are a problem. How can it be tackled?



A HIGH PILE-UP EVENT



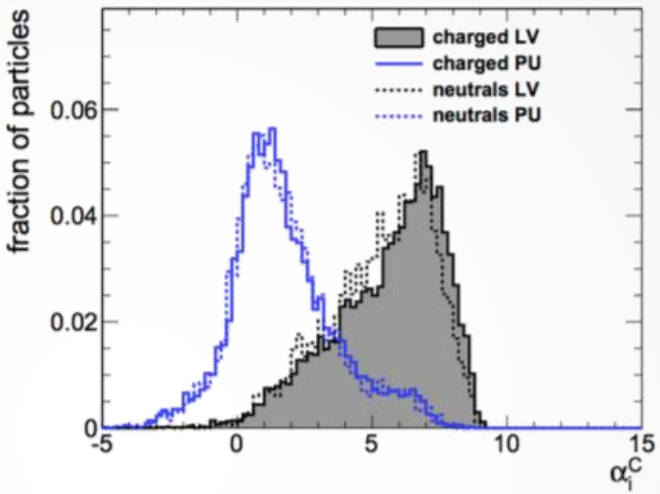
2.11 *GLOBAL PU MITIGATION

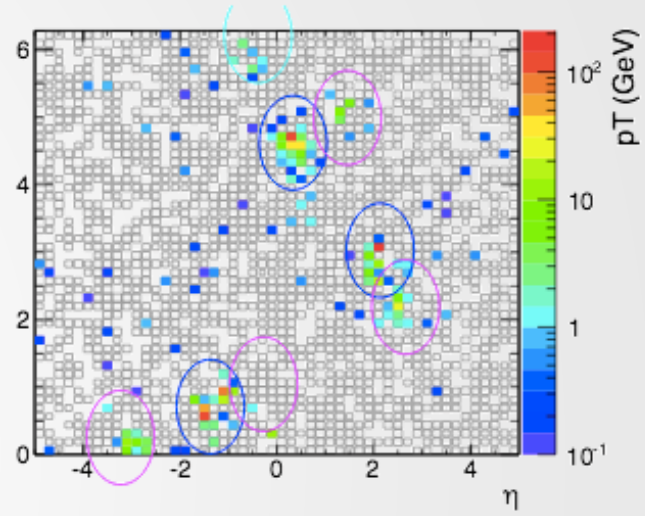
JHEP 1410 (2014) 59

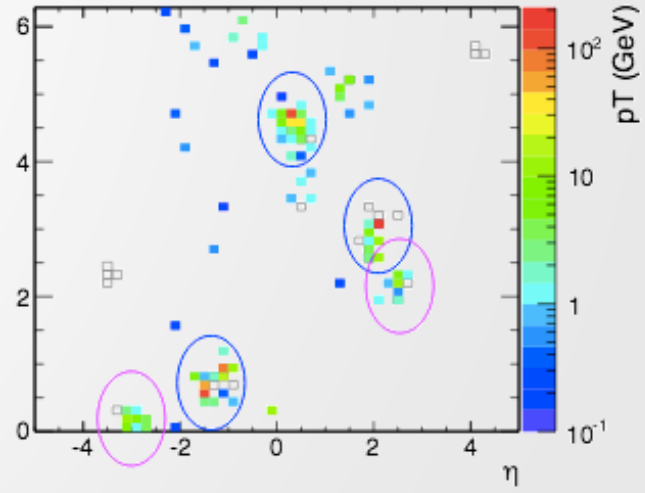
- Charged hadron subtraction (previous slide)
  - removes charged particles from PU vertices
  - does not work for neutral particles
- For neutrals, exploit that pile-up is random
- Algorithm: Pile-up per particle Id: 'Puppi'
  - Define

$$\alpha_i = \log \sum_{\substack{j \in Ch,\, L \\ j \neq i \quad V}} \left(\frac{p_{T,j}}{\Delta R_{ij}} \right)^2 \Theta(R_0 - \Delta R_{ij})$$

    which encodes the "non-PU-probability" of a particle.
  - distribution of α is *measured* using charged component in each event and *applied* to the neutral particles
  - reweight neutrals according to PU probability
  - effective in-event measure of the fluctuating PU



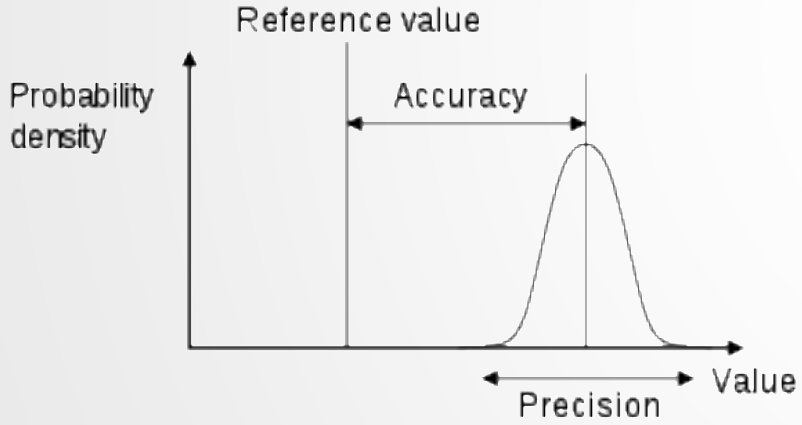




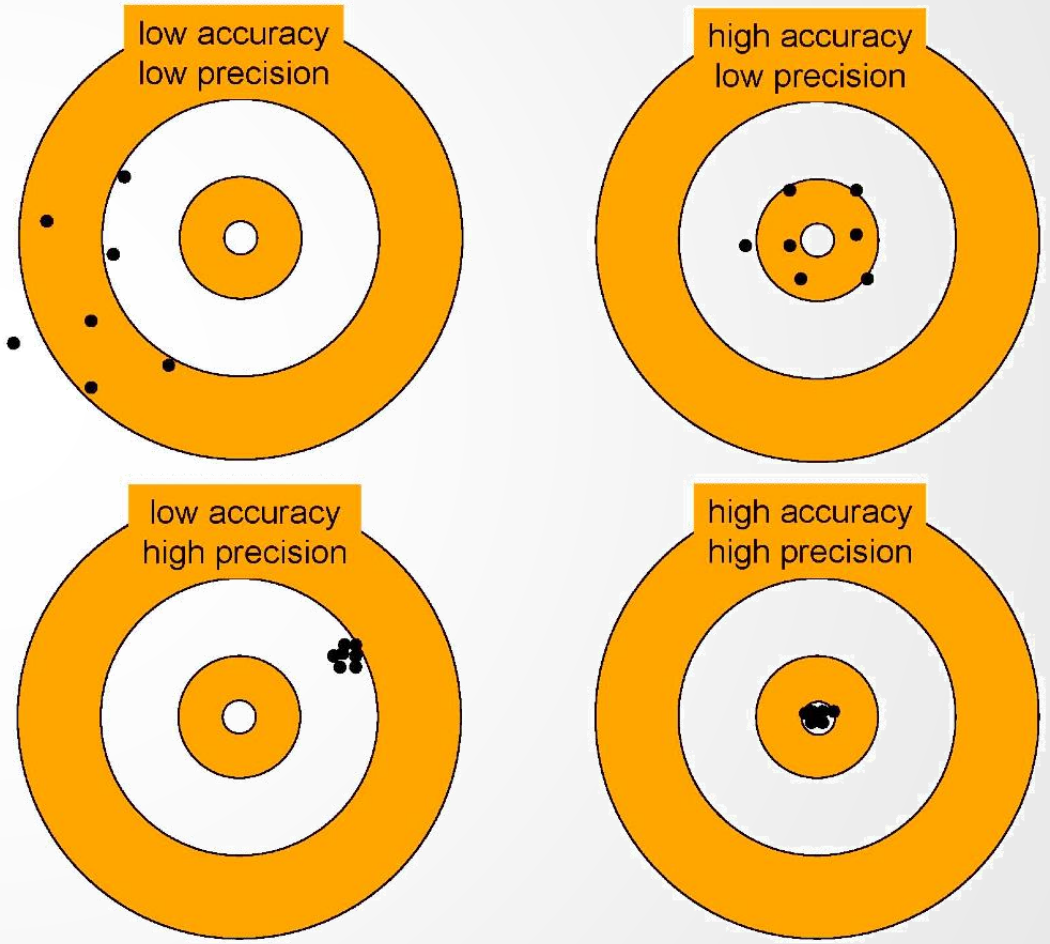
3. STATISTICAL AND SYSTEMATICAL UNCERTAINTIES

3.1 PRECISION AND ACCURACY

- Precision is a description of random errors, a measure of statistical variability.
- Accuracy describes systematic errors, that is, differences between the true and the measured value that are not probabilistic (or: bias).



- In particle physics, precision can be increased by accumulating more data
  - Equivalent to repeating the measurement



3.2 NORMAL DISTRIBUTION (GAUSSIAN)

- PDF: $g(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Mean: $\mathrm{E}(x) = \mu$
- Variance $V(x) = \sigma^2$
- "standard normal distribution" N(0,1): $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
- Cumulative distribution of N(0,1):

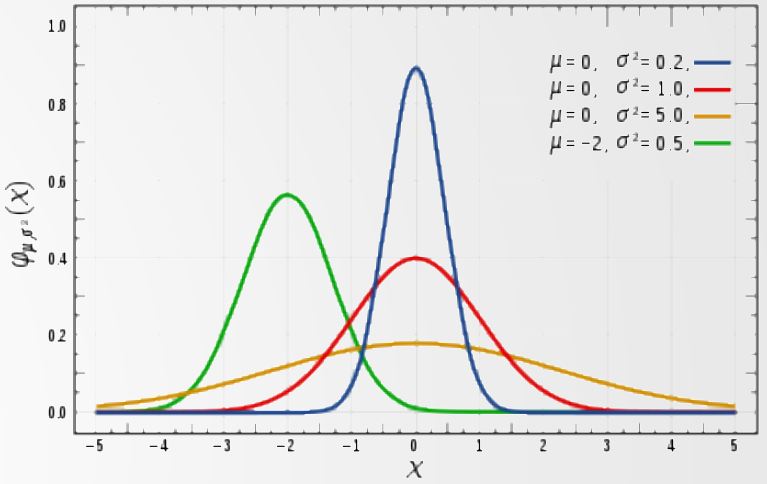
$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{d}z e^{-\frac{z^2}{2}} = \frac{1}{2}\left(\mathrm{erf}\left(\frac{x}{\sqrt{2}}\right) + 1\right)$$

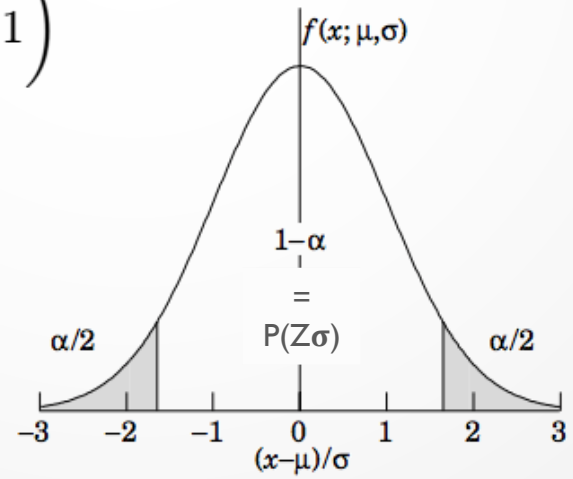
where 'erf' is the 'error' function.

$$P(Z\sigma) = \Phi(Z) - \Phi(-Z) = \mathrm{erf}\left(\frac{Z}{\sqrt{2}}\right)$$

- p-value: probability that a random process produces a measurement thus far, or further, from the true mean: $\alpha = 1\text{-}P(Z\sigma)$

Important property:
If x_1, x_2 follow Normal distr. with μ_1,σ_1, and μ_2,σ_2, then x_1+x_2 follows Normal distr. with $\mu = \mu_1 + \mu_2$ and $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$.





p-value	
α	δ
0.3173	1σ
4.55×10^{-2}	2σ
2.7×10^{-3}	3σ
6.3×10^{-5}	4σ
5.7×10^{-7}	5σ
2.0×10^{-9}	6σ

p-value	
α	δ
0.2	1.28σ
0.1	1.64σ
0.05	1.96σ
0.01	2.58σ
0.001	3.29σ
10^{-4}	3.89σ

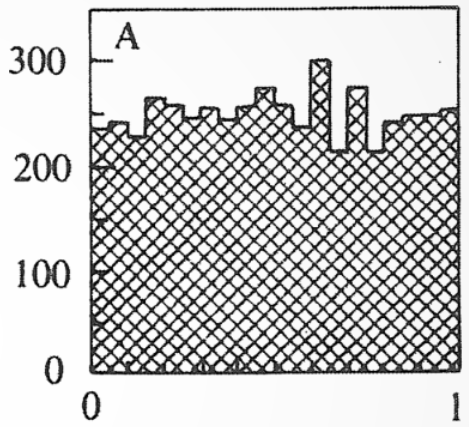
3.3 THE CENTRAL LIMIT THEOREM

- Central limit theorem:
  - When independent random variables are added, their properly normalized sum tends toward a normal distribution even if the original variables themselves are not normally distributed.
- More specifically: Consider n random variables with finite σ^2 and arbitrary pdf:

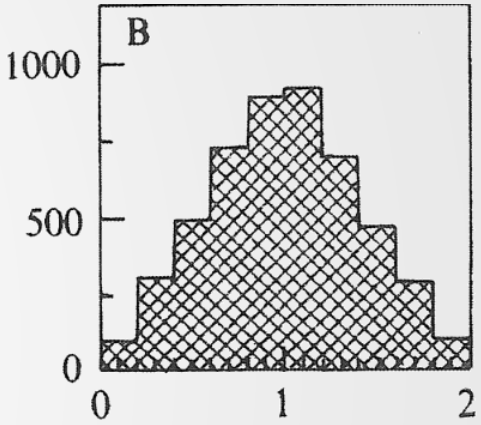
$$y = \sum_{i=1}^{n} x_i \qquad \overset{n\to\infty}{\longrightarrow} \qquad \mathrm{E}(y) = \sum_{i=1}^{n} \mu_i \,, \quad \mathrm{V}(y) = \sum_{i=1}^{n} \sigma_i^2$$

- Measurement uncertainties are often the sum of many independent contributions. The underlying pdf for a measurement can therefore be assumed to be a Gaussian.
- Many convenient features in addition, e.g., sum or difference of two Gaussian random variables is again a Gaussian.

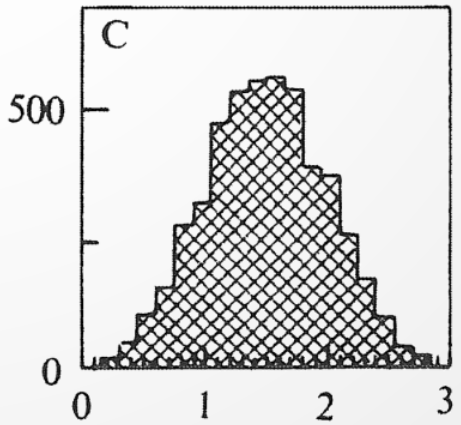
A: $x \sim U[0,1]$, $\mu=1/2$, $\sigma=1/12$
N=5000 events



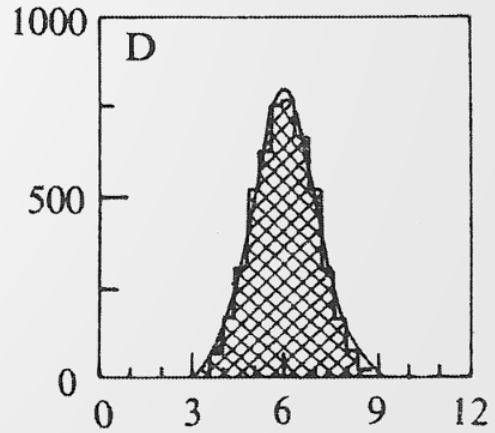
x_1+x_2 from A,
N=5000 events



$x_1+x_2+x_3$ from A,
N=5000 events



D: $x_1+x_{2+\ldots}+x_{12}$ from A,
N=5000 events, almost Gaussian



3.4 POISSON DISTRIBUTION

https://en.wikipedia.org/wiki/Poisson_distribution

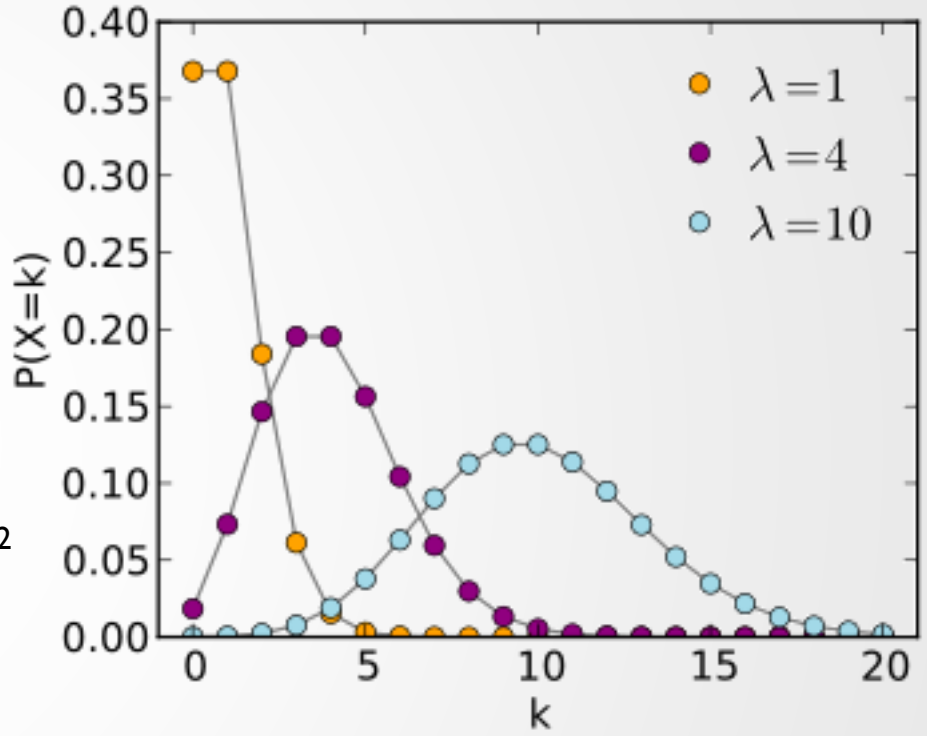
- Probability of a given number of events to occur in a fixed interval (of time or space) if these events occur with a known constant rate and independently of each other.

$$p(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$$\mathrm{E}(k) = \lambda$$

$$\mathrm{V}(k) = \lambda$$

Important property:
If k_1, k_2 follow Poisson distr. with λ_1, λ_2
$\rightarrow$ k_1+k_2 follows Poisson distribution $\lambda_1+\lambda_2$



- Can be approximated by a Gaussian for large λ
- Examples:
  - Clicks of a Geiger counter in a given time interval
  - Number of Prussian cavalrymen killed by horse-kicks
  - Number of particle interactions of a certain type produced in a given time interval or for a given integrated luminosity

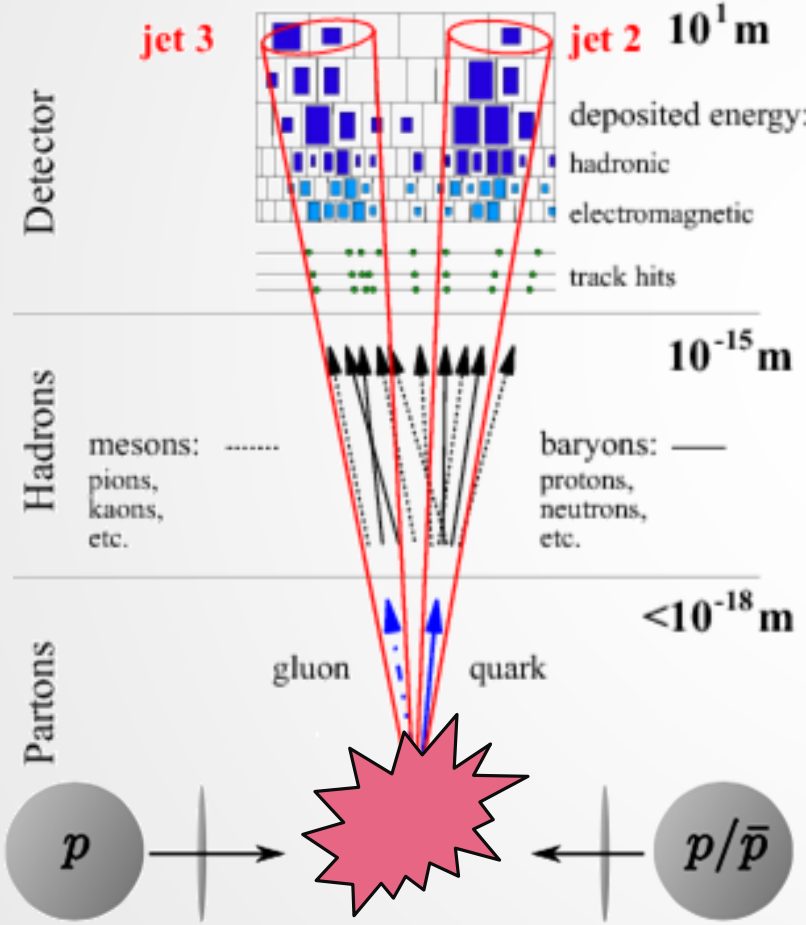
Number of deaths in 1 corps in 1 year	Actual number of such cases	Poisson prediction
0	109	108.7
1	65	66.3
2	22	20.2
3	3	4.1
4	1	0.6

3.5 STATISTICAL AND SYSTEMATIC UNCERTAINTIES

- Example: $x = 2.340 \pm 0.050$ (stat.) ± 0.025 (syst.)
- Statistical or random uncertainties
  - can be reliably estimated by repeating measurements
  - follow a known distribution (e.g.. Poisson or a Gaussian) that can be measured by repetition
  - Relative uncertainty reduces as $1/\sqrt{n}$ where n is the sample size
  - Main HEP use case: Expect λ events in a search region, and observe n. The measurement error on λ is $\sqrt{n}$.
- Systematic uncertainties
  - Cannot be calculated solely from 'sampling' fluctuations (=repeated measurements)
  - In most cases don't reduce as $1/\sqrt{n}$ (but often also become smaller with larger n because more data allows better auxiliary measurements)
  - Difficult to determine, in general less well known than the statistical uncertainty. (HEP: typically >90% of the work)
  - Systematic uncertainties ≠ mistakes (a bug in your computer code is not a systematic uncertainty)
- quoting stat. and syst. uncertainty separately gives an idea whether taking more data would be helpful

3.6 (SOME) TYPICAL HADRON COLLIDER SYSTEMATICS

- Some sources of systematic uncertainties:



physical evolution

work of the physicist

1. Low level / detector calibration
   - Response: fC ⟶ GeV conversion factors per detector cell, map of dead cells
   - timing of hit (this or the previous BX?)
   - hit quality (physical hit or detector noise?)
2. Reconstruction / jet & lepton calibration
   - Detector alignment (time & B field dependent)
     - Relative position of hits in tracker, calorimetry, and muon system ⟶ spectrometer measurements!
   - Electrons: Shower development, γ conversion, incorrectly classified charged pions
   - Jets: pile-up contribution, EM vs. hadronic energy component in shower, dead cells
3. Theoretical uncertainties in hadronization and the parton distribution functions

4 PRINCIPLES OF DATA ANALYSIS

4.1 ANALYSIS WORKFLOW OVERVIEW

1. simulation software covering all detector cells, the readout electronics, object reconstruction

2. preliminary simulation (10^9-10^{10} events) at ~4min/evt

3. definition of calibration workflow, preliminary analysis design

4. data taking, 'prompt' reconstruction

5. alignment and calibration of the real detector

6. Re-reconstruction of simulation and data (blind: avoid signal regions)

7. estimate systematic uncertainties based on data control regions

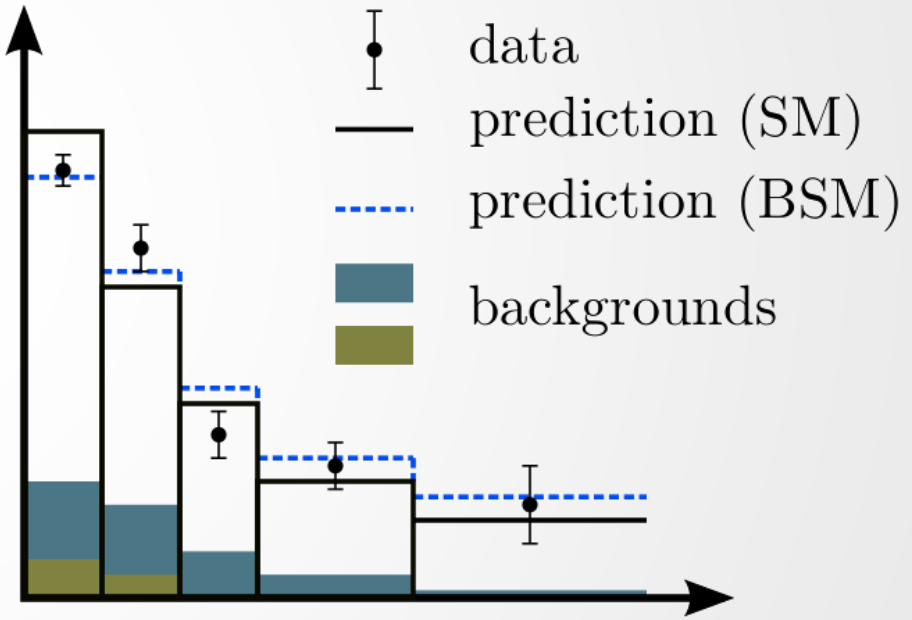
8. freeze & unblind (look at signal regions)

9. comparison of simulation and the data & data-driven background estimations

10. Limit setting / result / publication

4.2 COMPARING DATA WITH PREDICTION

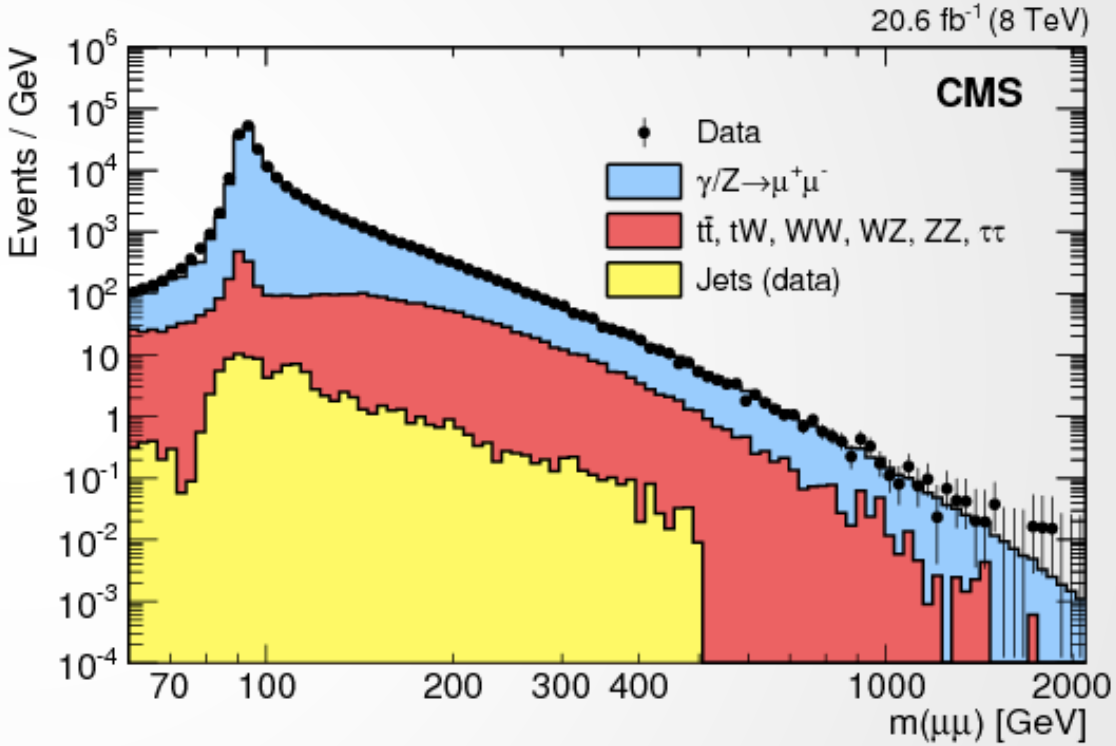
- BEFORE data is analyzed, a physics question must be formed
  - hypothesis test: Is the SM true, or rather a specific model beyond the SM (BSM)?
  - parameter measurement: What is the value of a certain parameter?
- Predictions can be obtained from
  - simulated events ("Monte Carlo" events)
  - other data regions (control regions) that have the same prediction based on a trusted fact
    - For example BR(W/Z⟶**μμ**) ≈ BR(W/Z⟶ee)
    - Often 'translation factors' are used to e.g. correct the efficiency differences between electrons and muons
  - theoretical calculations
- Very different systematic uncertainties!



- In a realistic analysis all variants occur, but often there is a central key idea related to the leading systematic.
- AFTER the analysis is frozen and the predictions were obtained, the data and prediction are compared

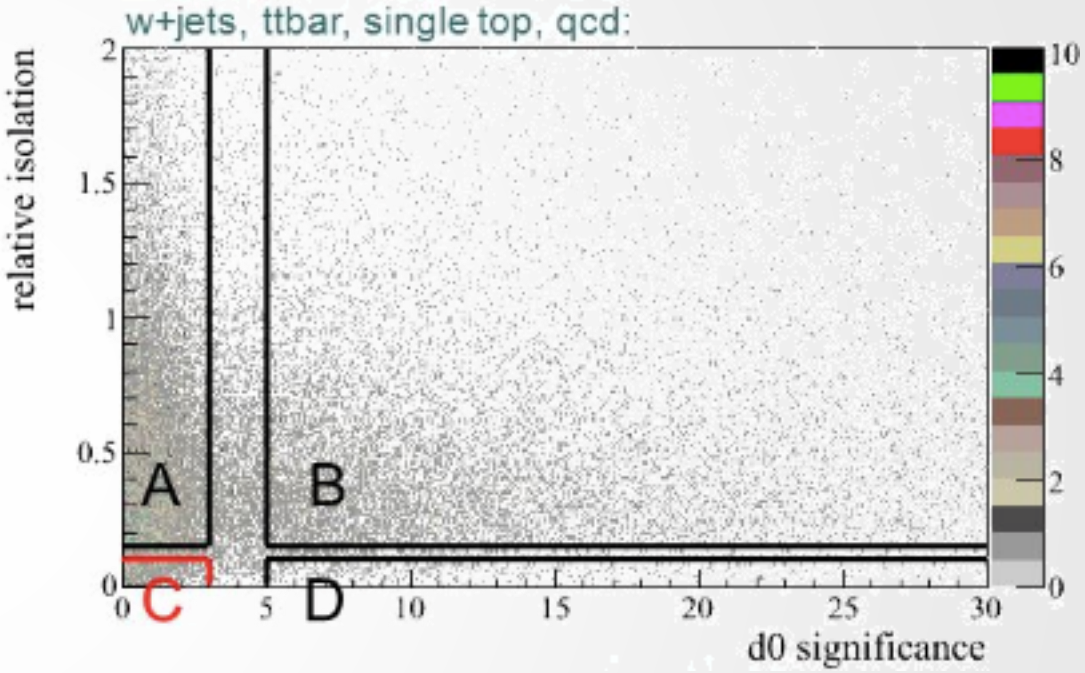
4.3 SIMULATION BASED ESTIMATION

- The analysis strategy should be designed such that systematic uncertainties are small
- Directly comparing data and simulation means
  - Every miscalibration, lack of theoretical understanding, directly affects the result
  - Systematic uncertainties in the background are then generally large
  - Need not be a problem if the background is small
- If the main source of a large uncertainty is known, the comparison can be reversed and, instead, a calibration can be obtained.
  - Here: The peak at m_Z = 91.2 GeV can be used to calibrate muon reconstruction
  - Once the muons are calibrated, we can use the same events to calibrate the recoiling jet!
  - However, loose m_Z as a measurement observable in the calibrated events (it was precisely measured at LEP)



4.4 DATA DRIVEN BKG ESTIMATION: ABCD METHOD

- If there are two uncorrelated variables, the background satisfies A/B = C/D
- This can be used to estimate $C_{est} = D \cdot A/B$ and this, if the signal is mostly in region C, the background can be predicted from A,B, and D
- Example: Multijet backgrounds to leptonic selections
  - Signal: Events with a single muon from W, top, etc.
  - Background: Muons from the hadronization of b-jets
    - There will be extra activity in the vicinity of the lepton ('relative isolation' variable)
    - The impact parameter of the lepton will populate high values in the background



- Almost no dependence on simulation – typically done at the early stages. Highly 'data driven'
- Extra uncertainties from the imperfect knowledge of the correlation (is it really negligible?)

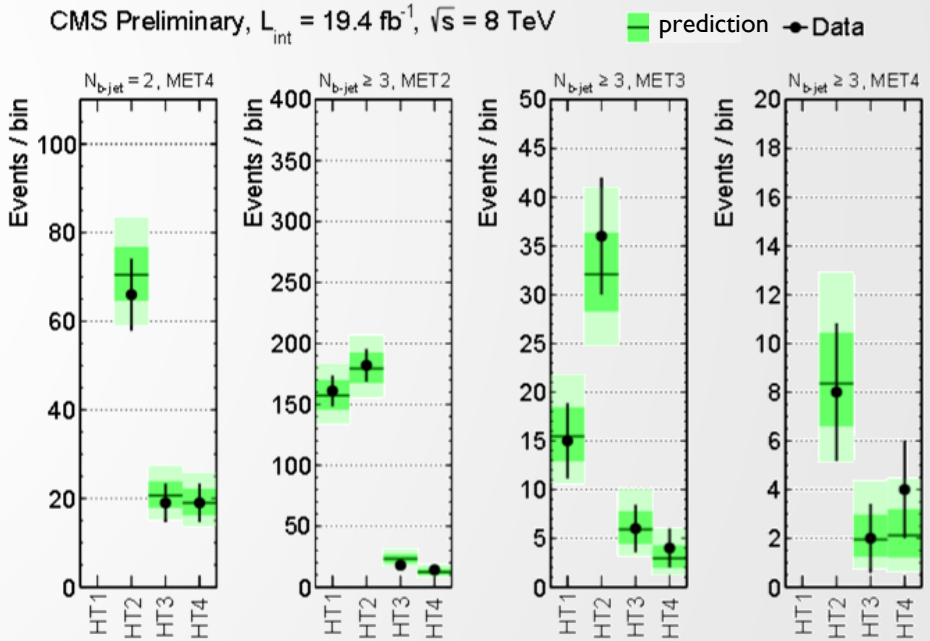
4.5 TRANSLATION FACTORS

- Sometimes, ratios of yields are more stable than yields because systematic uncertainties can cancel
- In this case, can take a ratio of background events between a 'signal' and a 'control' region from simulation (MC)

$$N_{\text{pred}}^{\text{signal}} = N_{\text{obs}}^{\text{control}} \times \underbrace{\frac{N_{\text{MC}}^{\text{signal}}}{N_{\text{MC}}^{\text{control}}}}_{r = \text{translation factor}}$$

- Can multiply with observed control region yield to arrive at background prediction
  - Similar to ABCD method, but less restricted because the translation factor can be measured in any suitably defined region (no rectangular cuts required)
  - Can be corrected for known differences
  - Systematic uncertainties on r can be estimated by comparing e.g. different simulations

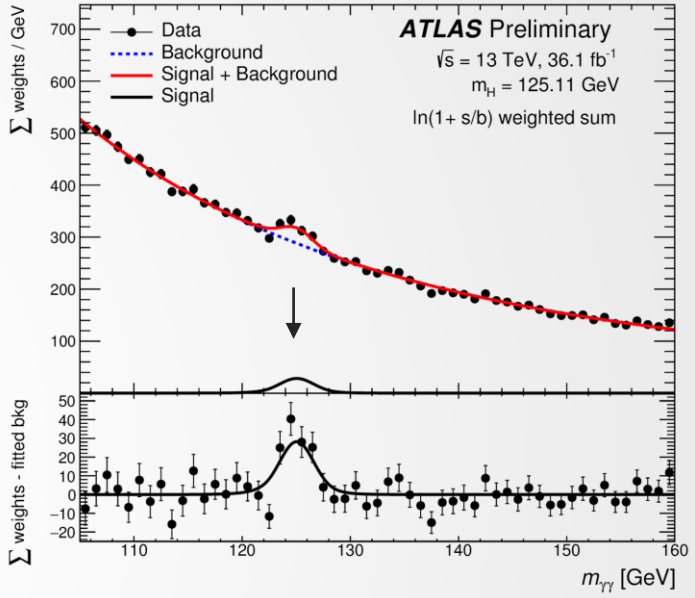
Example of data validation regions

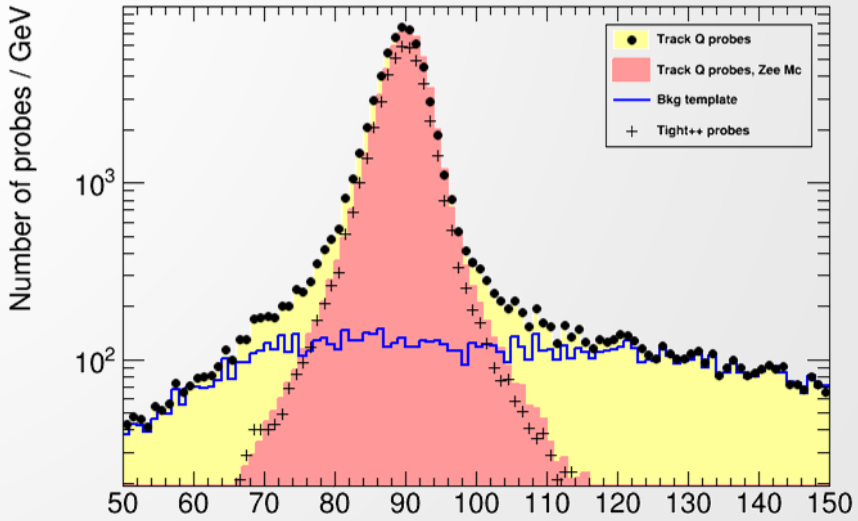


- Often, the applicability of a k-factor must be demonstrated in a signal-free validation region in real data

4.6 SIDEBAND SUBTRACTION

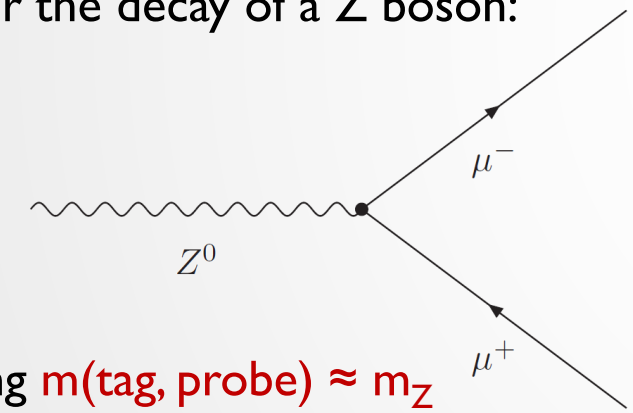
- If the shape of the background in the distribution is known
  - either from an assumed functional parametrization, or
    - Extra systematic: e.g. compare different parametrization
  - a simulated background template, or
    - extra systematic: e.g. compare different simulations
  - an independent measurement (in a control region)
    - extra systematic: simulated differences between control and signal region
- Sideband subtraction is used to 'subtract' the background contribution in a signal region
- Mostly, the background template is normalized in a background dominated region (i.e. high and low invariant mass) and then the normalized template is integrated in the signal region to obtain the prediction.





4.7 TAG AND PROBE

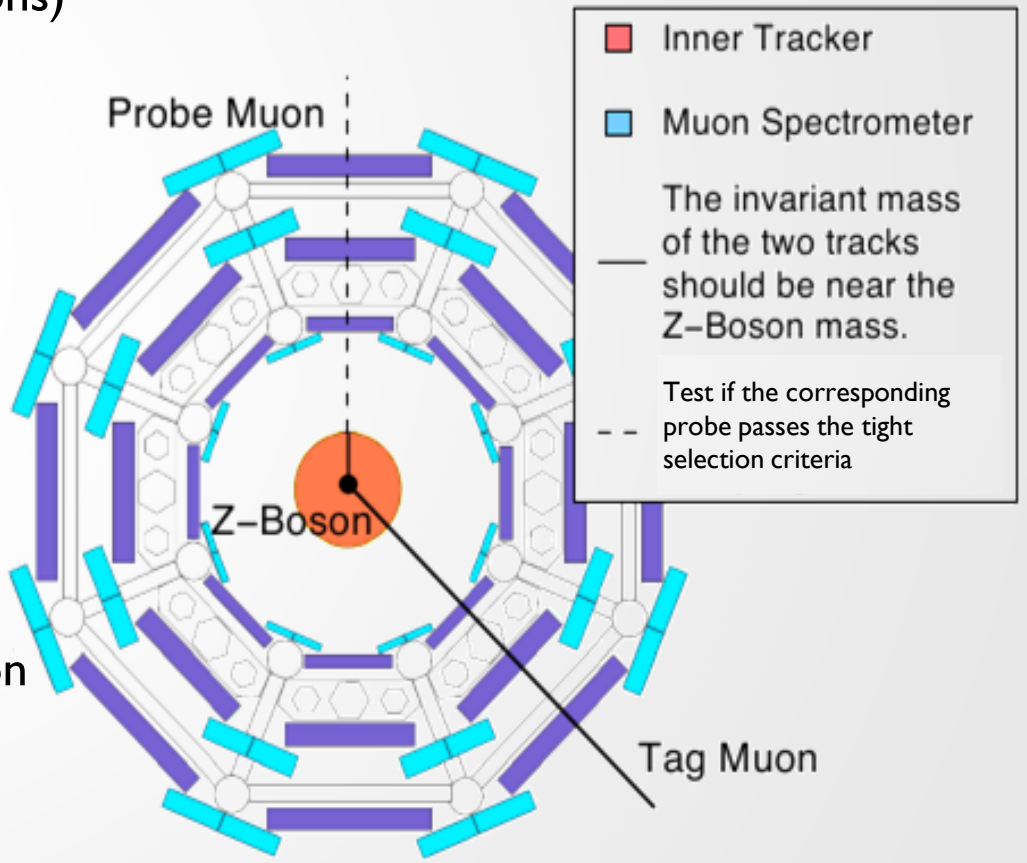
- Definition of efficiency: The probability that an existing object is reconstructed accordingly.
- Need to measure the efficiency of e.g. muon reconstruction in a sample of genuine muons, i.e. with negligible contribution from 'fake' muons (e.g. hadrons misidentified as muons)
  - Given a set of 'tight' selection criteria, what is the muon efficiency?
- How to ensure a pure selection of genuine muons?
  Consider the decay of a Z boson:



Tag muon with tight definition

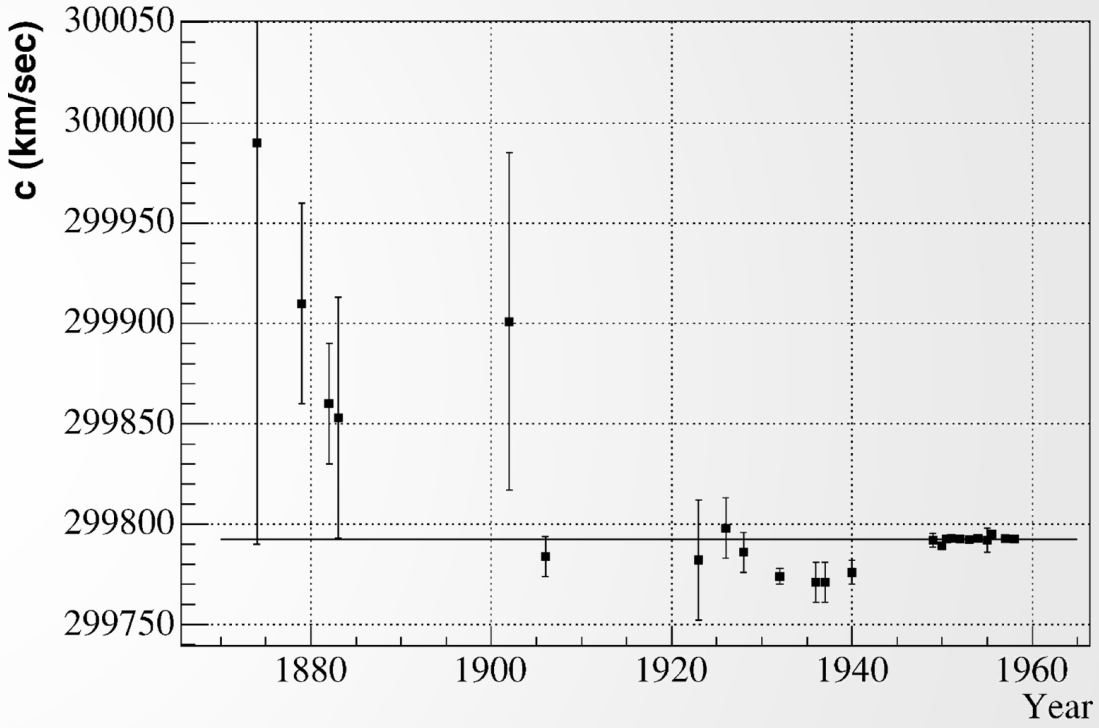
Probe muon with very loose definition (e.g. just a track)

- Requiring m(tag, probe) ≈ m_Z ensures that the probe muon is genuine despite its loose selection
- Can now measure the probability of a probe muon to pass the tight event selection threshold.



4.8 BLINDING

- History of the measurements of the speed of light vs. time and their uncertainties:
- The fluctuations are likely not from a random process
- It's conceivable previous measurements un(consciously) affect the outcome
- Issue: How to understand data well enough, without looking at signal?
  - Look at control regions!
- Strategies:
  - full blinding (W mass measurement in Z mass w)
  - partial blinding (allow every 5th event)
  - biasing of data with unknown value (not always possible)

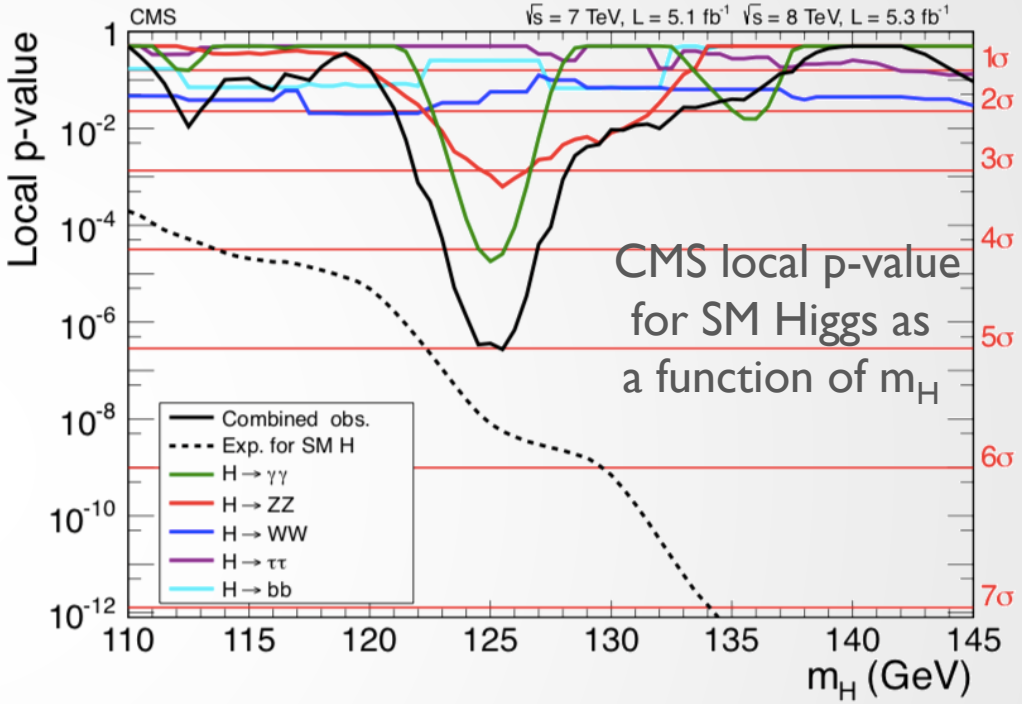


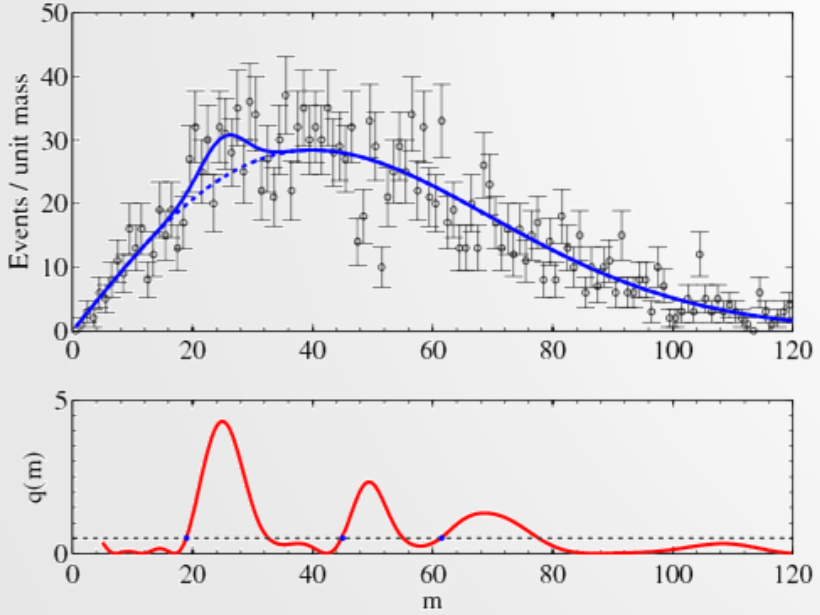
Klein JR, Roodman A. 2005.
Annu. Rev. Nucl. Part. Sci. 55:141–63

- Blinding is equally important in searches for rare (B)SM processes and in measurements.

4.9 LOOK ELSEWHERE EFFECT

- In a search for e.g. new effects, one needs to take into account how many "bins" are considered.
- For a 10^3 search bins (=trials), we expect to see on average one deviation that occurs with a probability of 10^{-3} or less (≈ 3 sigma)

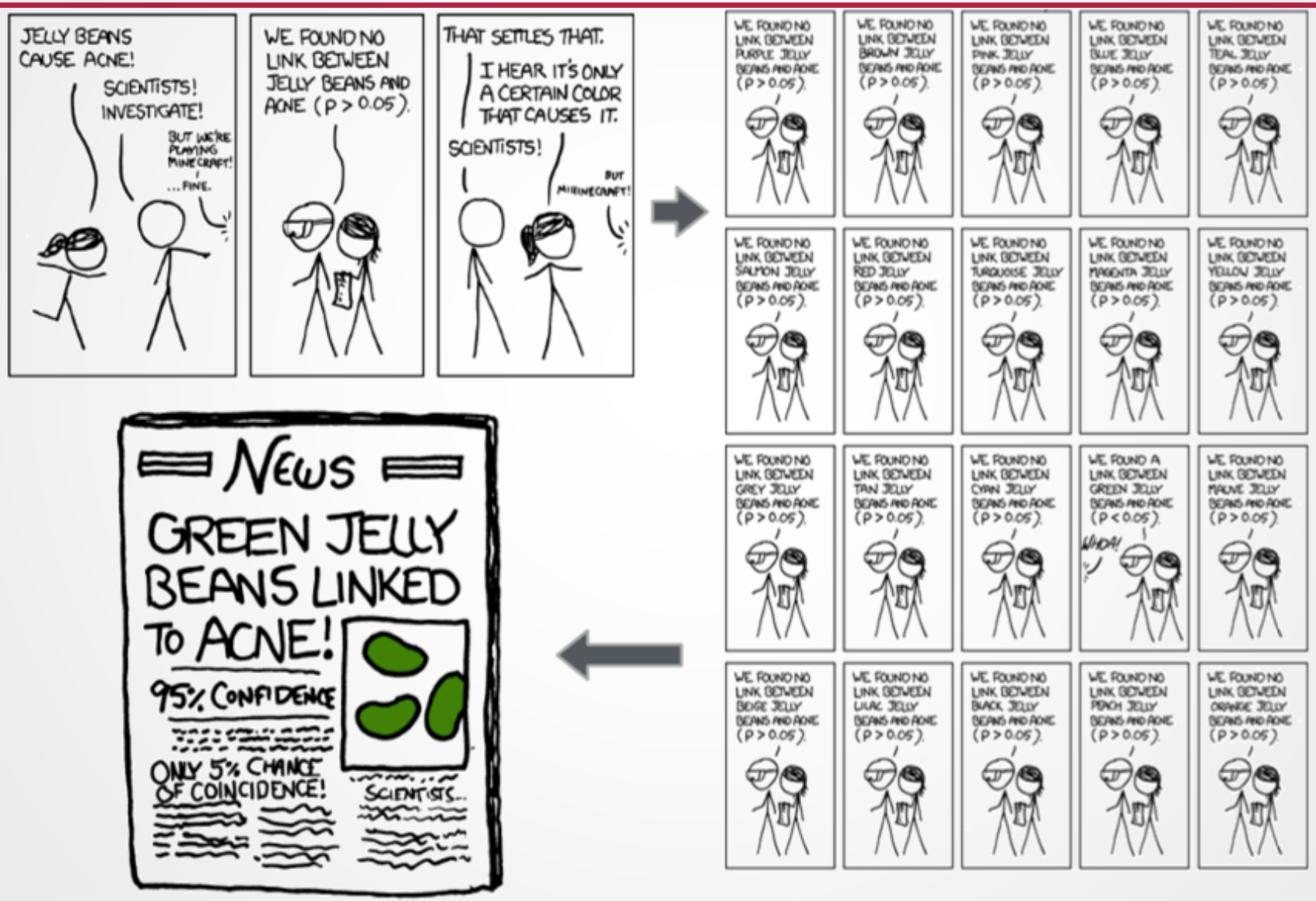




- If one is performing multiple tests then obviously a p-value of 1/n is likely to occur after n tests
- Solution: "trials penalty" or "trials factors", i.e. make threshold a function of n (more stringent threshold for larger n)

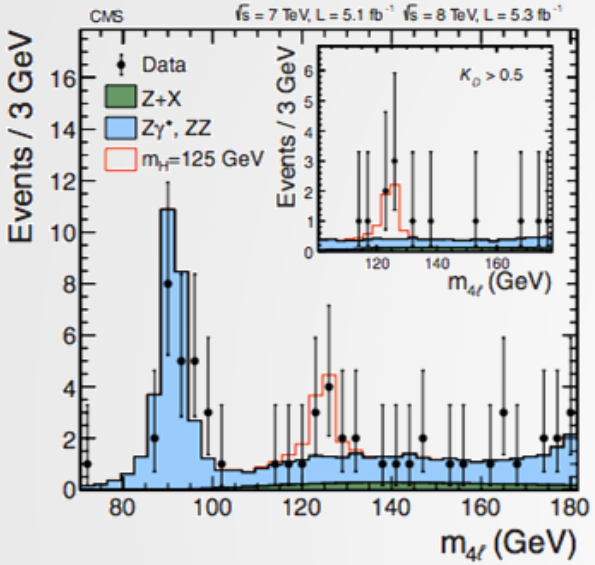
- Local p-value corresponds to 5σ
- Global p-value for mass range 110 –145 GeV corresponds to 4.5σ
- Problem: Need to estimate the number of trials; depends on detector resolution, signal assumptions

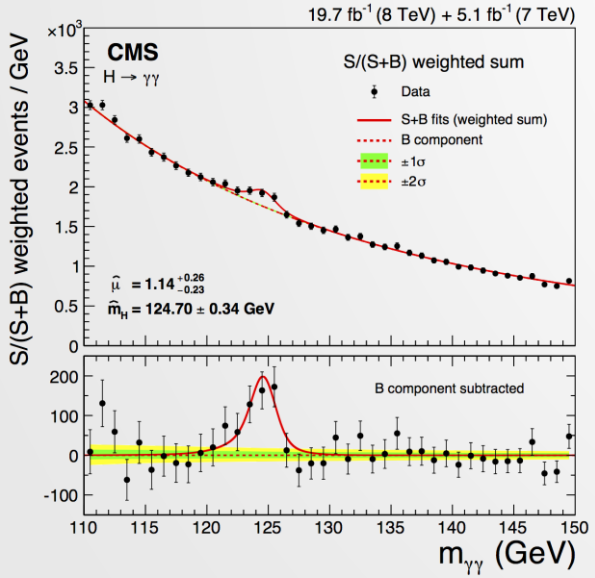
4.9 LOOK ELSEWHERE EFFECT / P-VALUE HACKING https://xkcd.com/882



WRAP-UP

WE'VE GONE THROUGH A LOT





1. Which experiment at which facility is the result from? Which dataset (energy, luminosity)? What is limiting in machine operations?
2. How is the proton described in the parton model and how does our knowledge on the PDFs interplay with the LHC predictions (particularly for the Higgs boson)?How are the particles produced in hadronisation, how do they decay and how are the decay products measured?
3. How do the meta-stable final-state particles interact with the detector?
4. What are the detector concepts and technologies? What are limiting factors for timing, energy, and momentum resolution? What properties can be measured, what level of particle identification is possible when combining the various sub-detectors?
5. How were the detectors assembled to experiments such that a measurement can be done? How are the events reconstructed and how, in principle, can backgrounds be estimated?