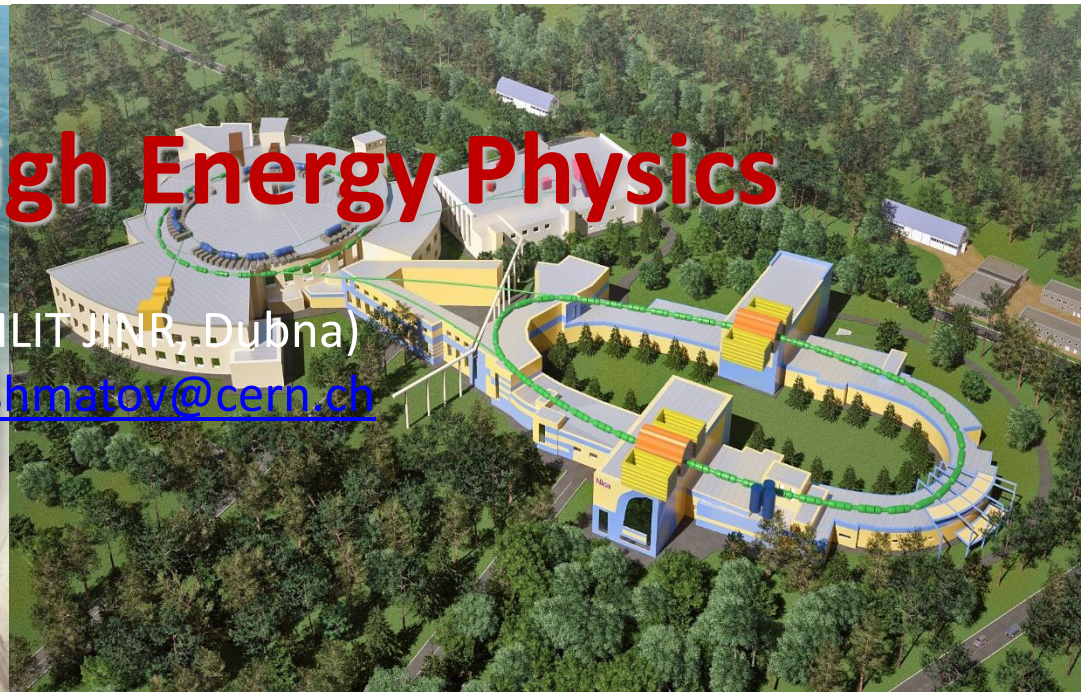




Computing in High Energy Physics

Sergei Shmatov (MLIT JINR, Dubna)

shmatov@jinr.ru shmatov@cern.ch



MLIT, JINR, Dubna
7 October, 2024




People have long asked,

- "What is the world made of?"
- "What holds it together?"

Some eternal questions

Physicists hope to fill in their answers to these questions through the analysis of data from High Energy Physics experiments

Particle physics have focused on the inner space frontier, pursuing the questions of the construction of matter and the fundamental forces at the smallest scale accessible.





Wavelength of probe radiation should be smaller than the object to be resolved

$$\lambda \ll \frac{h}{p} = \frac{hc}{E}$$

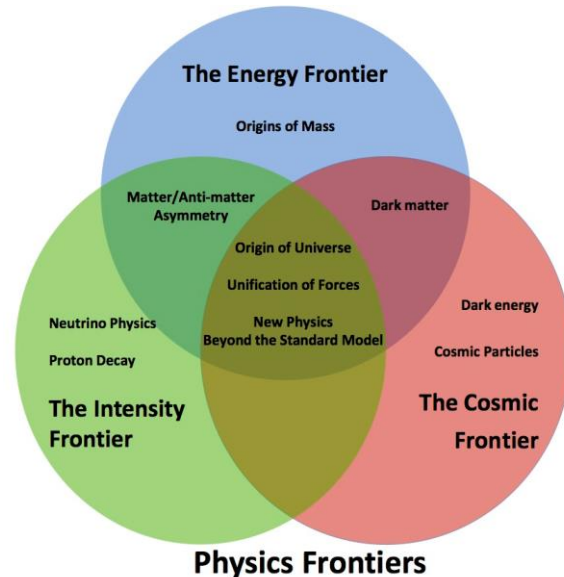
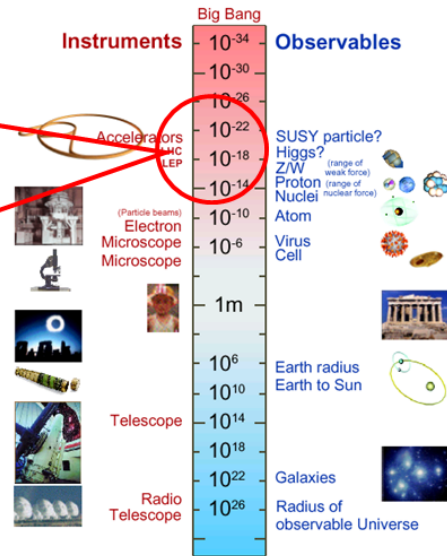
Object	Size	Energy of Radiation
Atom	10^{-10} m	0.00001 GeV (electrons)
Nucleus	10^{-14} m	0.01 GeV (alphas)
Nucleon	10^{-16} m	0.1 GeV (electrons)
Quarks	?	> 1 GeV (electrons)

Radioactive sources give energies in the range of MeV

Need accelerators for higher energies.

"electronic eyes"



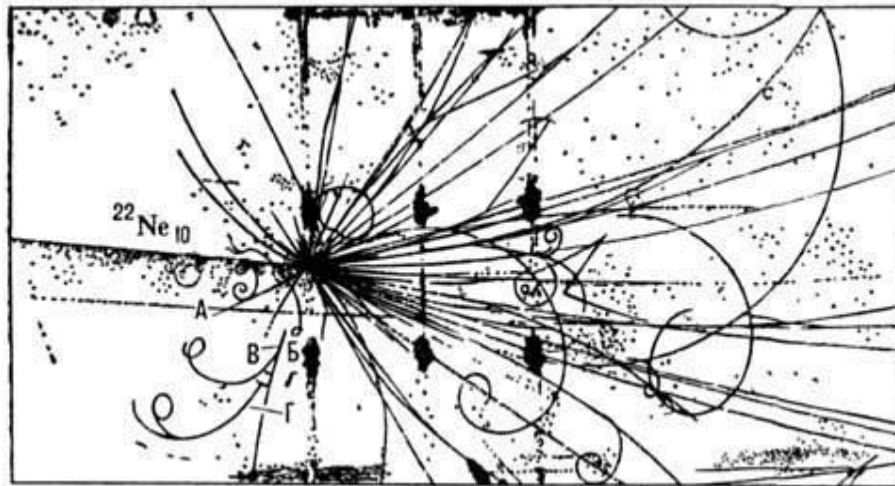
Computing is any goal-oriented activity requiring, benefiting from, or creating computing machinery (c) Wiki

- general-purposed devices (computers/laptops/mobile..) and software
- dedicated tools

Bubble chamber, Synchrophasotron (JINR),

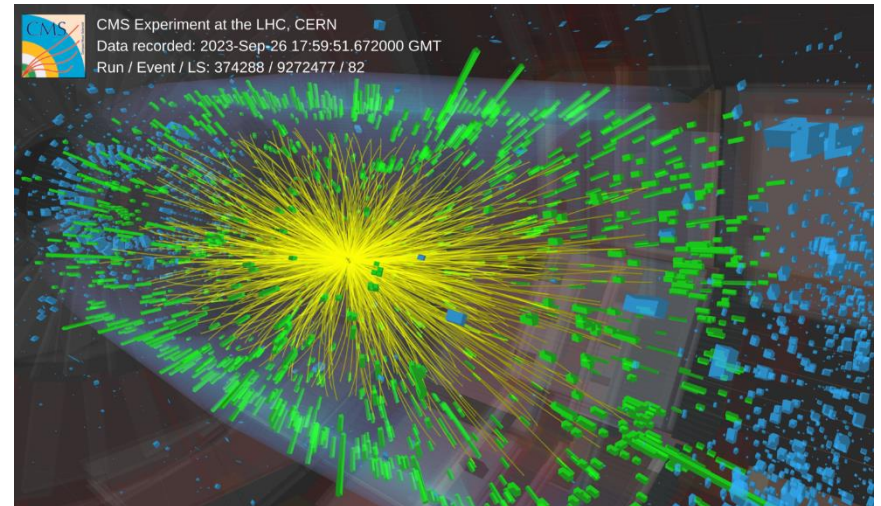
^{22}Ne ($p=92,4 \text{ ГэВ/c}$) + Ta

- 50 particles
- only one photo per second



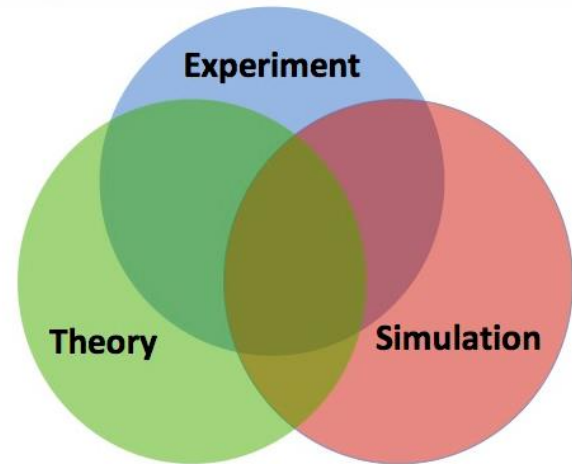
CMS @ LHC, PbPb (5,5 TeV), RUN3

- thousands particles
- 140 million electronic channels
- 3-dimensional “camera” able to shot 40 million “pictures” per second



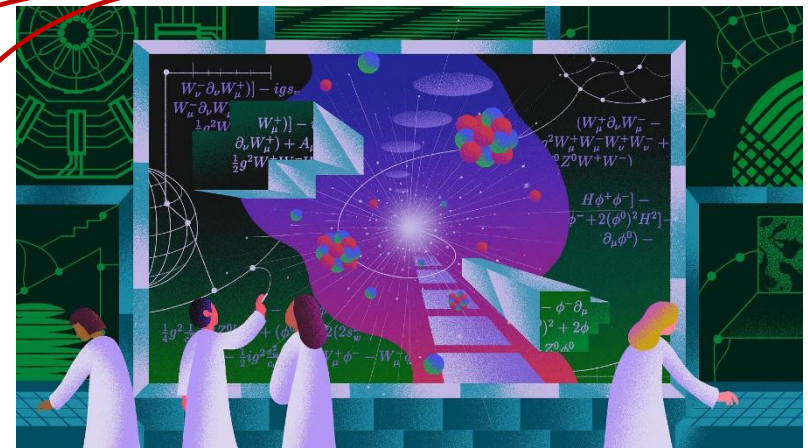
Particle physics or high energy physics is the study of fundamental particles and forces that constitute matter (c) Wiki

- Where can I get elementary particles?
 - ✓ in Nature (cosmic sources, earth sources, i.e. natural radioactivity)
 - ✓ man-made sources (reactors, accelerators)
- How can you catch particles \Rightarrow detector facilities
- What is needed for data processing?
 - ✓ algorithms and software for reconstruction of physics objects and processes
- What is needed for data analysis?
 - ✓ Theory
 - ✓ Monte Carlo Tools
 - ✓ Statistics Tools

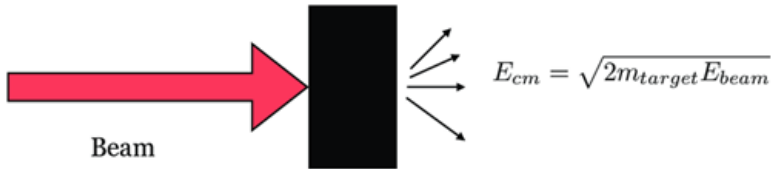


Along Three Paths

Information Technologies

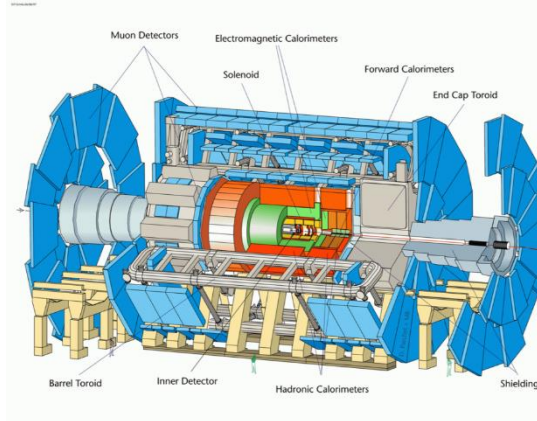


$$E_{cm} = \sqrt{4E_1 E_2}$$

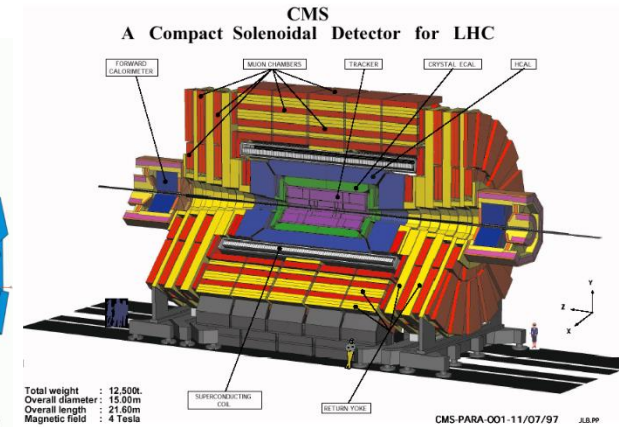


$$E_{cm} = \sqrt{2m_{target} E_{beam}}$$

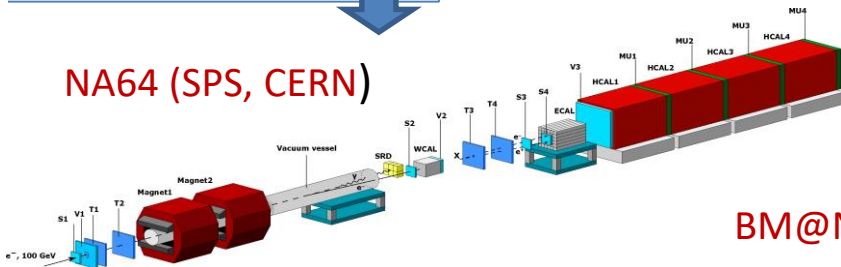
ATLAS (LHC, CERN)



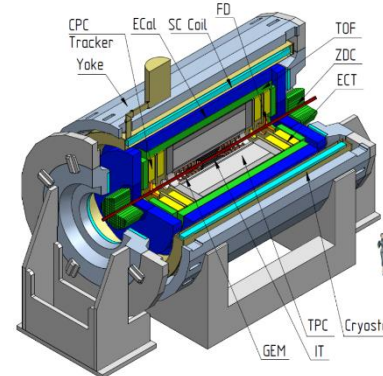
CMS (LHC, CERN)



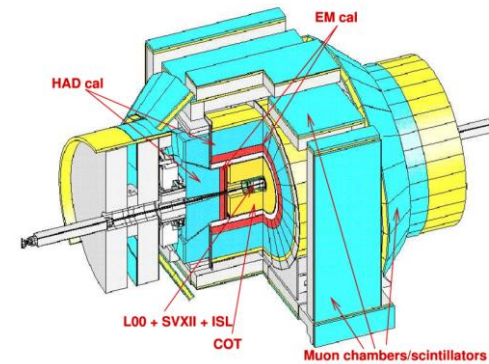
NA64 (SPS, CERN)



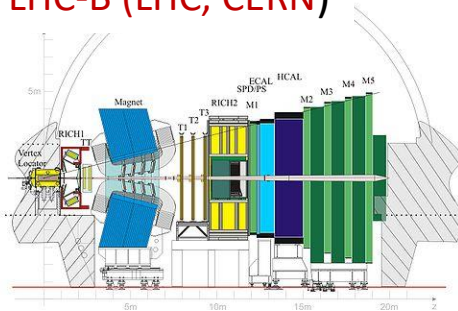
MPD (NICA, JINR)



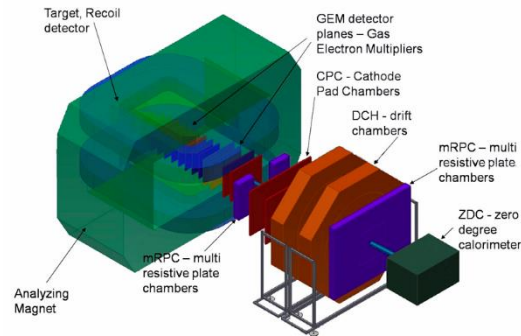
CDF (Tevatron, FLab)



LHC-B (LHC, CERN)

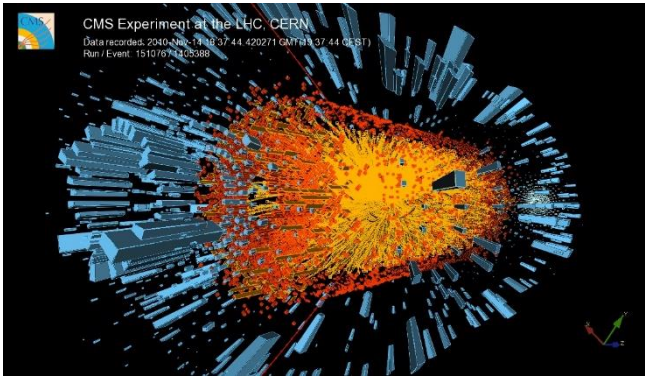


BM@N (NICA, JINR)

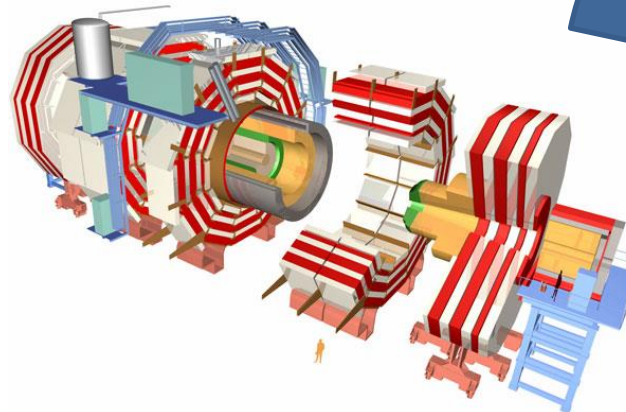


Scale without preservation of proportions

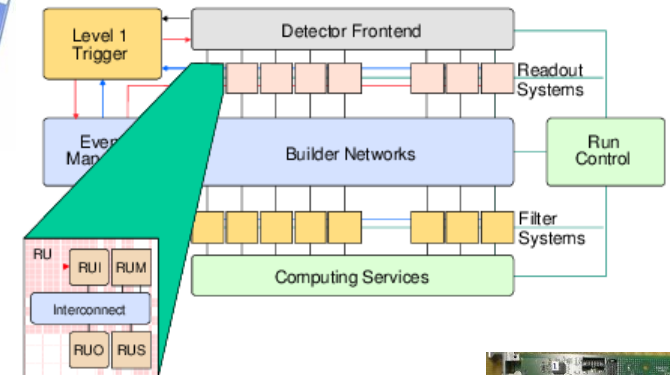
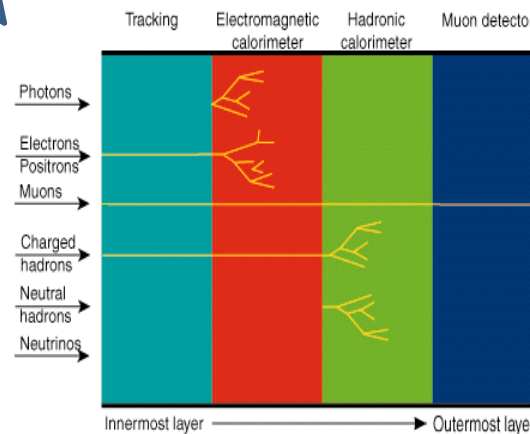
LHC Collisions



Interactions with
detector elements



Readout of detector electronics



What do physicist want to see?

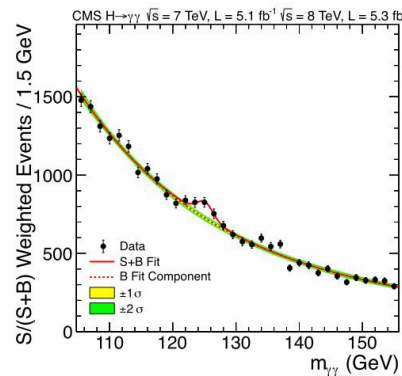
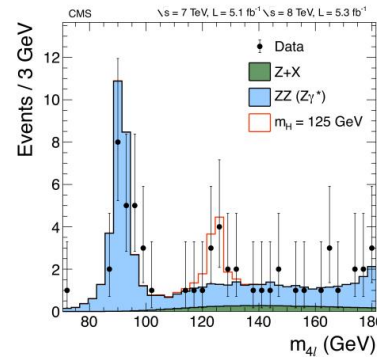
Higgs Boson



From design



to discovery



4 July 2012

Higgs announcement at CERN

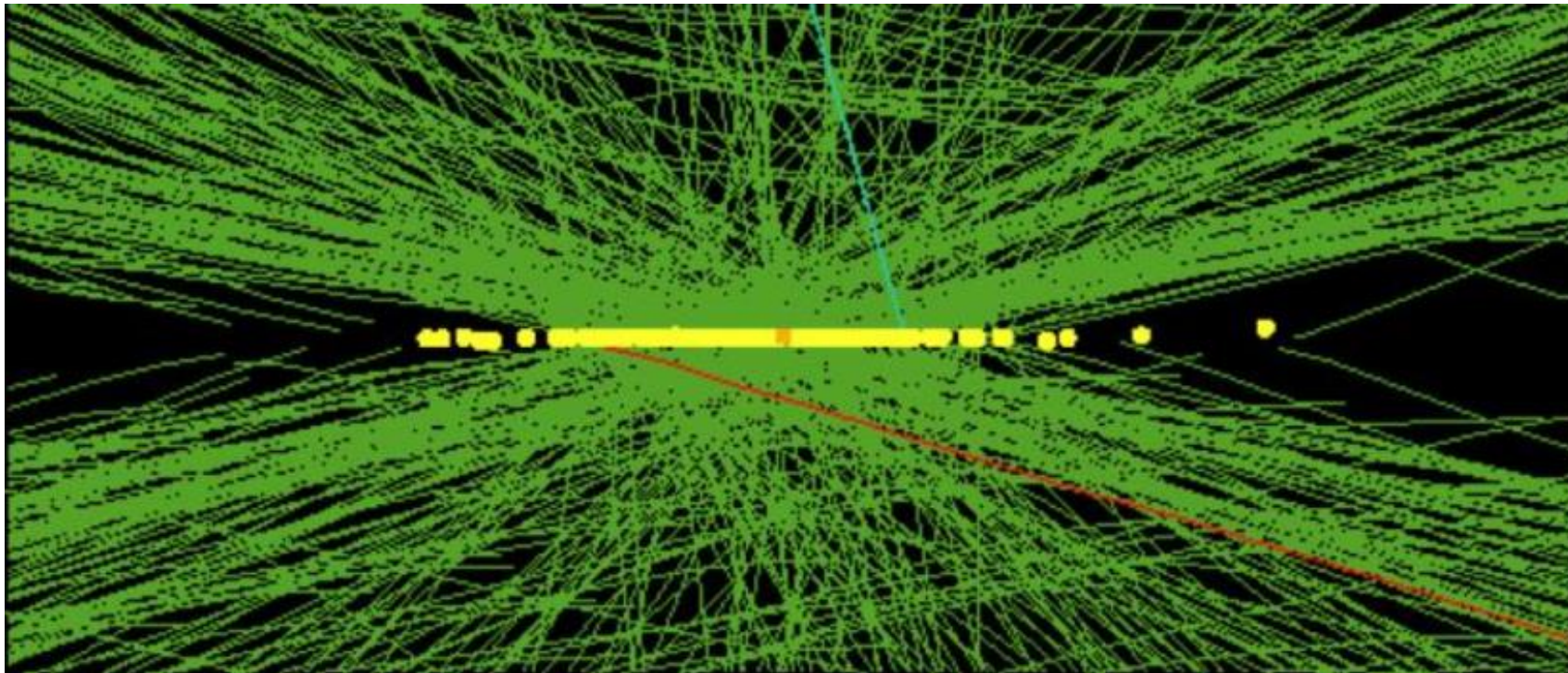


	Int. Luminosity at 7, 8 TeV	m_H [GeV]	Expected [st. dev.]	Observed [st. dev.]
ATLAS	10.7 fb ⁻¹	126.0 ± 0.6	4.6	5.0
CMS	10.4 fb ⁻¹	125.3 ± 0.6	5.9	4.9

What do they actually see?

Real CMS Event with High Pile-up

High pileup event with 78 reconstructed vertices taken in 2012 by CMS

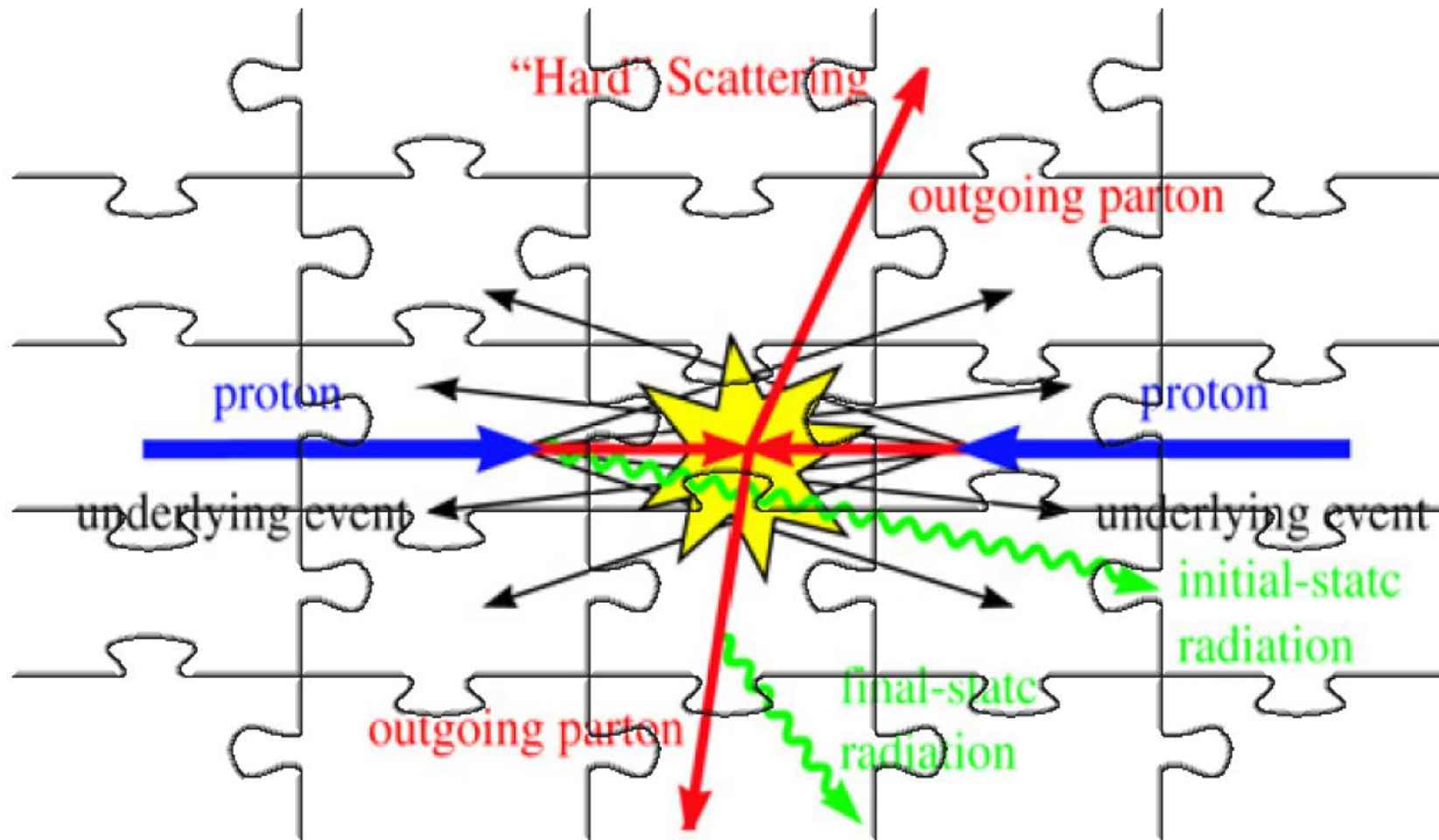


What is happening and and what we can do about it...

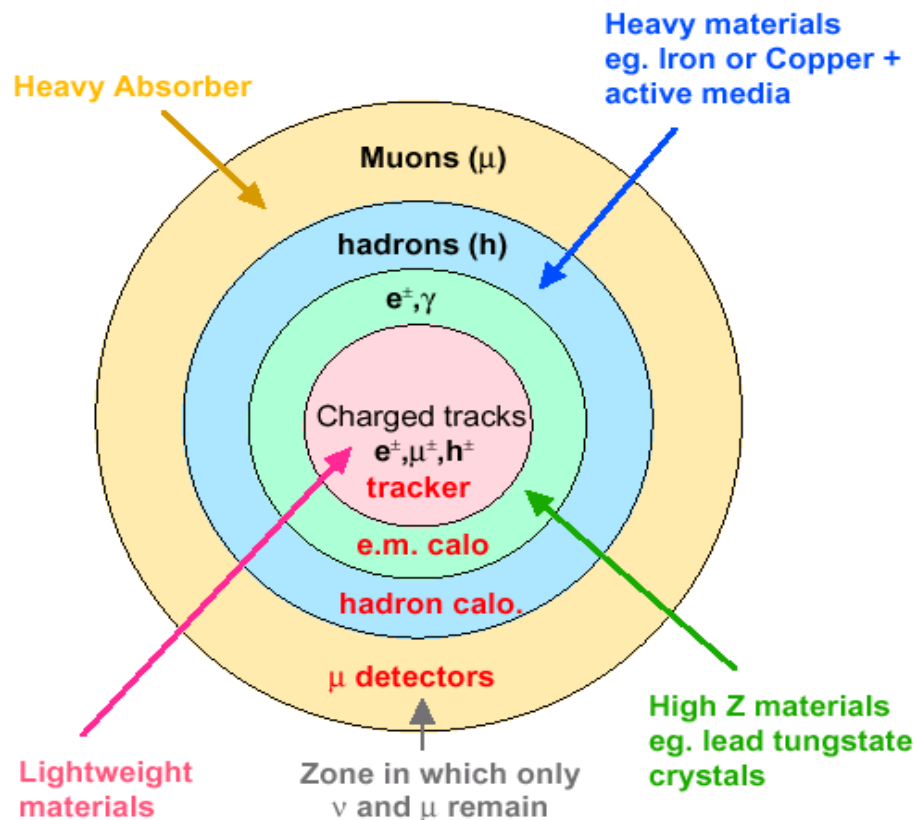
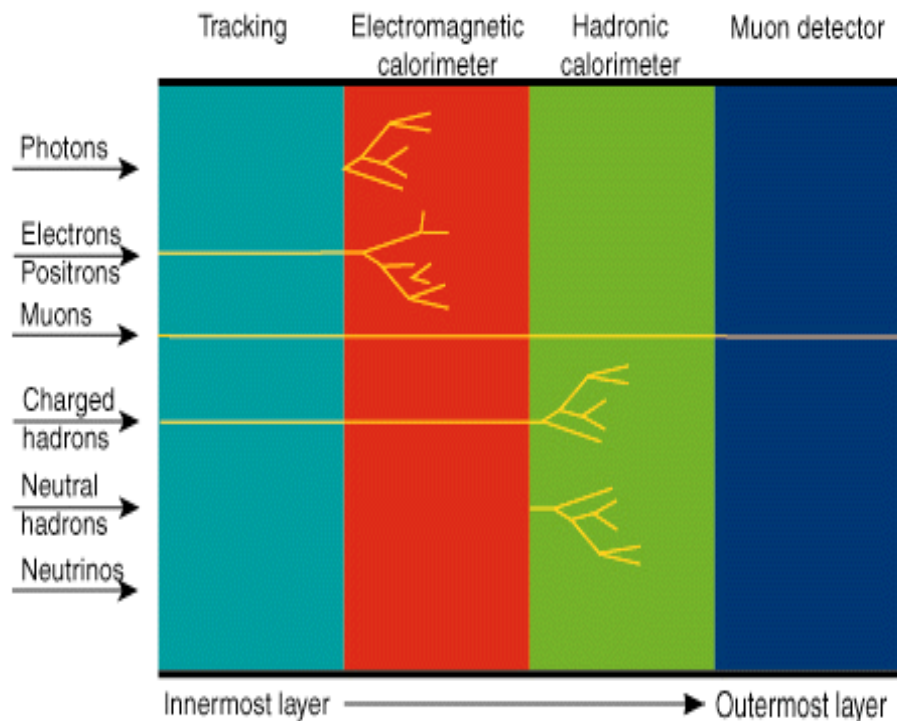
- Physics objects
- Event Selection
- Reconstruction and Processing
- Data Analysis



Data Analytics
[ˈdā-tə ə-nə-ˈli-tiks]
The science of analyzing raw data to make conclusions about that information.



Onion structure of detector layers placed in B-field

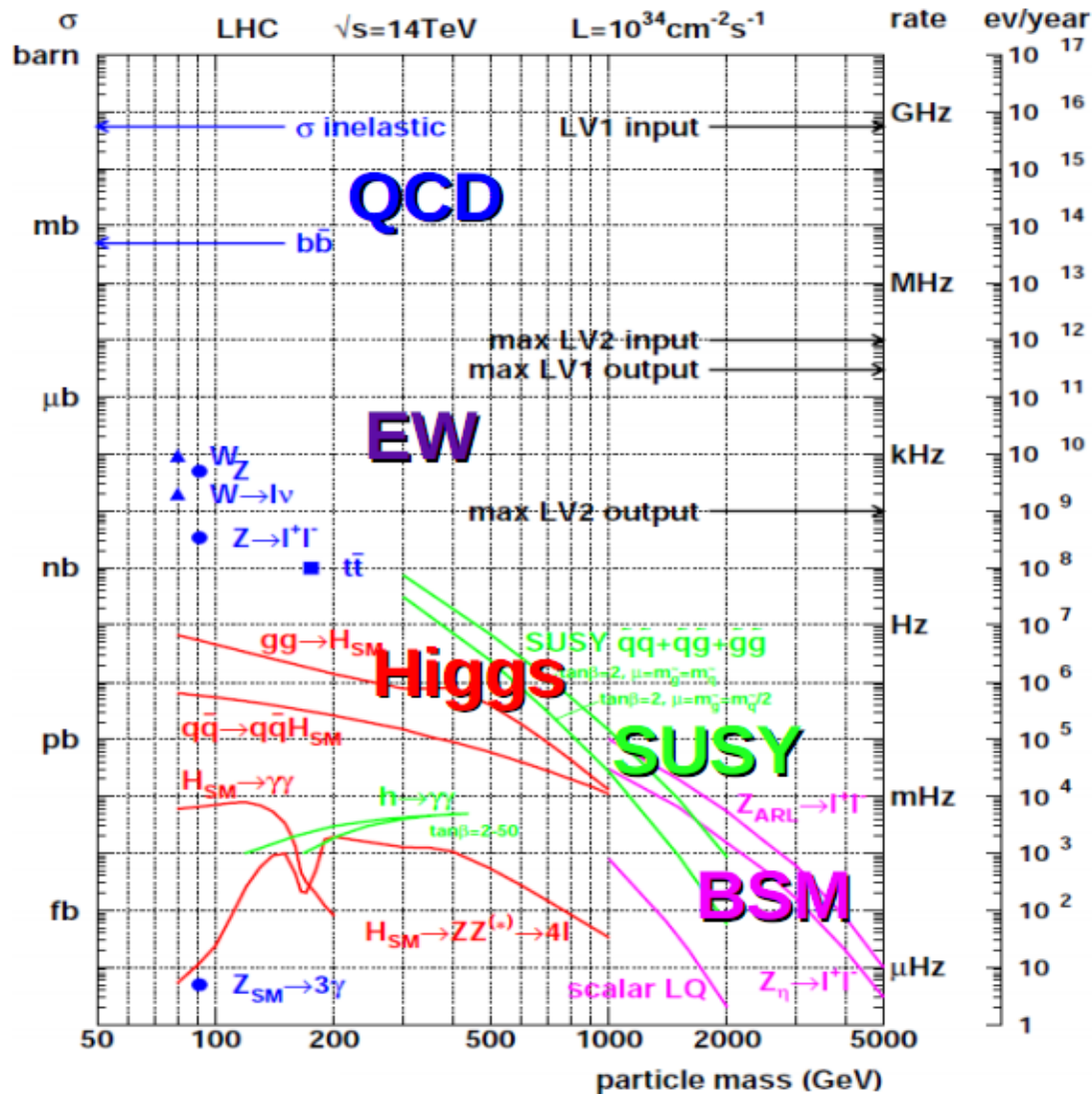


Each layer identifies and measures (or remeasures) the energy of particles unmeasured by the previous layer

No single detector can determine identity and measure energies/momenta of all particles

Event Selection and Data Flow





■ SM processes:

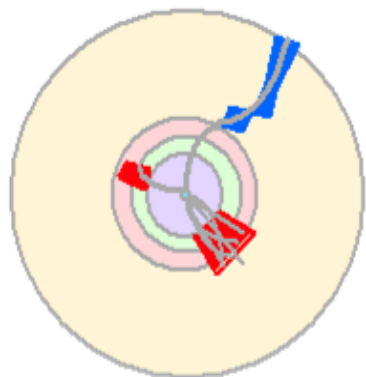
$$\sigma \sim 1/(100 \text{ MeV})^2$$



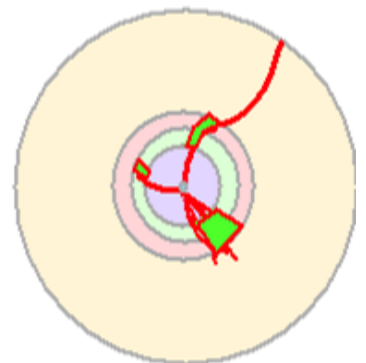
$10^{-8} !$

■ New Physics:

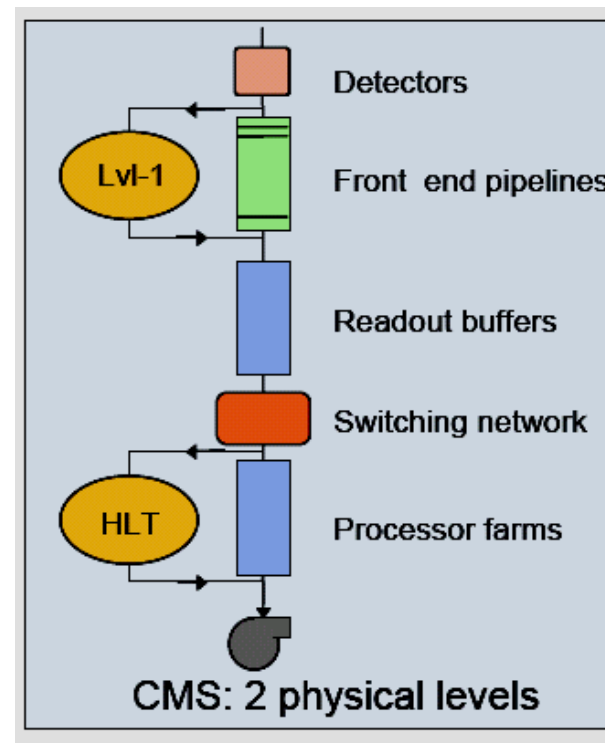
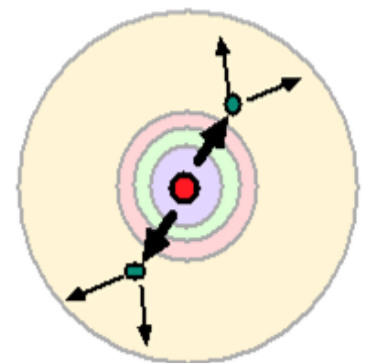
$$\sigma \sim 1/(1 \text{ TeV})^2$$

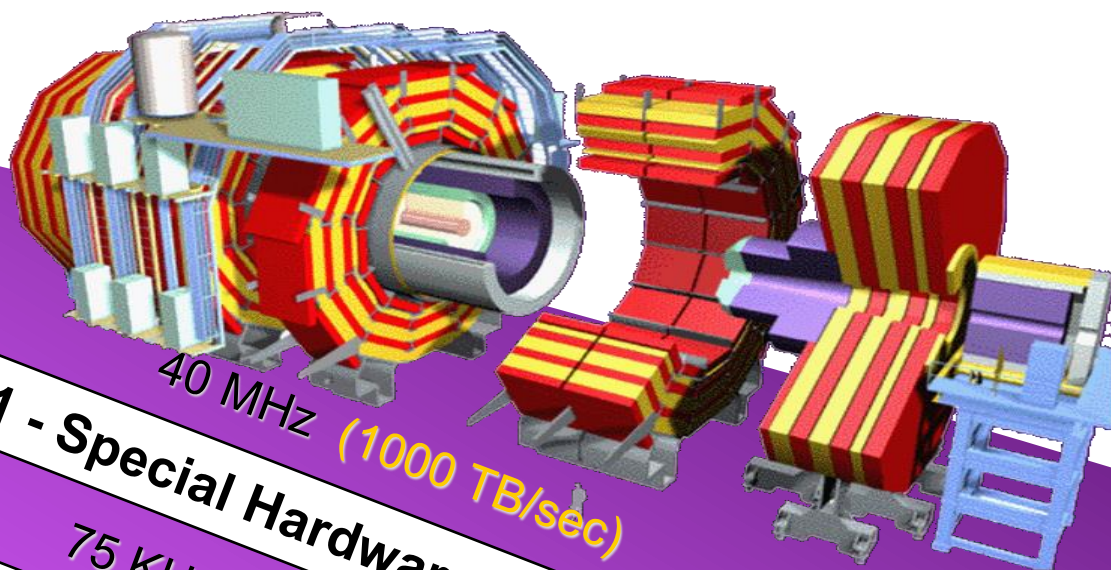


- Level-1:
Hardware selection is comprised of custom electronics that process data from detectors, rough cutoffs



- High Level Trigger:
Software selection based on reconstruction of physics objects, event topology



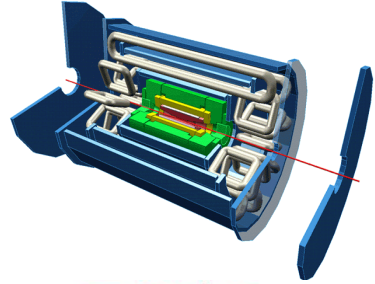


Level 1 - Special Hardware
40 MHz (1000 TB/sec)

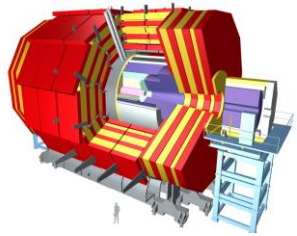
Level 2 - Embedded Processors
75 KHz (75 GB/sec)

Level 3 - Farm of commodity CPUs
5 KHz (5 GB/sec)

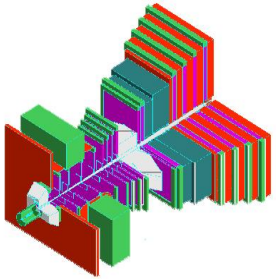
100 Hz (100 MB/sec)
Data Recording & Offline Analysis



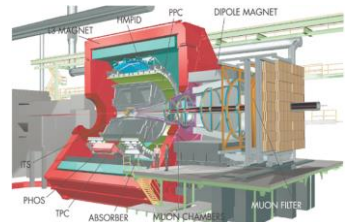
	Level-1 kHz	Event MByte	Storage MByte/s	
ATLAS	100	1	100	~ 3PB/year



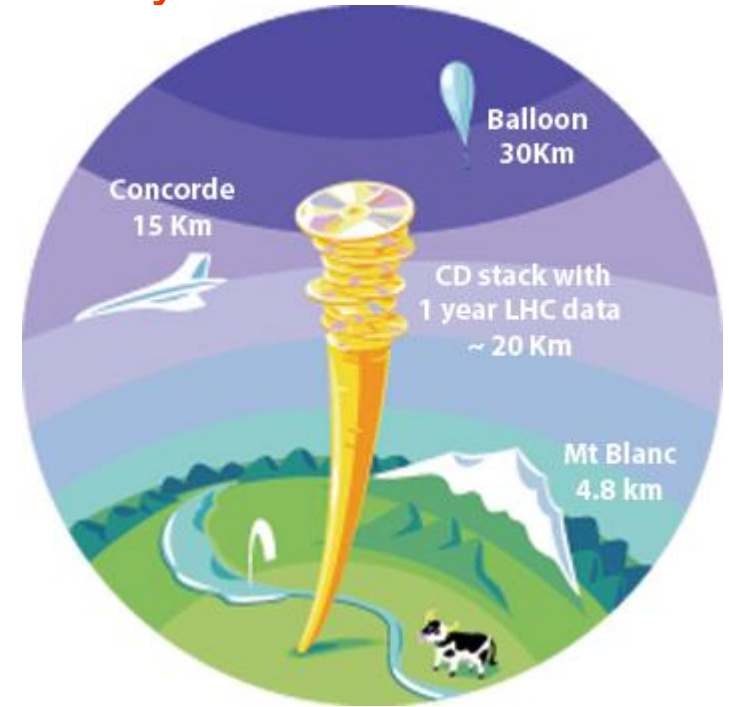
CMS	100	1	100	
------------	-----	---	-----	--



LHCb	400	0.1	20	
-------------	-----	-----	----	--

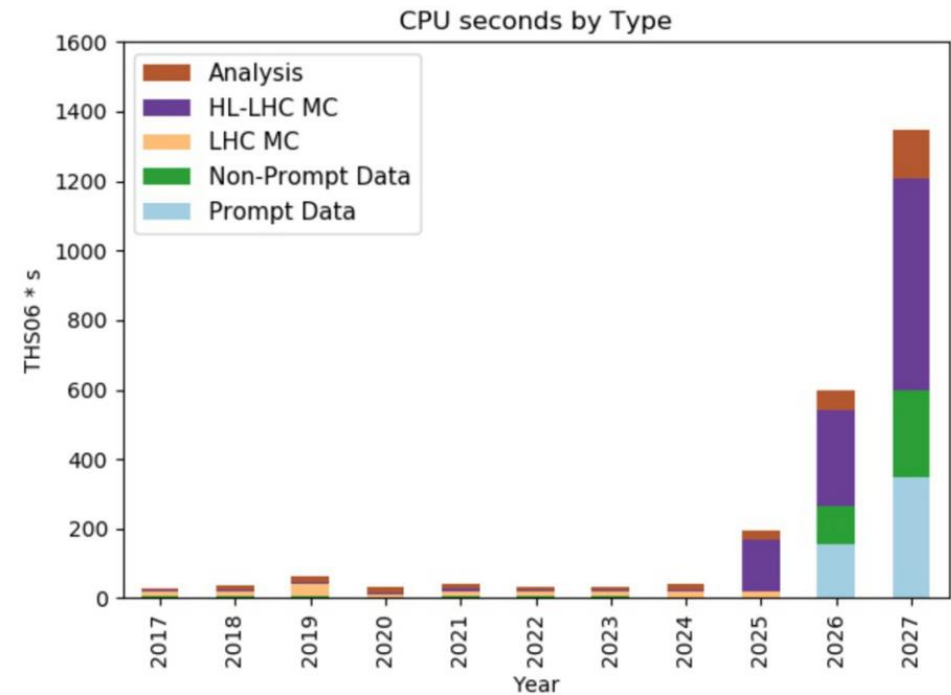
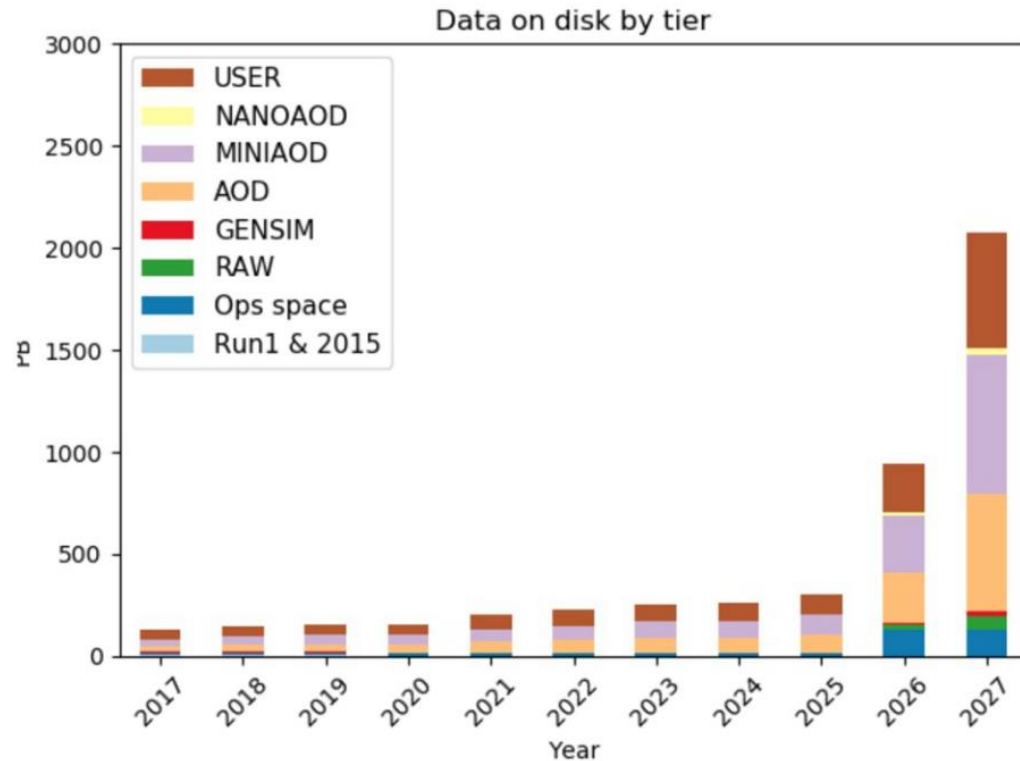


ALICE	1	25	1500	
--------------	---	----	------	--



Big Data/GRID Computing

CMS @ LHC Example

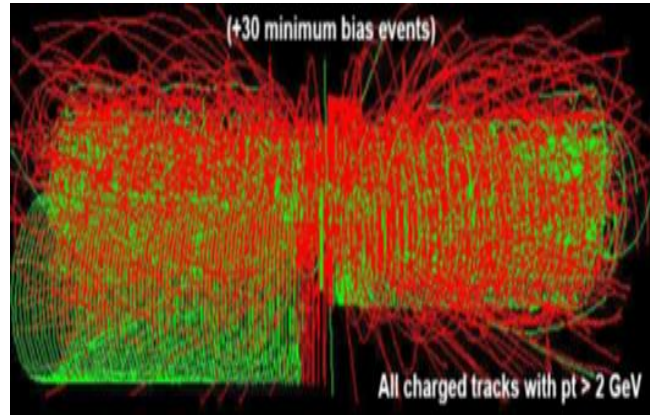


- Facing up to the exabyte (10^{18} bytes) era \Rightarrow required computing capacity is roughly 10 times higher than today
- an improvement of around a factor 10 in processing capabilities

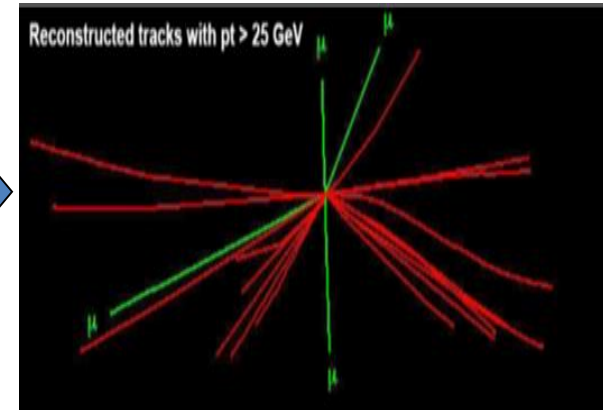
RAW Data



Reconstruction

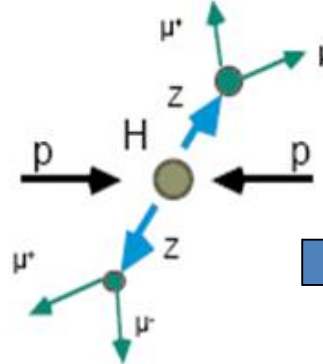


Event Selection

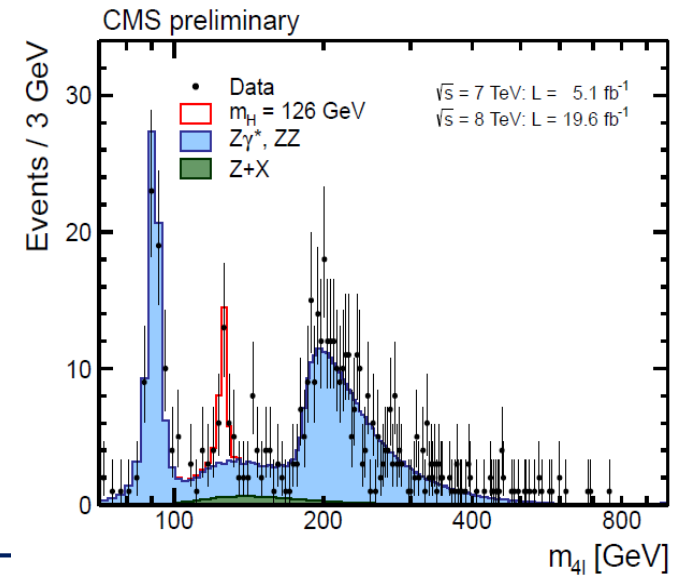
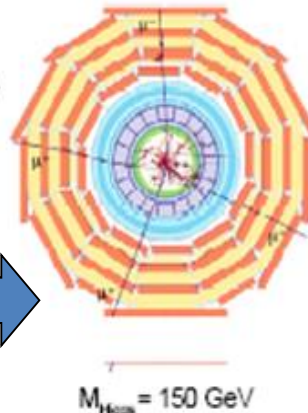


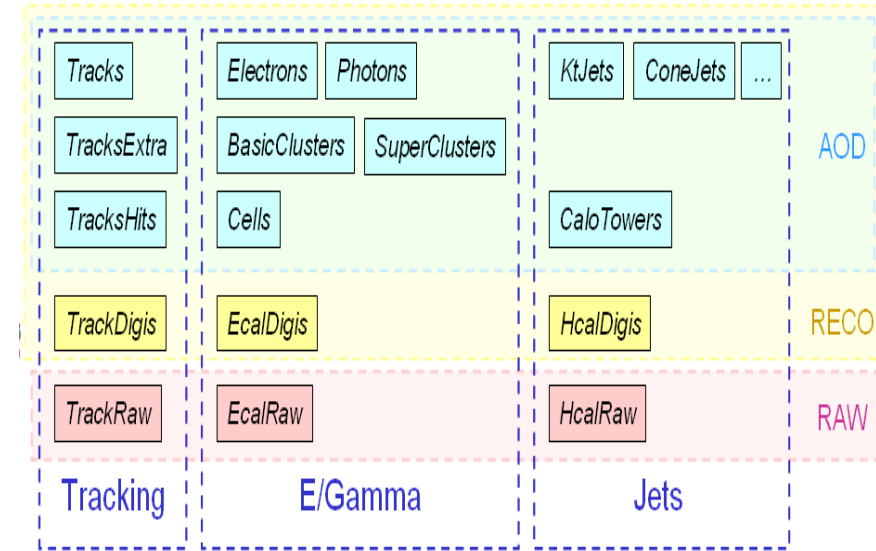
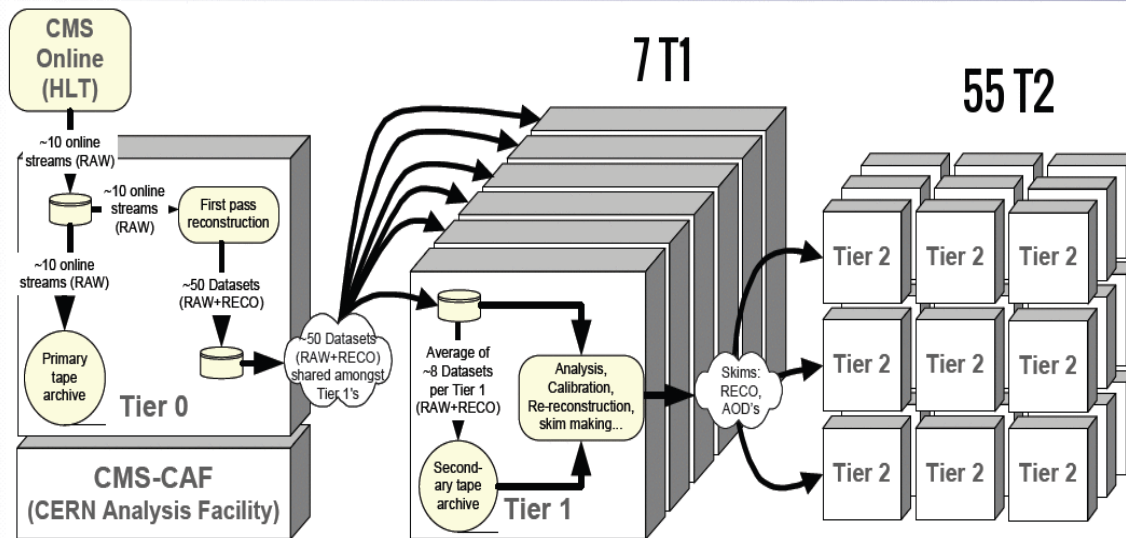
Calibration/Condition/etc
Data Bases

Theory/ Monte Carlo Simulation

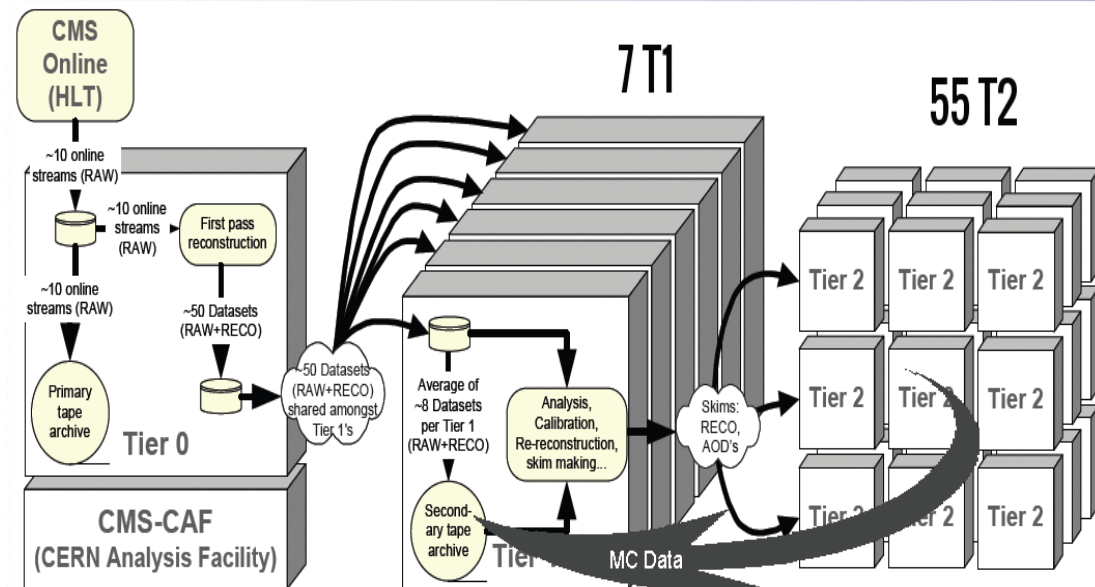


Detector Response Simulation



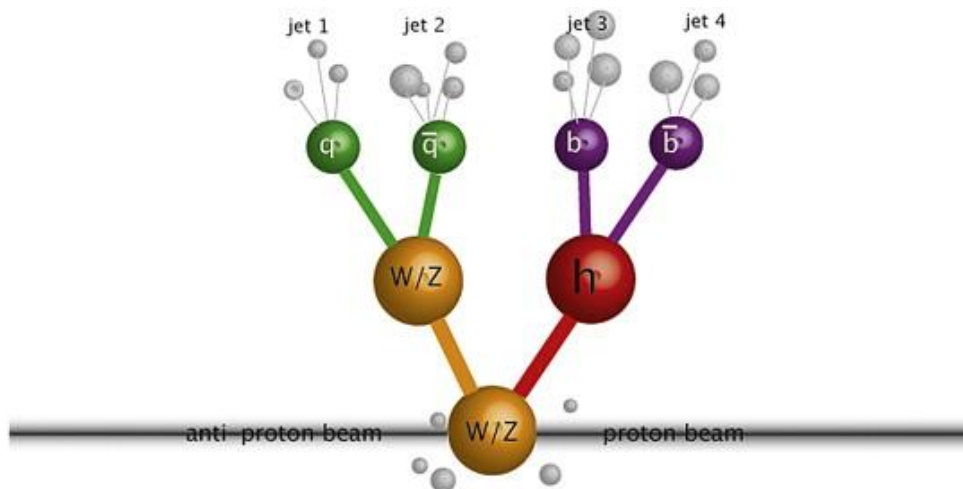


- T0 ⇒ T1**
 - ✓ scheduled, time-critical, will be continuous during data-taking periods
 - ✓ reliable transfer needed for fast access to new data, and to ensure that data is stored safely
- T1 ⇒ T1:**
 - ✓ redistributing data, generally after reprocessing (e.g. processing with improved algorithms)
- T1 ⇒ T2:**
 - ✓ Data for analysis at Tier-2s

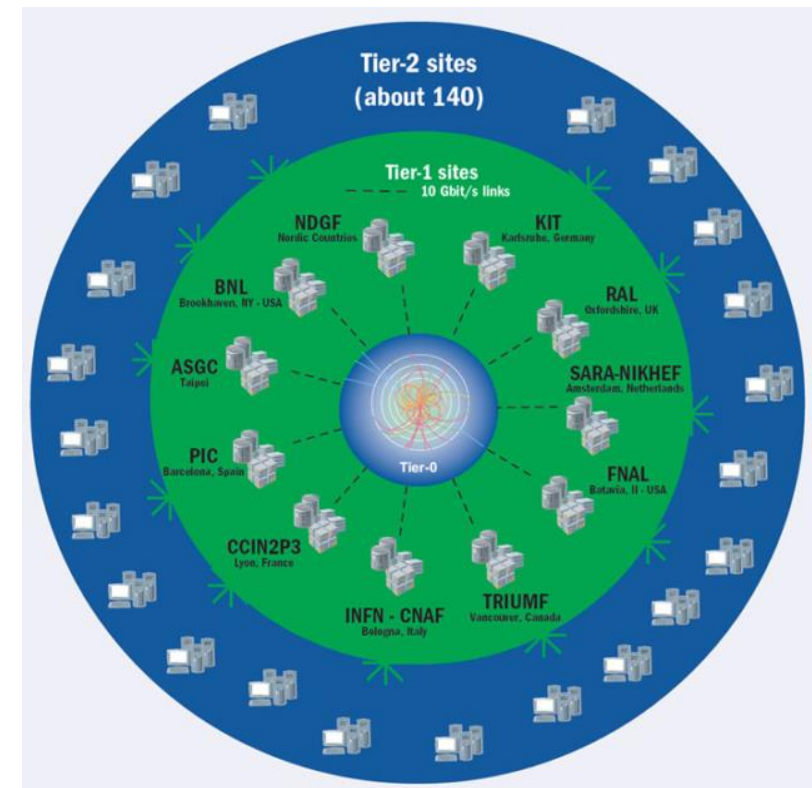


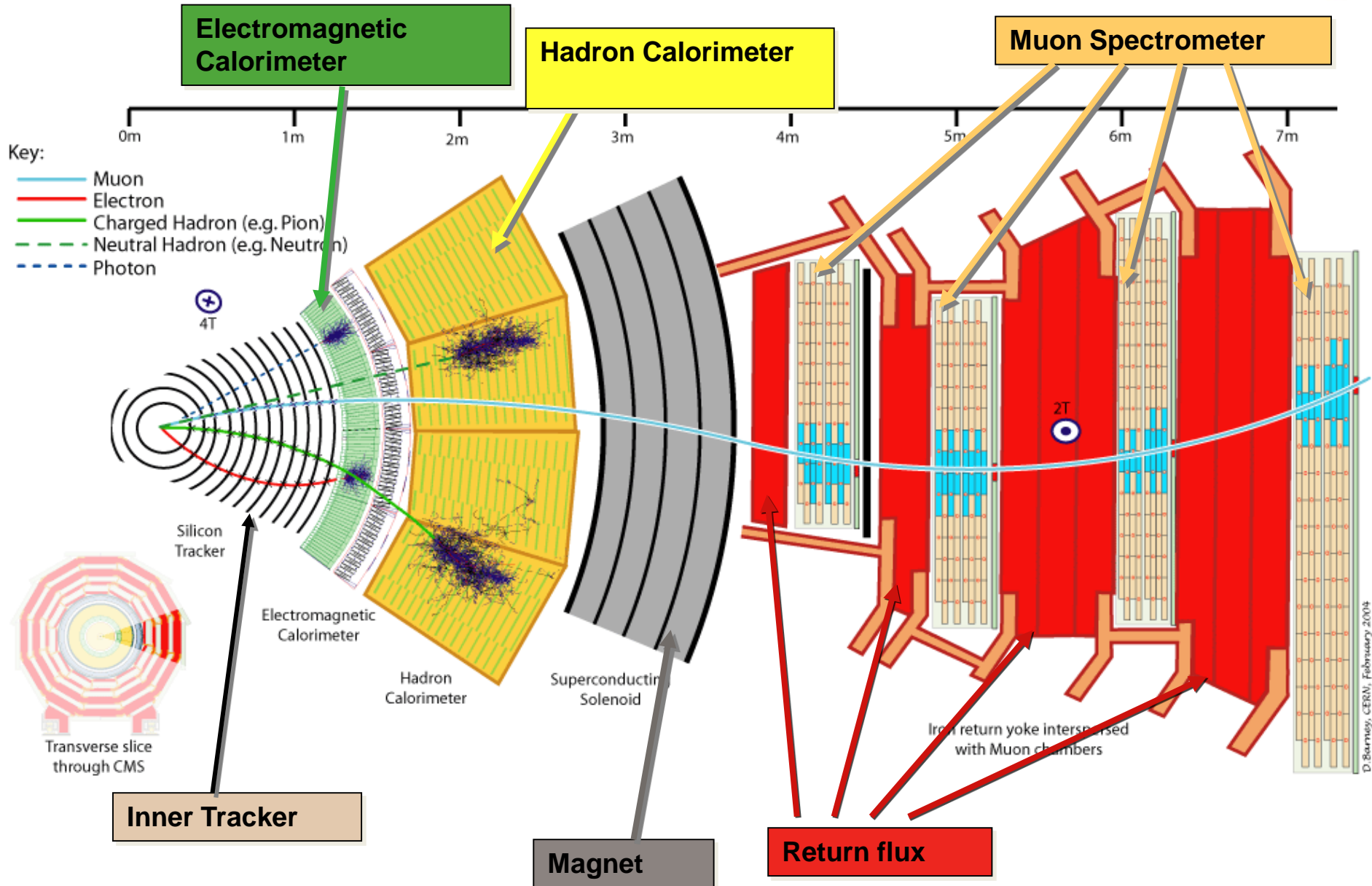
Event Reconstruction

- Reconstruction (mathematical methods/algorithms/SW)
 - ✓ physics objects - stable particles (e , μ , γ), clusters of particles (energy), vertexes, etc
 - ✓ unstable particles/ physics processes

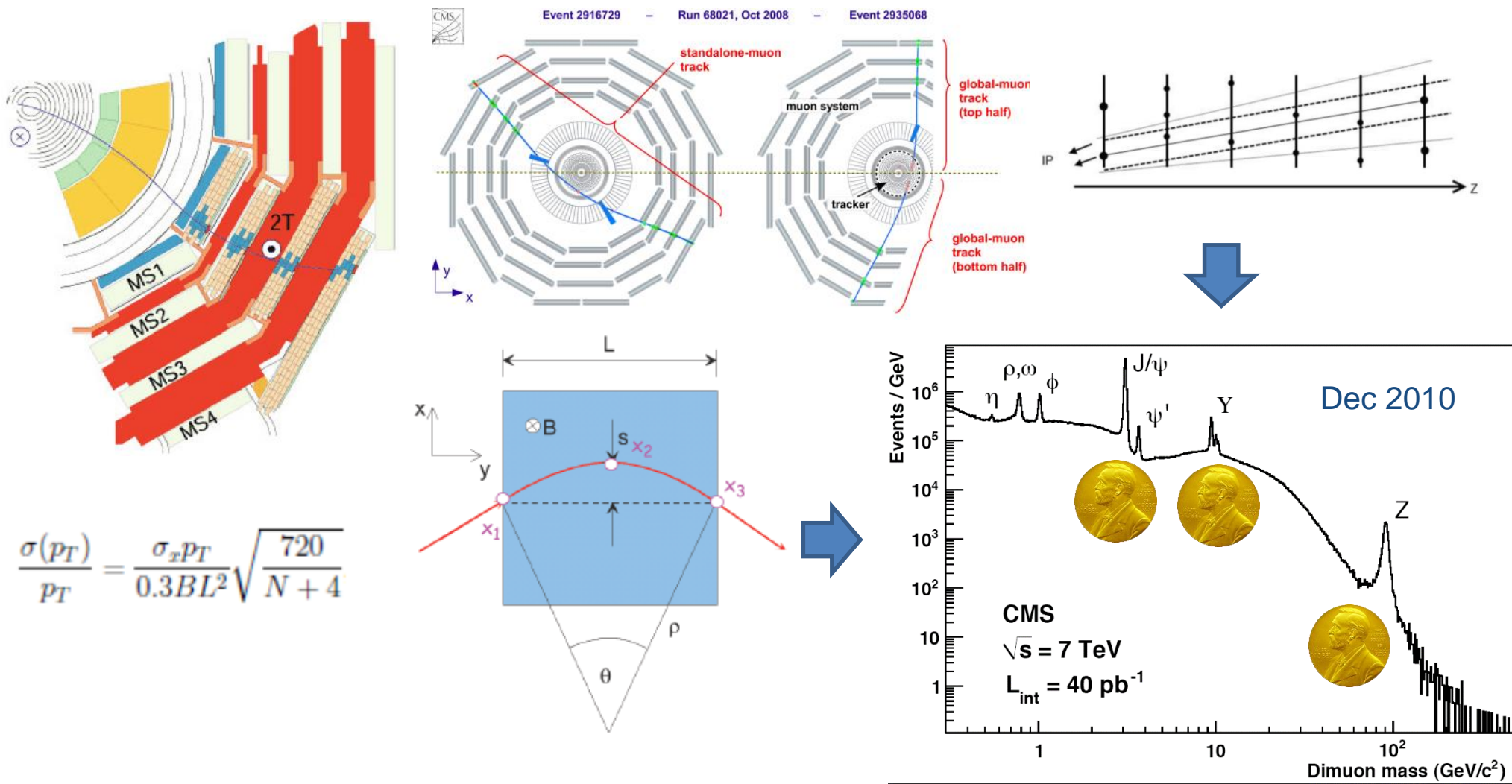


- Data Processing

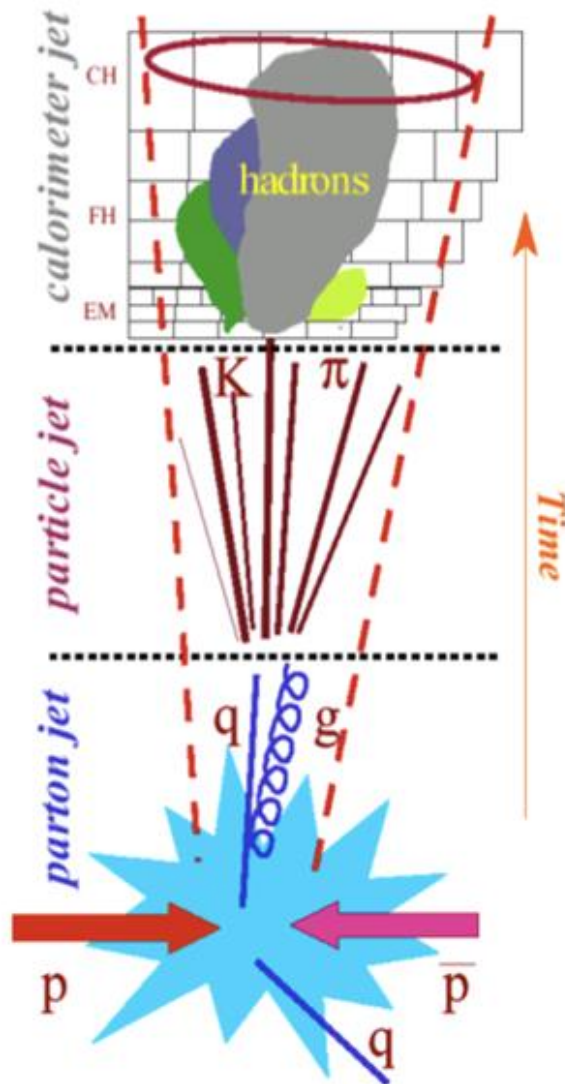




CMS Muon System shows a excellent performance to detect different resonances



<https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsMUO>



• Calorimeter jet (cone)

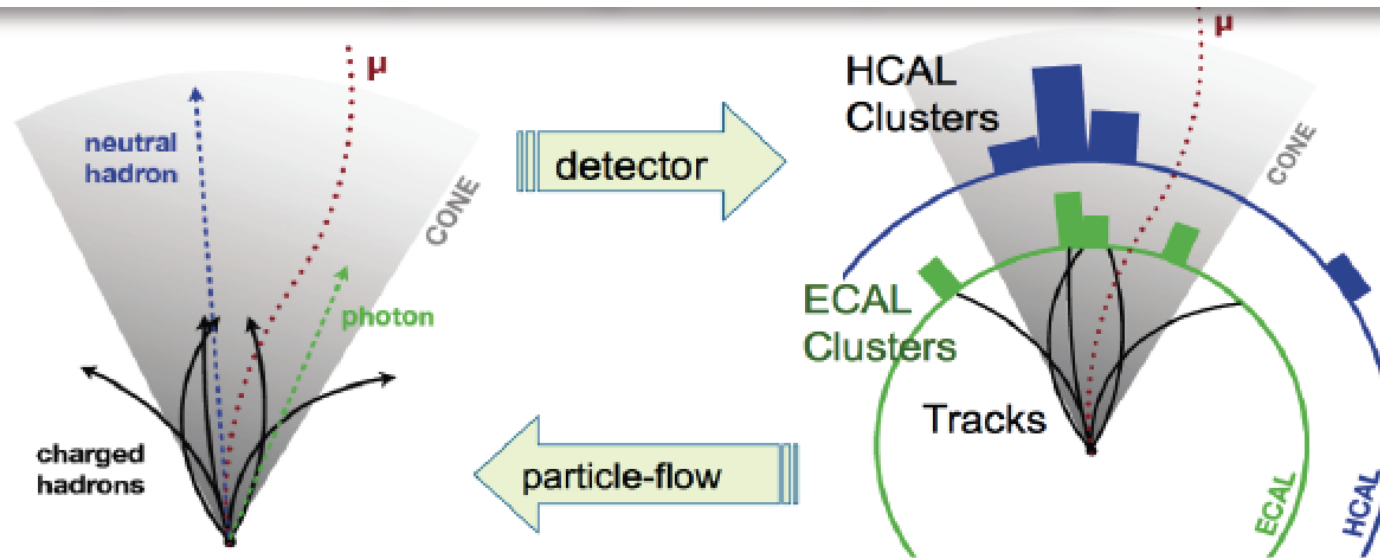
- ◆ jet is a collection of energy deposits with a given cone R : $R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$
- ◆ cone direction maximizes the total E_T of the jet
- ◆ various clustering algorithms

- correct for finite energy resolution
- subtract underlying event
- add out of cone energy

• Particle jet

- ◆ a spread of particles running roughly in the same direction as the parton after hadronization

Using all information of the detector together for optimal measurement



- Optimal combination of information from all subdetectors
- Returns a list of reconstructed particles
 - e, μ, γ , charged and neutral hadrons
 - Used in the analysis as if it came from a list of generated particles
 - Used as building blocks for jets, taus, missing transverse energy, isolation and PU particle identification

HL-LHC: elephant in the room

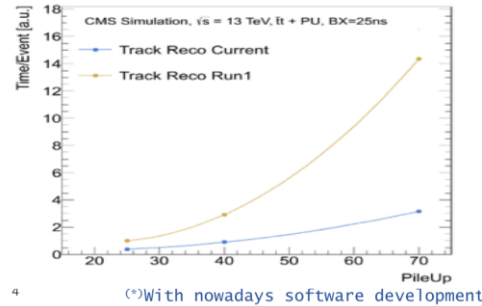


- ▶ ~40 collisions/event
- ▶ ~10 sec/event processing time
- ▶ (at best) Same computing resources as today
- ▶ ~200 collisions/event
- ▶ ~minute/event processing time^(*)
- ▶ (at best) Same computing resources as today

Flat budget vs. more needs = current rule-based reconstruction algorithms will not be sustainable

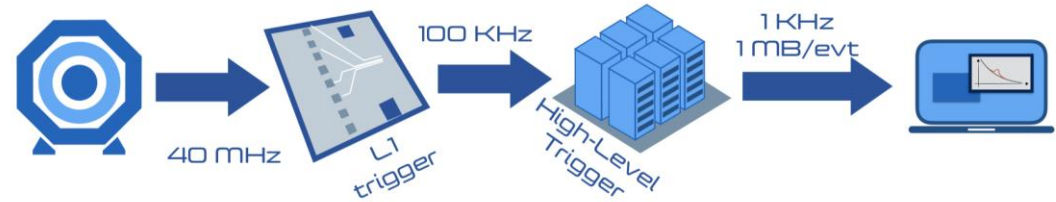
Adopted solution: more granular and complex detectors → more computing resources needed → more problems

Modern Machine Learning might be the way out



The LHC Big Data Problem

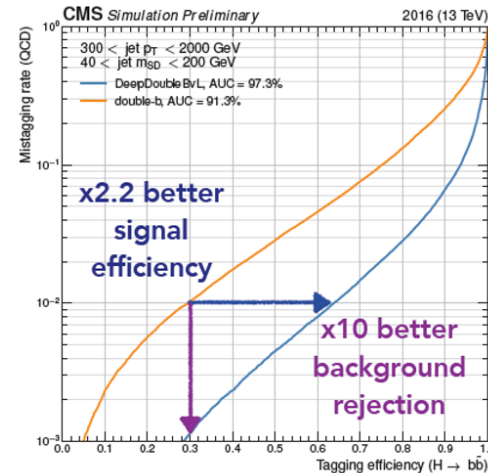
Too many data, too large data → need to filter online



The solution to the HL-LHC problem: modern Machine Learning as a fast shortcut between the data and the right answer (the outcome of our traditional & slow algorithms)

DP-2018/033 DEEP DOUBLE-B TAGGER

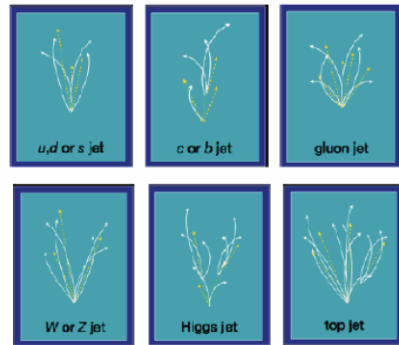
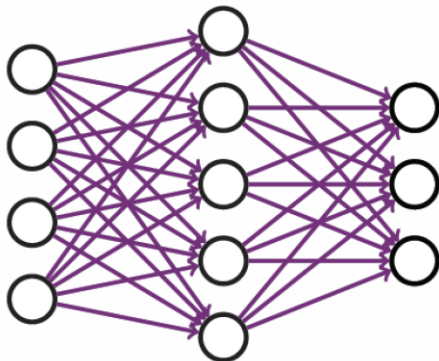
- Large performance gain over previous algorithm



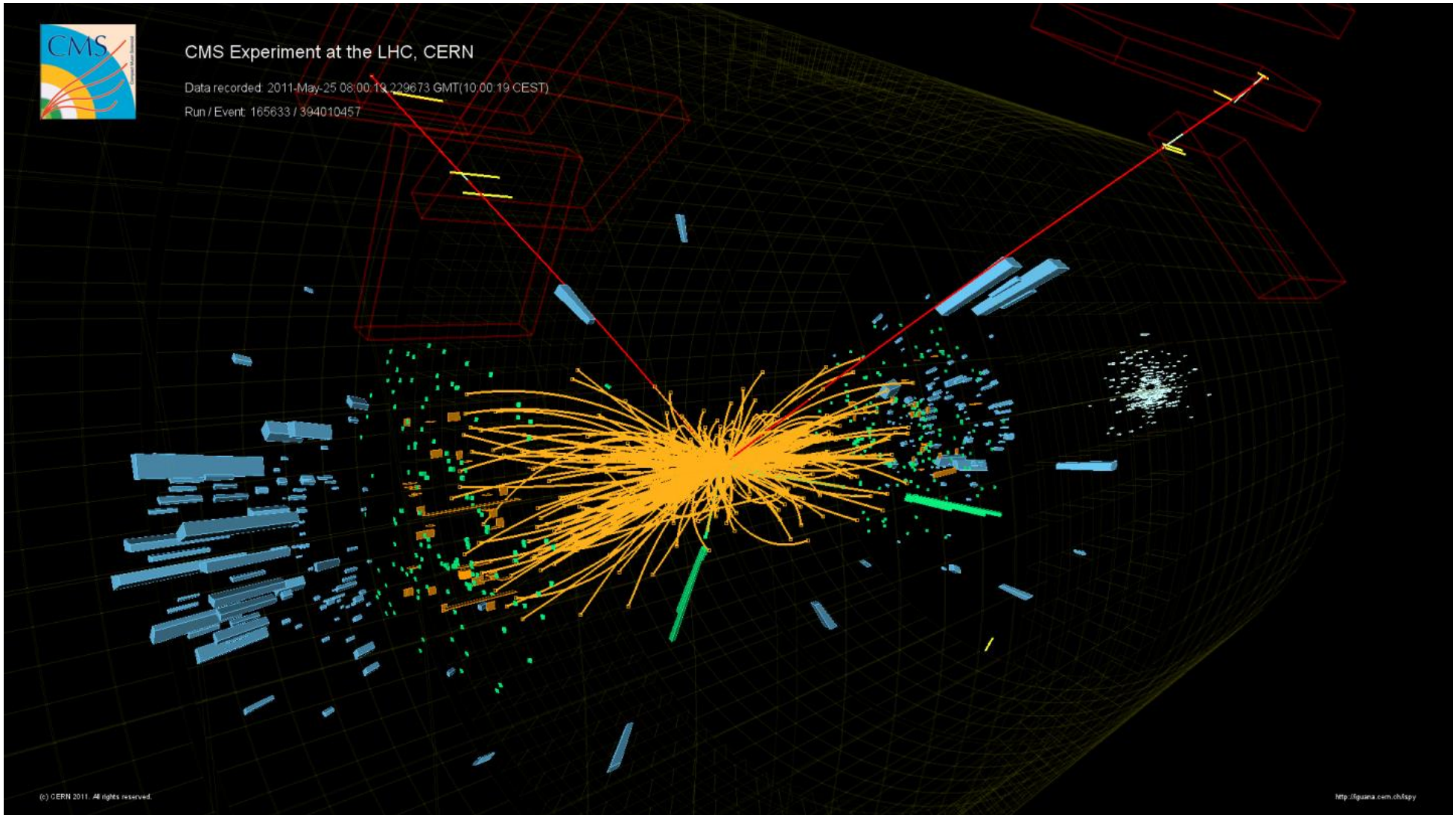
DEEP LEARNING TECHNIQUES

Deep neural networks

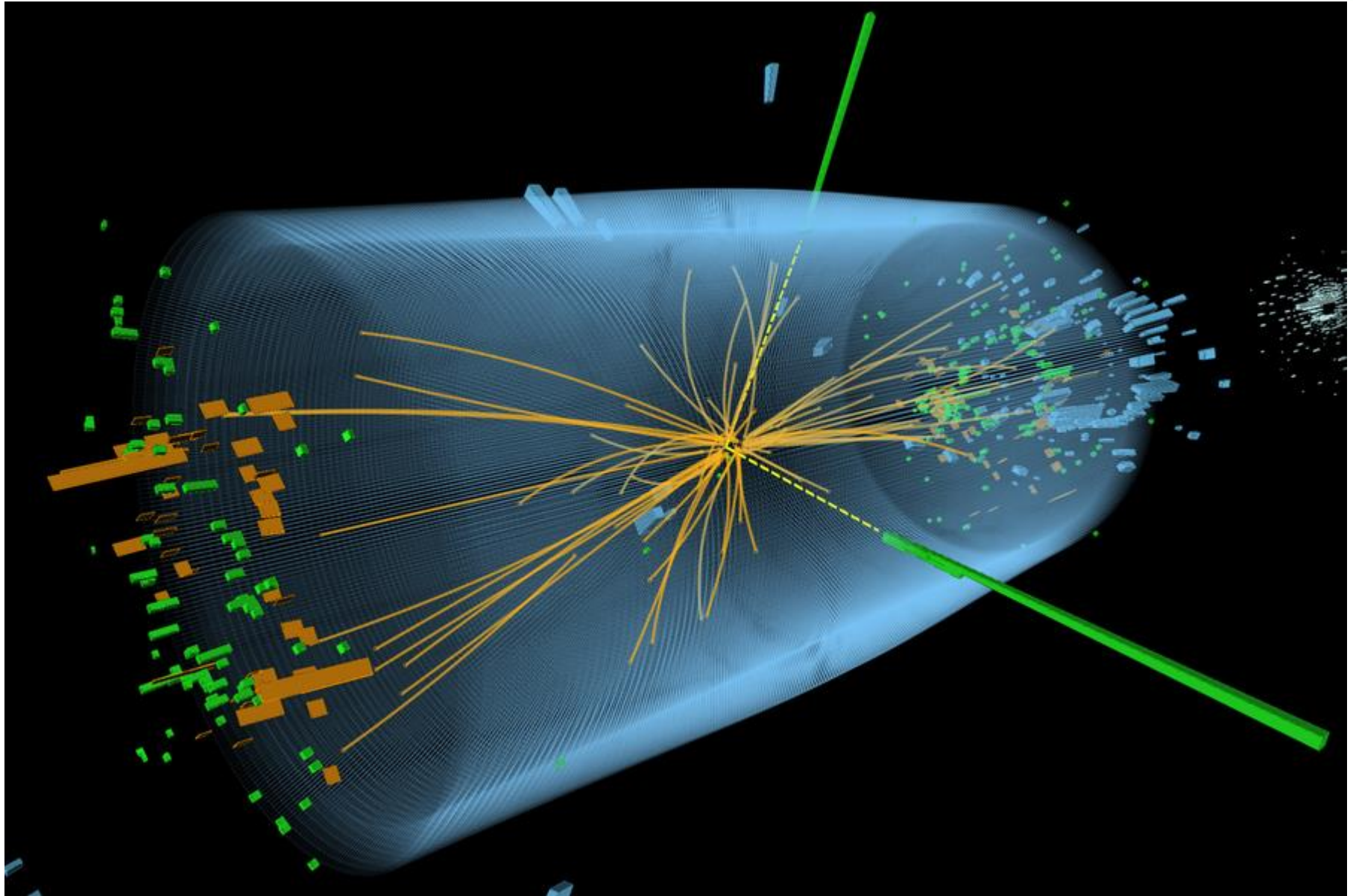
based on many low-level features with large training data sets to classify jets



Example of $h \rightarrow ZZ \rightarrow 2e 2\mu$

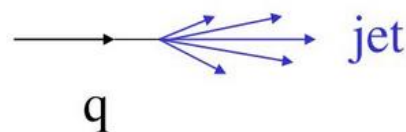
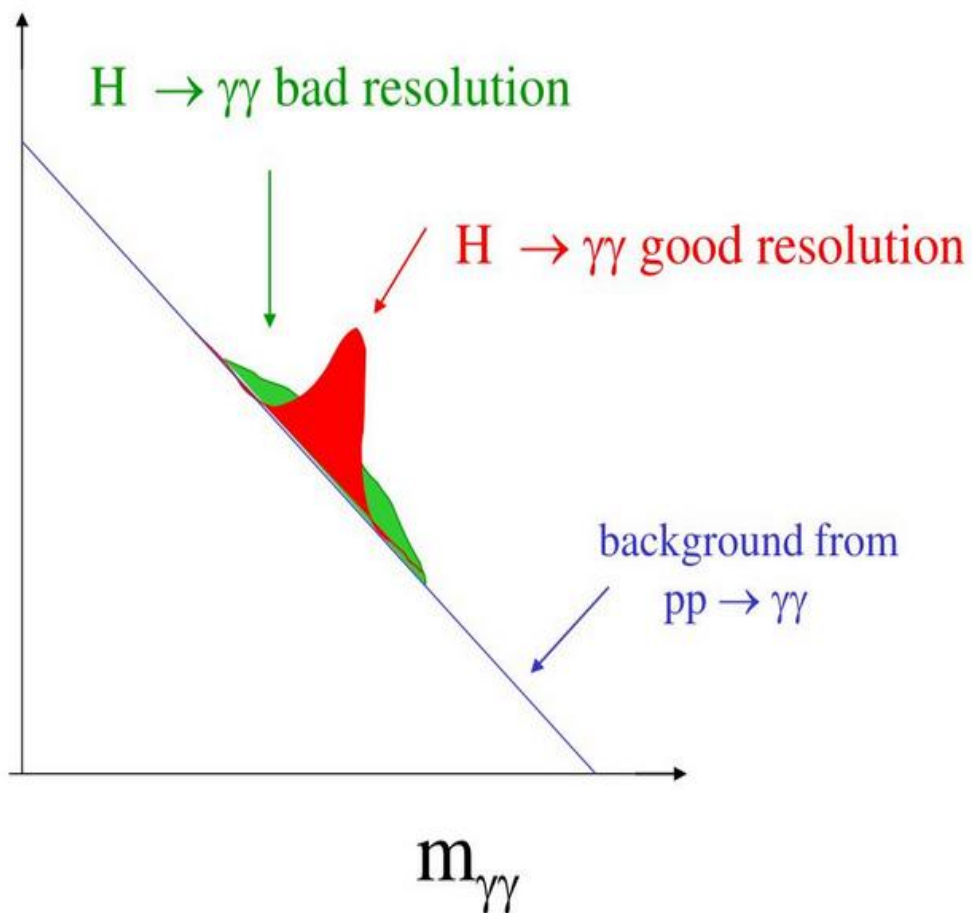


Example of $h \rightarrow 2\gamma$

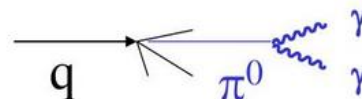




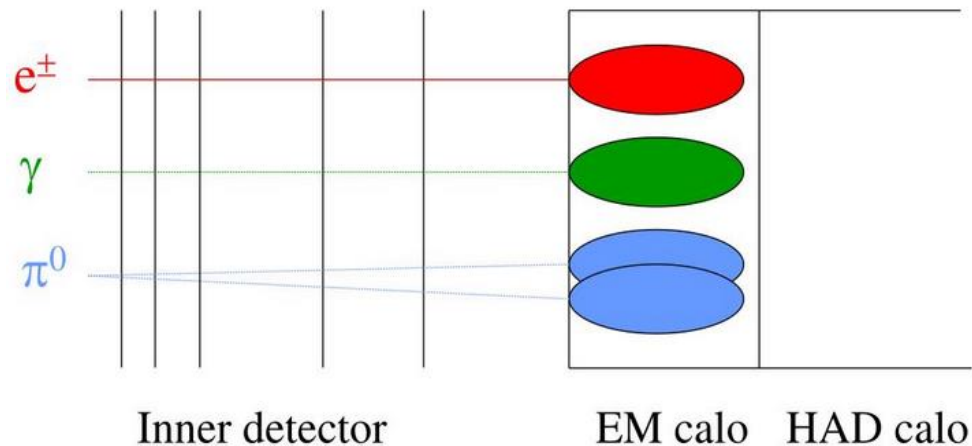
Example : $H \rightarrow \gamma\gamma$



number and p_T of hadrons in a jet have large fluctuations



in some cases: one high- p_T π^0 ; all other particles too soft to be detected

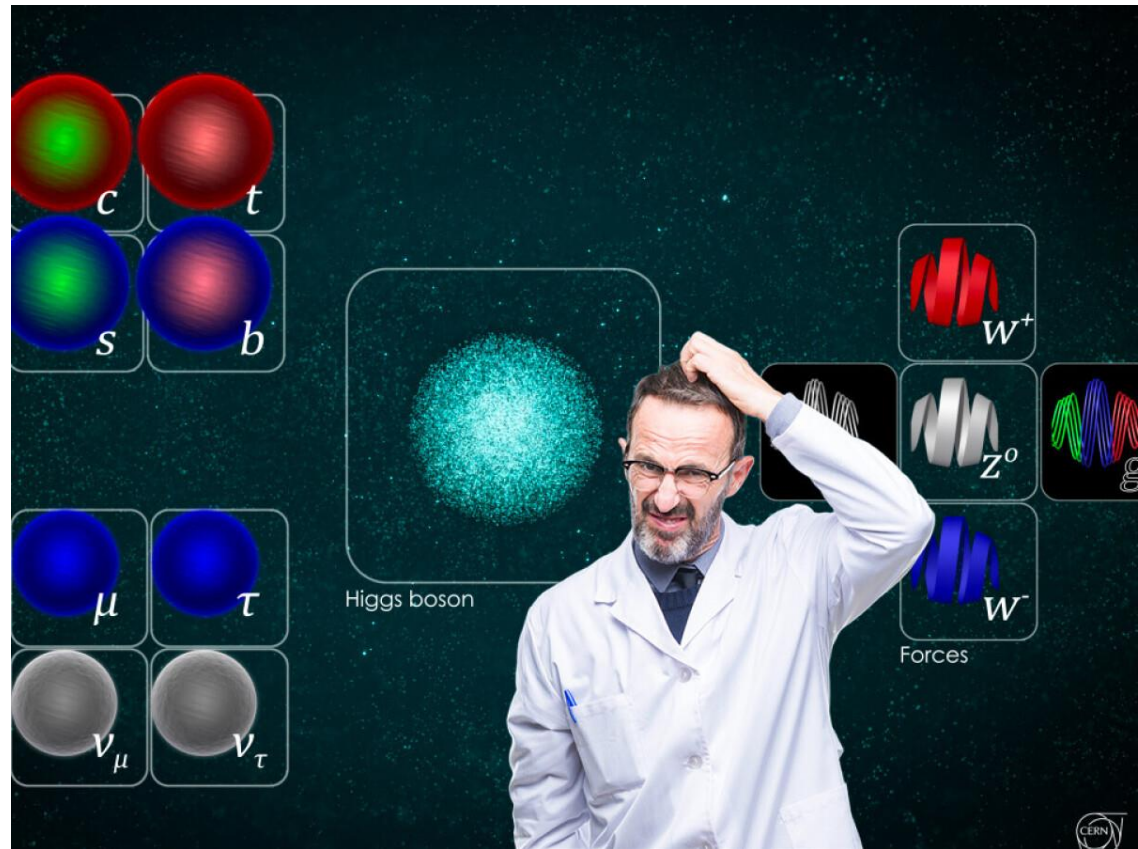
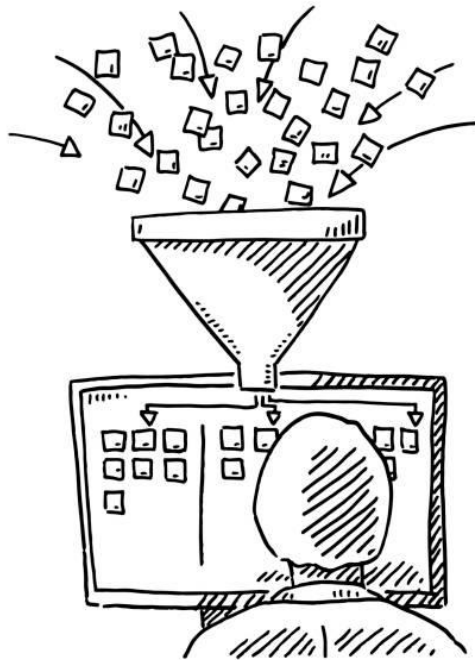


$d(\gamma\gamma) < 10$ mm in calorimeter \rightarrow QCD jets can mimic photons. Rare cases, however:

$$\frac{\sigma_{jj}}{\sigma(H \rightarrow \gamma\gamma)} \sim 10^8 \quad m_{\gamma\gamma} \sim 100 \text{ GeV}$$

Data Analysis

- Data vs Theory \Rightarrow which theories you believe vs. reject
- Significance of final results \Rightarrow do you trust your analysis or not?



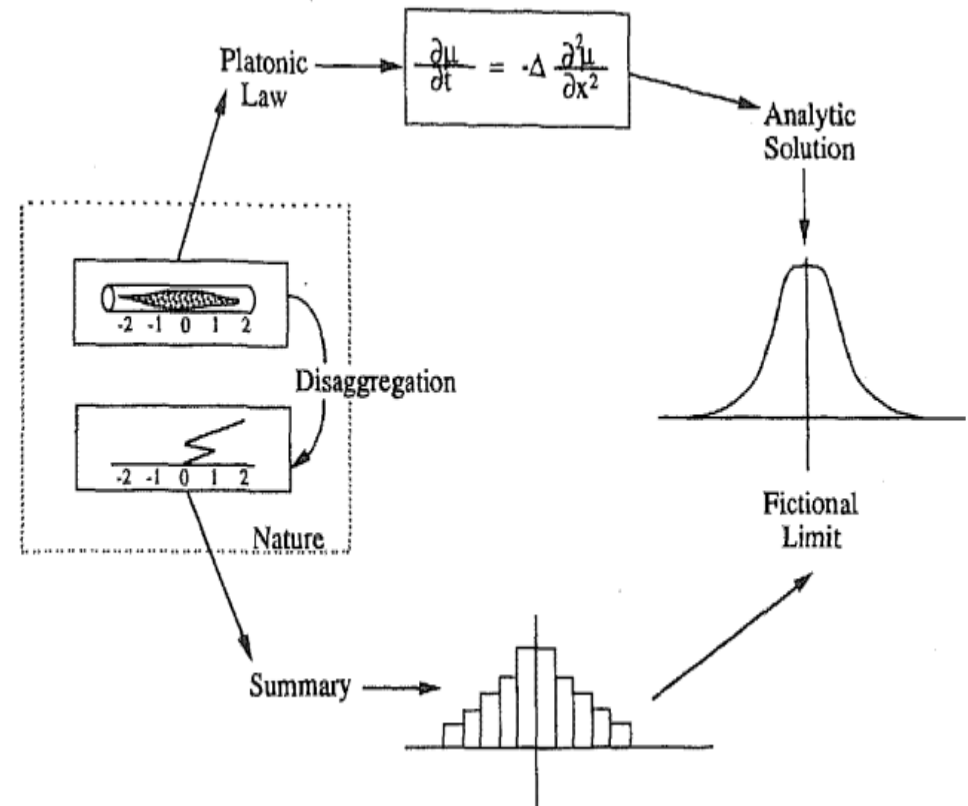
Data Analysis: Theory and Modeling (Monte Carlo Simulation)

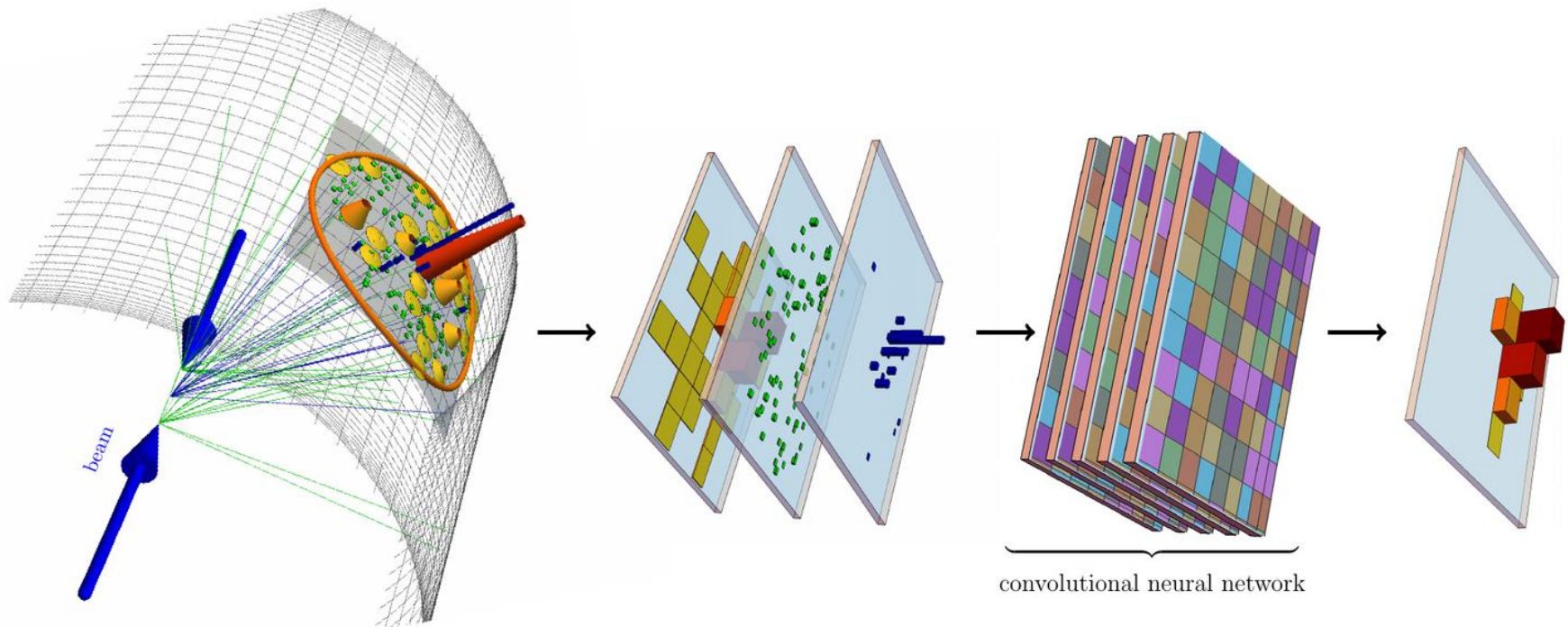
Three main goals

- experiment planning
- algorithm's training
- data/MC comparison

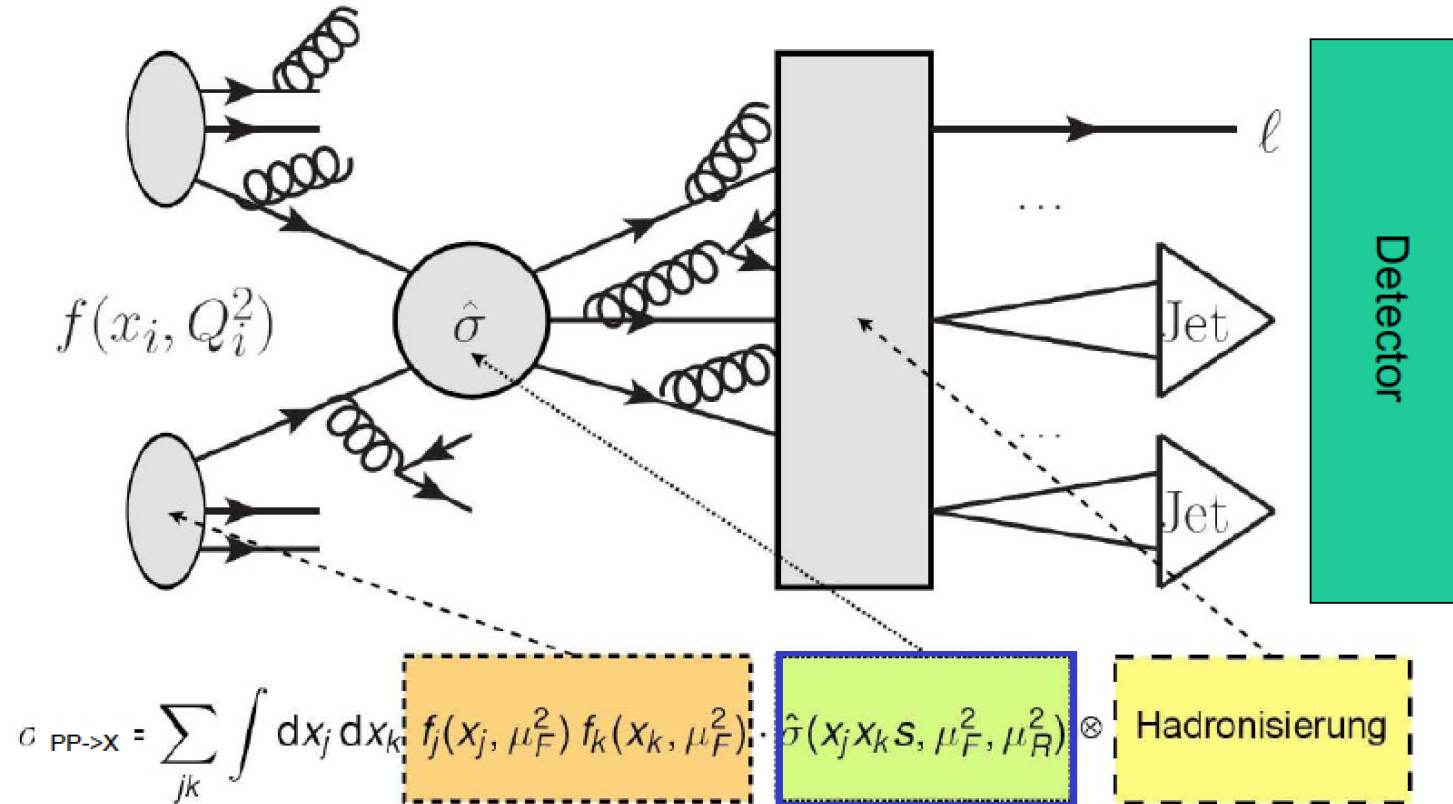
Digital Twin of Experiments

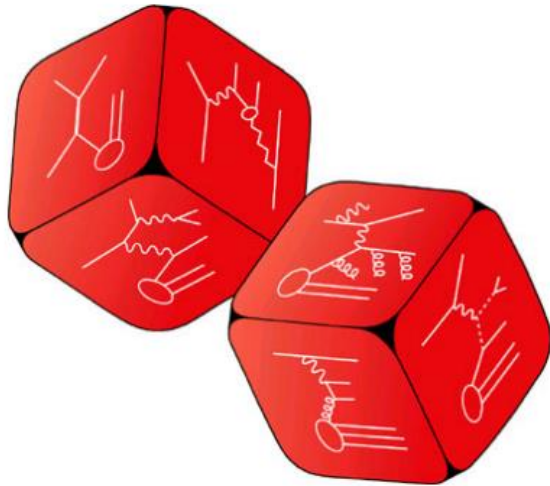
- physics in a collision point
- models of detector systems
- response from detectors including digitization
- processing of MC data (simulation of data flow)





Cross Section = PDFs X Sub Process X Hadronisation





Three general-purpose generators:

- HERWIG
- PYTHIA
- SHERPA

Many others good/better at some specific tasks.

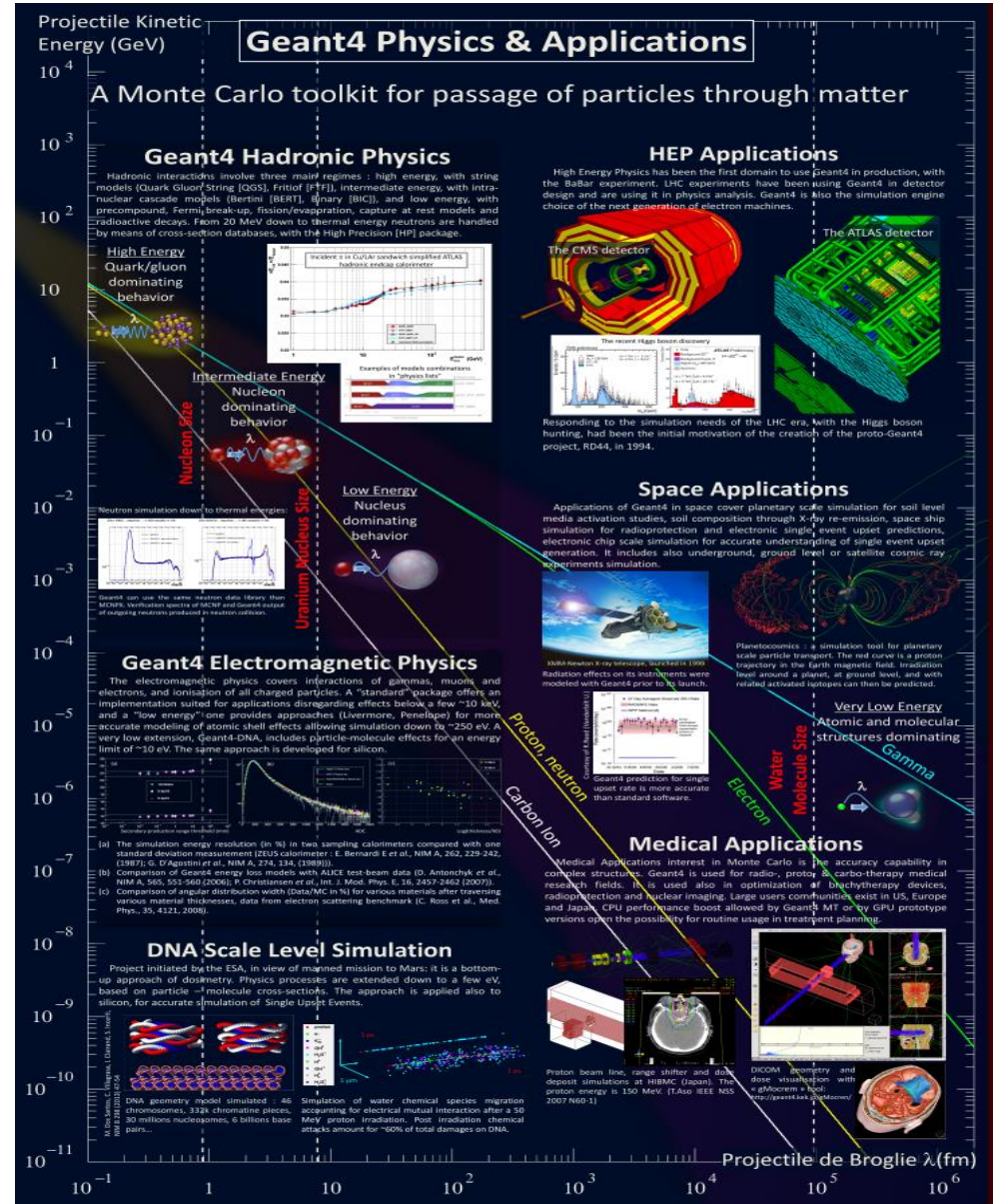
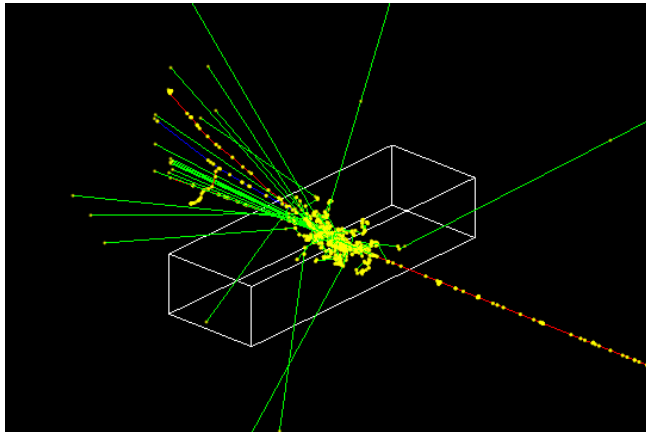
Generators to be combined with detector simulation (GEANT)

accelerator/collisions \Leftrightarrow event generator
 detector/electronics \Leftrightarrow detector simulation

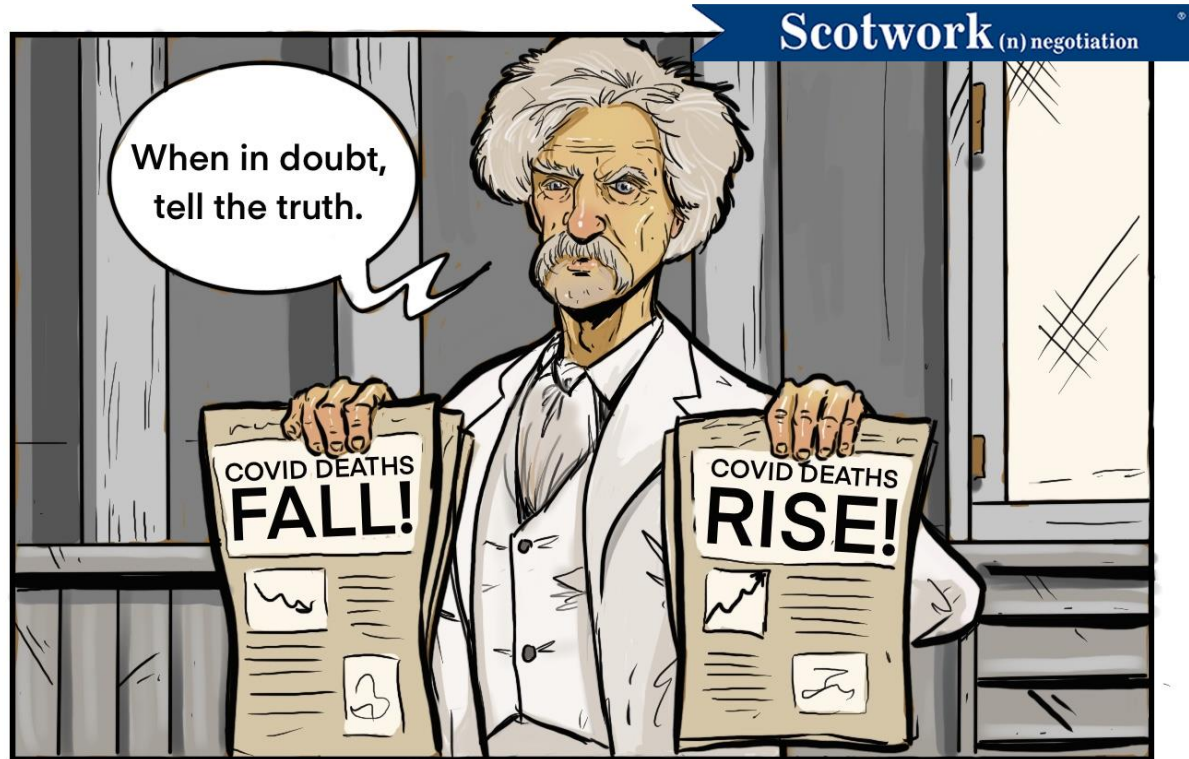
- to be used to
- predict event rates and topologies
 - simulate possible backgrounds
 - study detector requirements
 - study detector imperfections

GEANT4

- Toolkit created by CERN to simulate the passage of particles through matter.
- Designed to make the physics used transparent within the toolkit, handle a wide range of geometries, and enable an easy adaptation of different physics to fit the application.

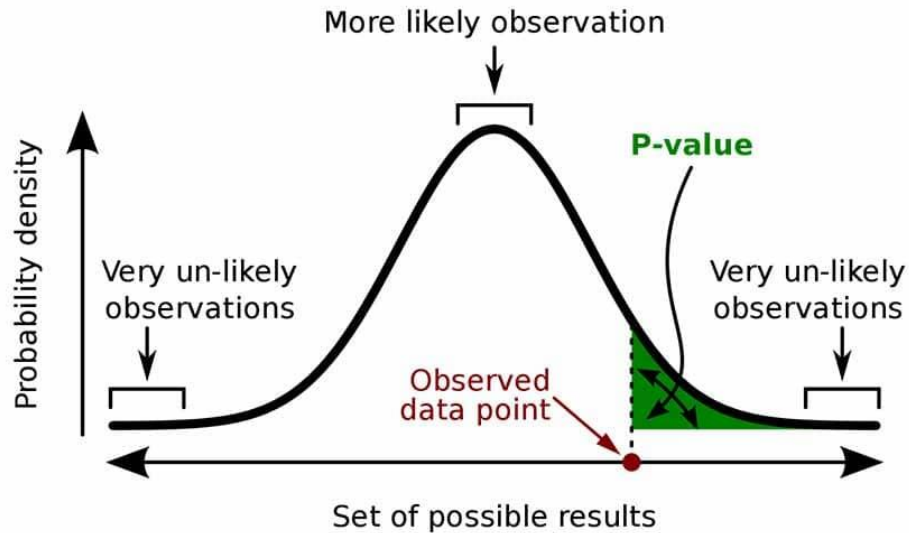


Data Analysis: Statistics

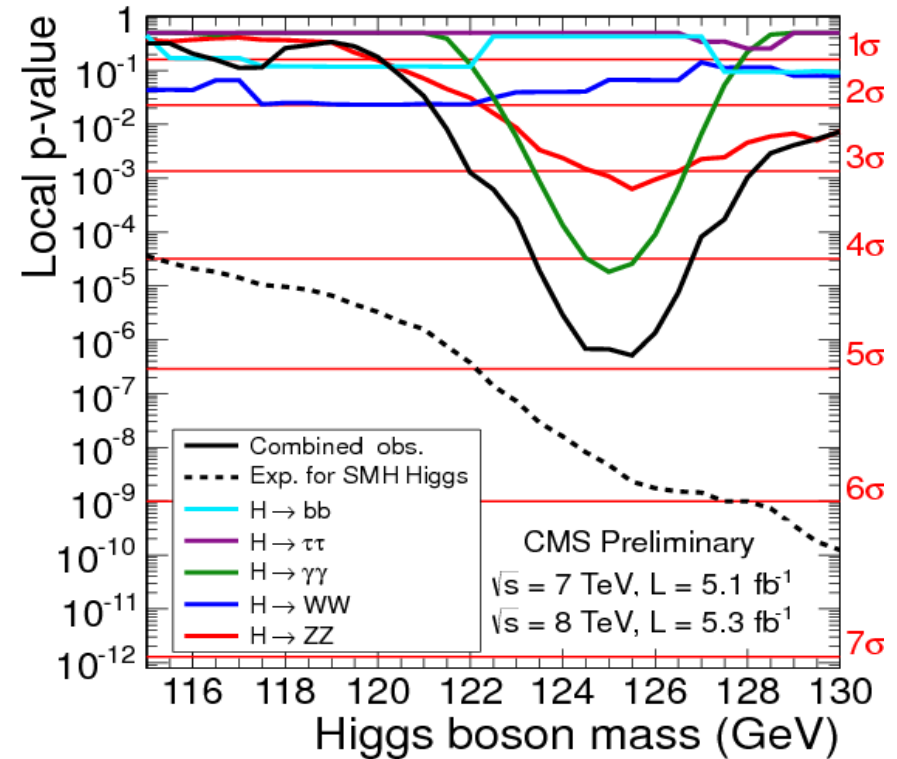


There are three kinds of lies: lies, damned lies, and statistics (c) Benjamin Disraeli

The probability that an observed excess was a statistical fluctuation of the background (p-value)

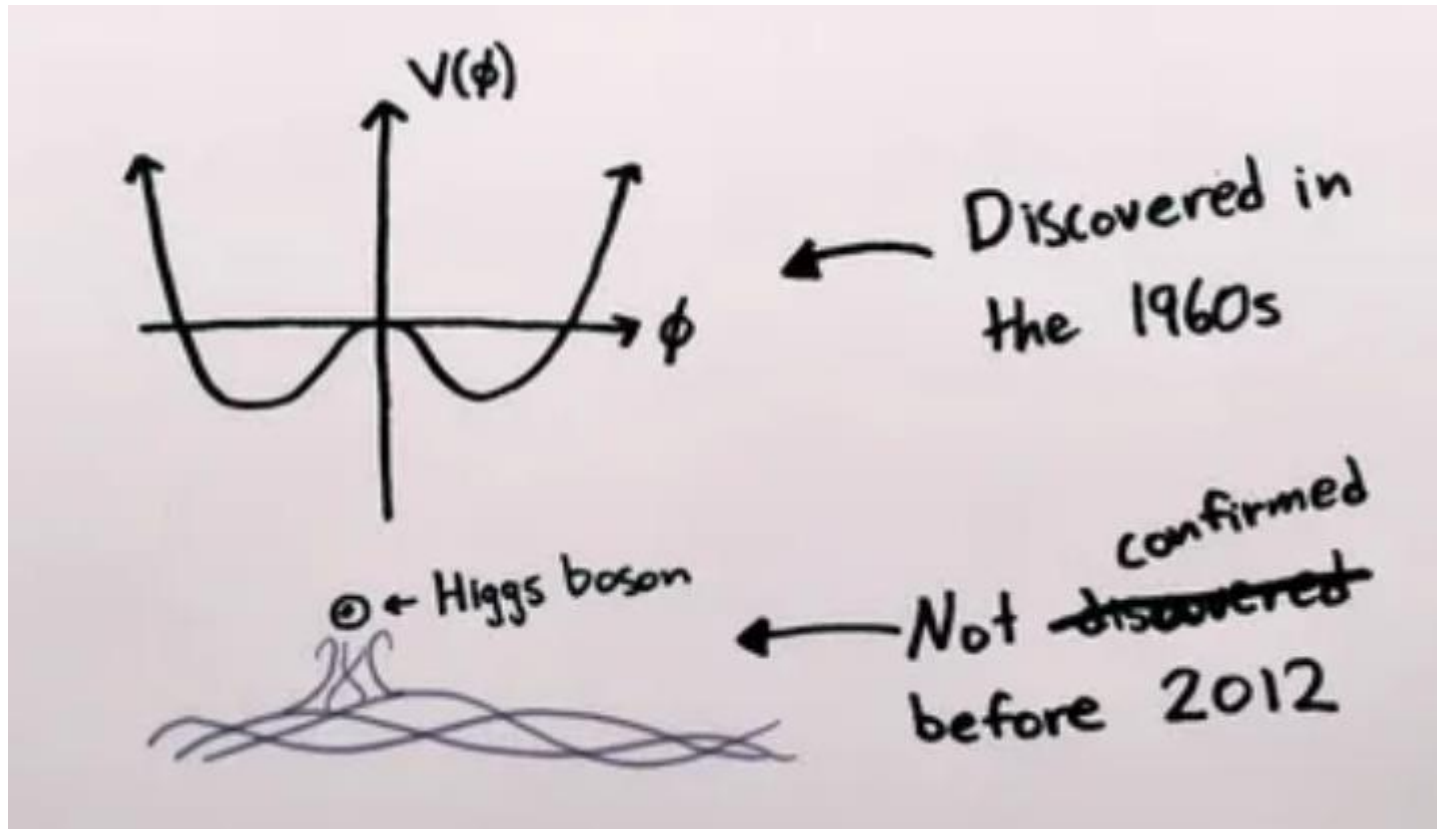


A **p-value** (shaded green area) is the probability of an observed (or more extreme) result assuming that the null hypothesis is true.



Notable values for an excess in particle physics are 3σ , or **p-value = 0.0013**; and 5σ , or **p-value = 2.87×10^{-7}** . When we have an excess of 3σ we talk about an evidence, and when we have an excess of 5σ , we are facing a discovery.

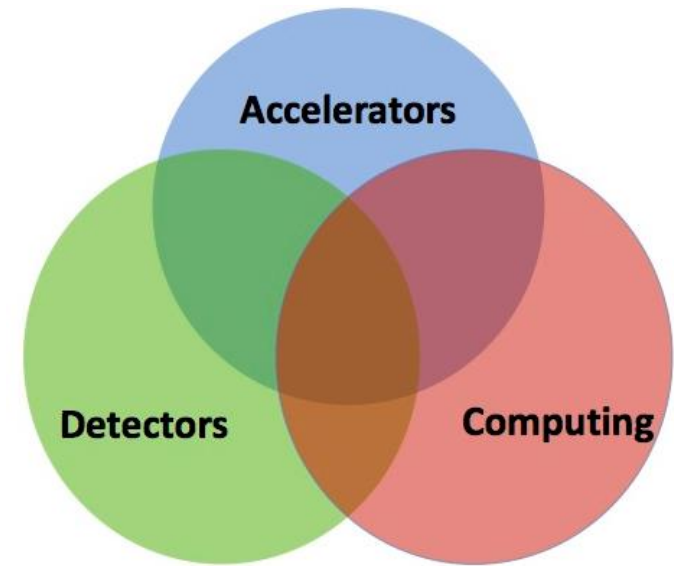
... and as a result...



Accelerators : powerful machines that accelerate particles to extremely high energies and bring them into collision with other particles

Detectors : gigantic instruments that record the resulting particles as they “stream” out from the point of collision.

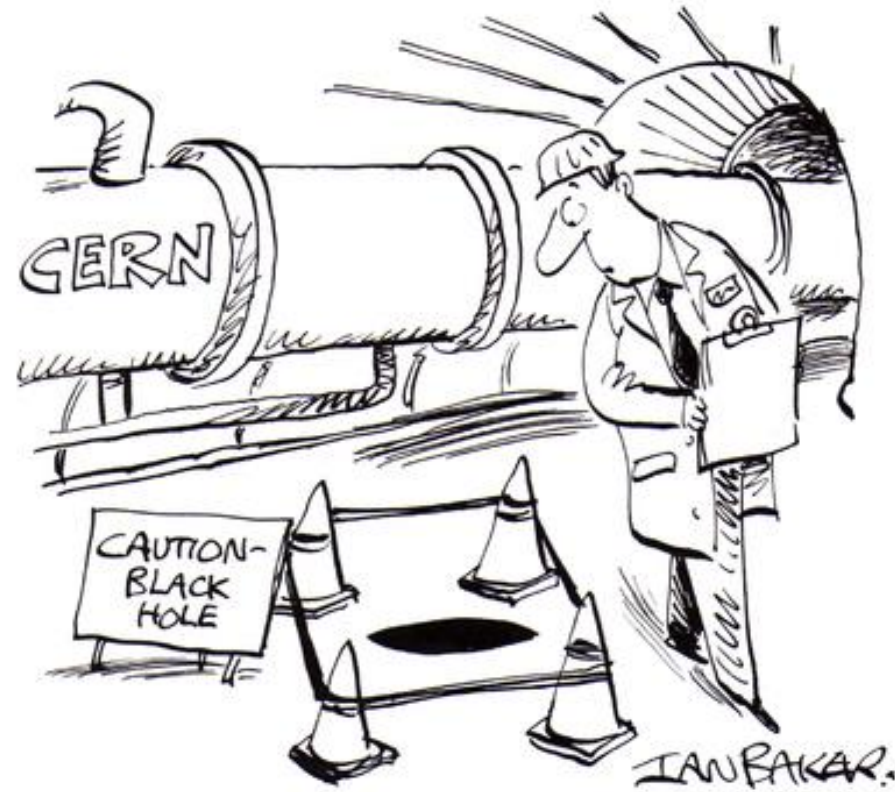
Computing : to collect, store, distribute and analyse the vast amount of data produced by these detectors



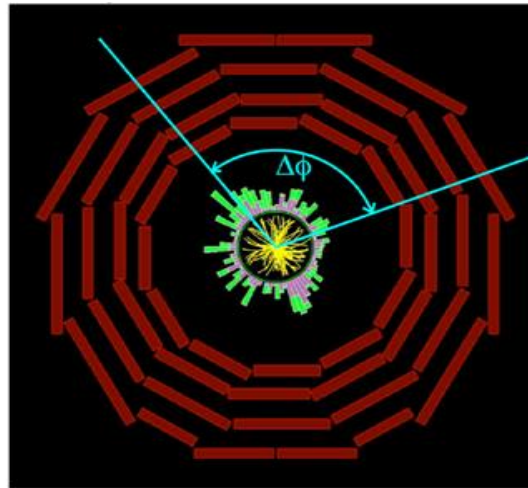
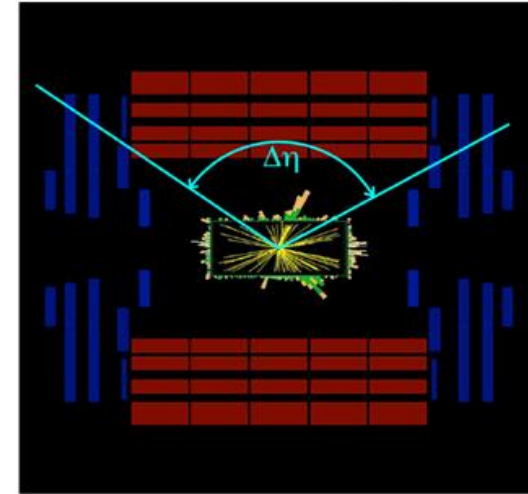
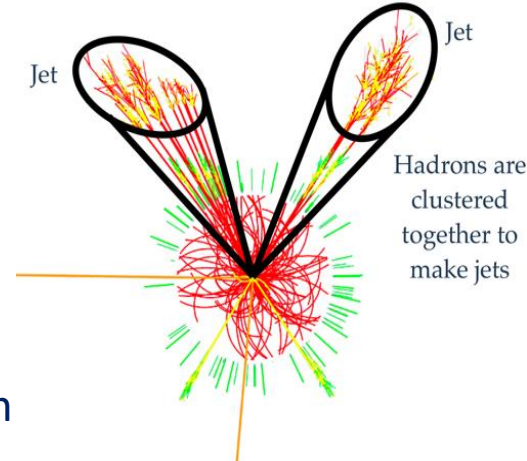
It's been a global effort, a global success. It has only been possible because of the extraordinary achievements of the experiments, infrastructure and the grid computing“
 (c) Rolf Heuer, the Director General of CERN, when the discovery of the Higgs

Collaborative Science on Worldwide scale : thousands of scientists, engineers, technicians and support staff to **design, build and operate these complex “machines”**.

THANK YOU FOR YOUR ATTENTION!

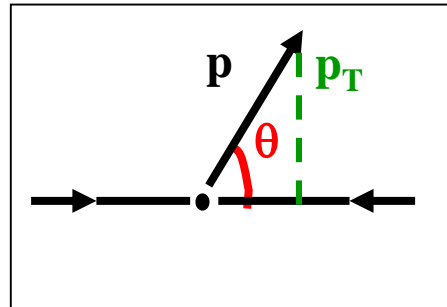


- Muons (transverse momentum p_T)
- Electrons (energy and tr. momentum p_T)
- Photons (energy)
- Jets (energy and coordinates)
-
- Missing energy and p_T
 - vectorial sum of all transverse momentum



Kinematic Variables

- Transverse momentum p_T (energy)
 - particles that escape detection have $p_T=0$
 - total visible $p_T = 0$
- Longitudinal momentum p_z and energy E_z
 - particles that escape detection have $p_T=0$
 - visible p_z is not conserved (not so usefull variable)



Angles

- azimuthal and polar angles
- polar angle θ is not Lorentz invariant \Rightarrow
- rapidity y
- or (or $m=0$) pseudorapidity η

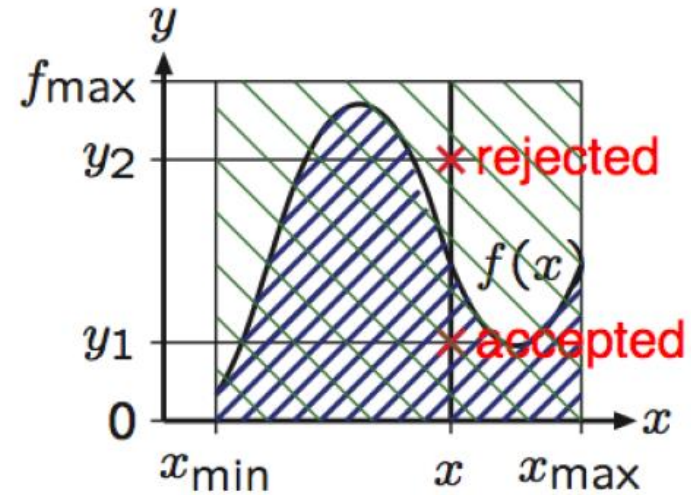
$$y \equiv \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right)$$

$$\eta = - \ln \left[\tan \left(\frac{\theta}{2} \right) \right]$$

4 π -experiments cover 360° over ϕ and large pseudorapidity range, $|\eta| \leq 5.0$ (0.8°)

If $f(x) \leq f_{\max}$ in $x_{\min} < x < x_{\max}$
use **interpretation as an area**

- ① select
 $x = x_{\min} + R(x_{\max} - x_{\min})$
- ② select $y = R f_{\max}$ (new R !)
- ③ while $y > f(x)$ cycle to 1



Integral as by-product:

$$I = \int_{x_{\min}}^{x_{\max}} f(x) dx = f_{\max} (x_{\max} - x_{\min}) \frac{N_{\text{acc}}}{N_{\text{try}}} = A_{\text{tot}} \frac{N_{\text{acc}}}{N_{\text{try}}}$$

Binomial distribution with $p = N_{\text{acc}}/N_{\text{try}}$ and $q = N_{\text{fail}}/N_{\text{try}}$,
so error

$$\frac{\delta I}{I} = \frac{A_{\text{tot}} \sqrt{pq/N_{\text{try}}}}{A_{\text{tot}} p} = \sqrt{\frac{q}{p N_{\text{try}}}} = \sqrt{\frac{q}{N_{\text{acc}}}} \rightarrow \frac{1}{\sqrt{N_{\text{acc}}}} \text{ for } p \ll 1$$

Событие (результат) называется “**статистически значимым**”, если оно вряд ли произошло случайно

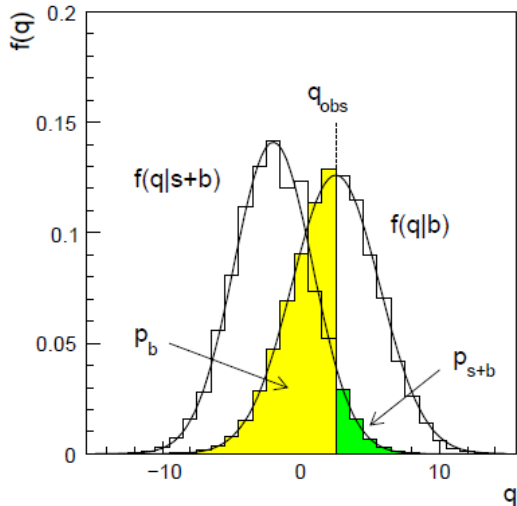
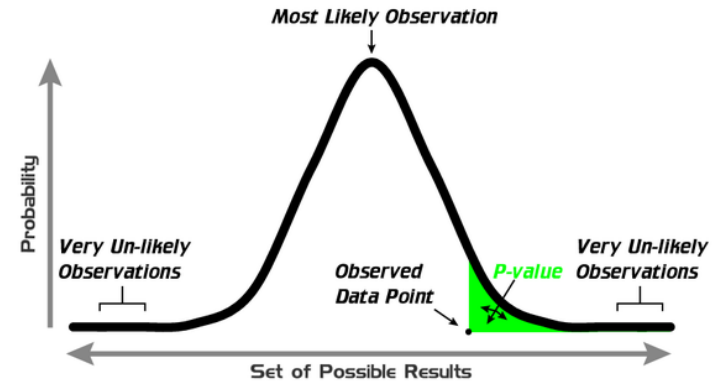
p-value - вероятность получить результат, такой как наблюдается (или выше) в предположении, что **нуль-гипотеза** верна

⇒ в нашем случае вероятность, того, что флуктуация фона достигли (или превысили) наблюдаемое значение

$$p = P(n \geq n_{\text{obs}} \mid b)$$

Нуль-гипотеза – основная проверяемая гипотеза (фон)

⇒ Нулевая гипотеза отвергается, когда значение p-value меньше уровня **стат. значимости** α (по соглашению < 0.05)



Масштабный фактор (strength factor)

$$\mu = \frac{\sigma}{\sigma_{\text{SM}}} < \mu^{95\%} \text{ at } 95\% \text{ C.L.}, \text{ e.g. } \mu^{95\%} = 1 \Rightarrow \text{exclusion}$$

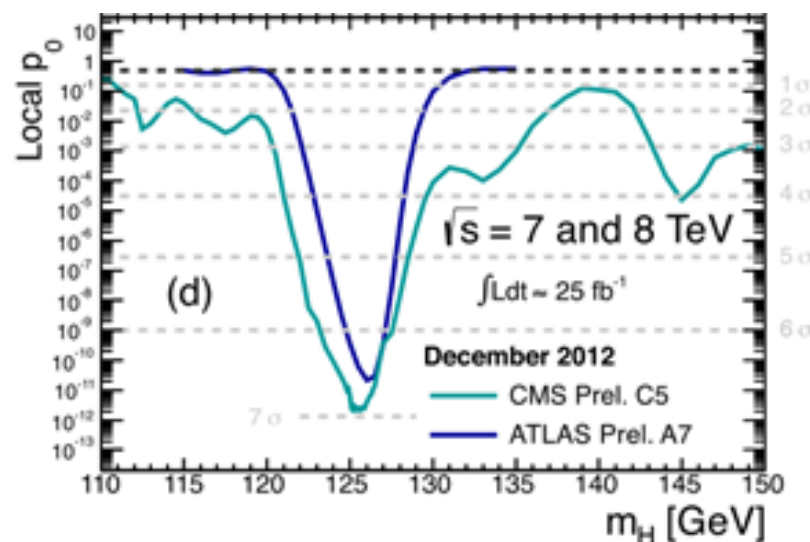
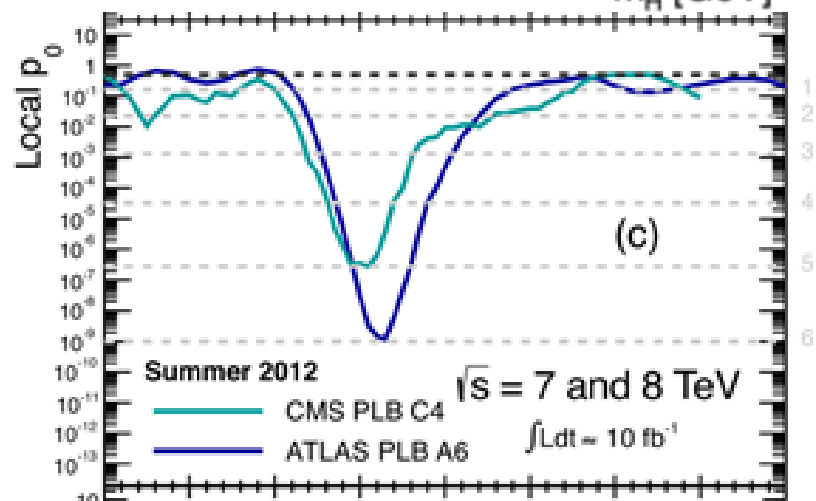
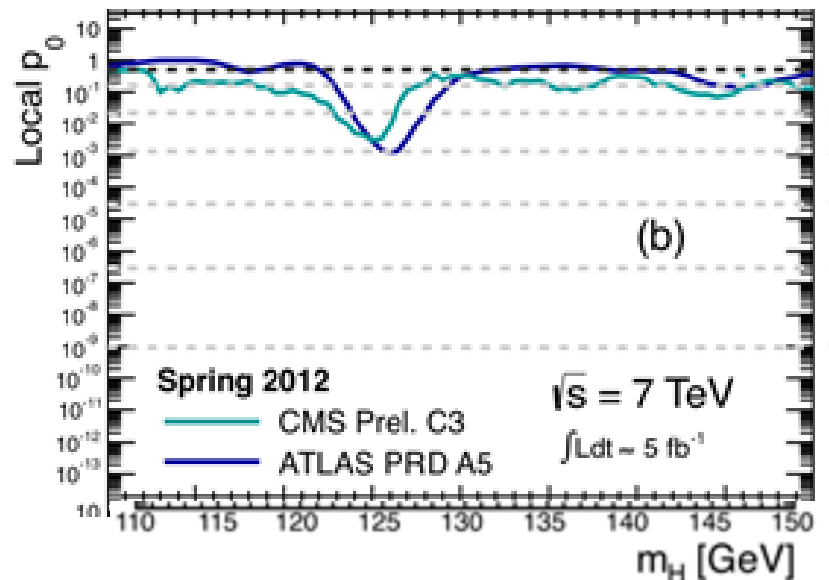
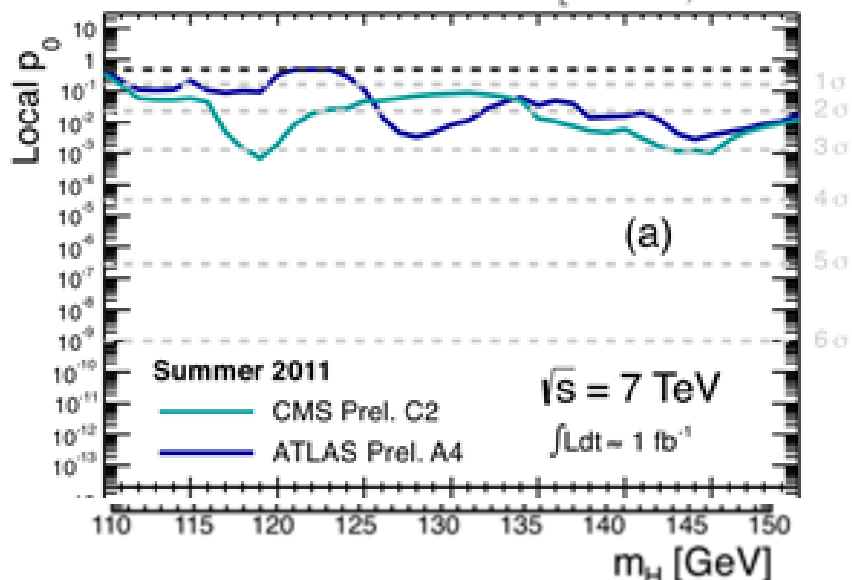
σ_{SM} – сечение бозона Хиггса в СМ, σ - гипотетическое сечение бозона Хиггса

$$\text{CL}_S(\mu^{95\%}) = \frac{\text{CL}_{S+B}}{\text{CL}_B} = \frac{P(q_\mu > q_\mu^{\text{obs}} \mid B + \mu^{95\%} \times S)}{P(q_\mu > q_\mu^{\text{obs}} \mid B)} = 0.05$$

$$q_\mu = -2 \ln \frac{\mathcal{L}(\text{data} \mid \mu, \hat{\theta}_\mu)}{\mathcal{L}(\text{data} \mid \hat{\mu}, \hat{\theta})}$$



[CERN, PDG 2013]



What does Brazilian Flag mean?



Dimuon example

$$R_\sigma = \frac{\sigma(pp \rightarrow Z' + X \rightarrow l^+l^- + X)}{\sigma(pp \rightarrow Z^0 + X \rightarrow l^+l^- + X)}$$

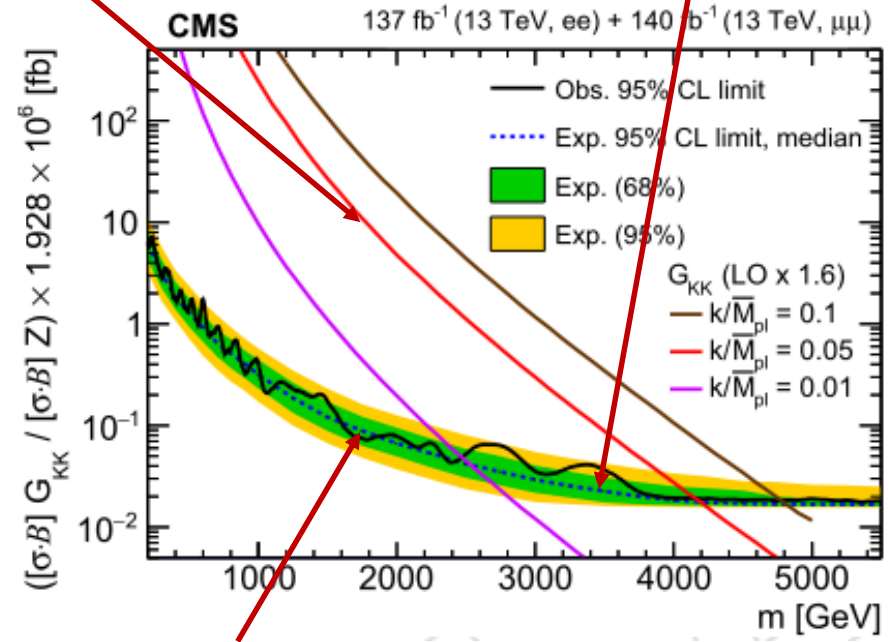
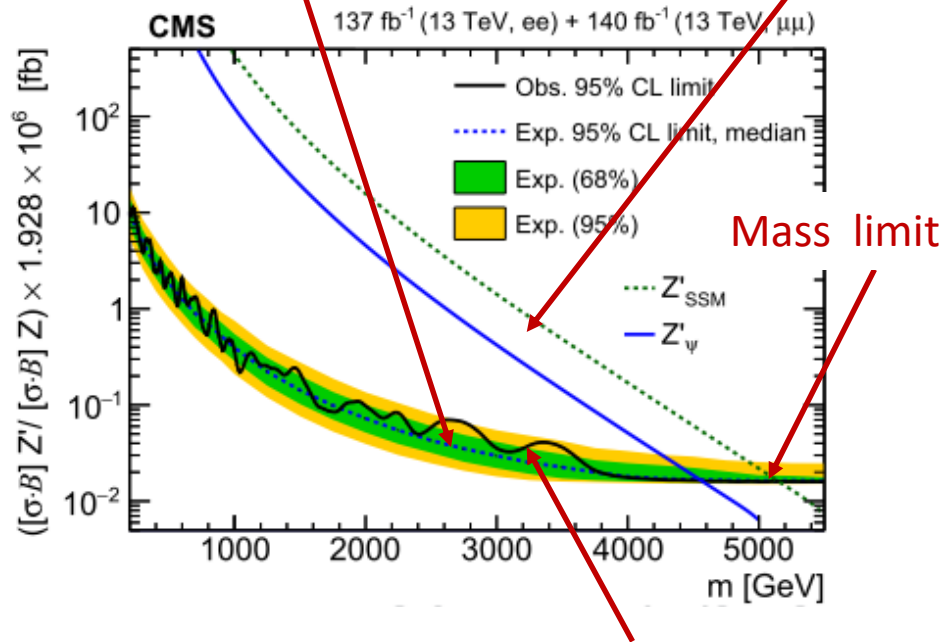
Extended gauge models

Models of low-energy gravity (RS1-type scenario of ED)

SM predictions

BSM predictions

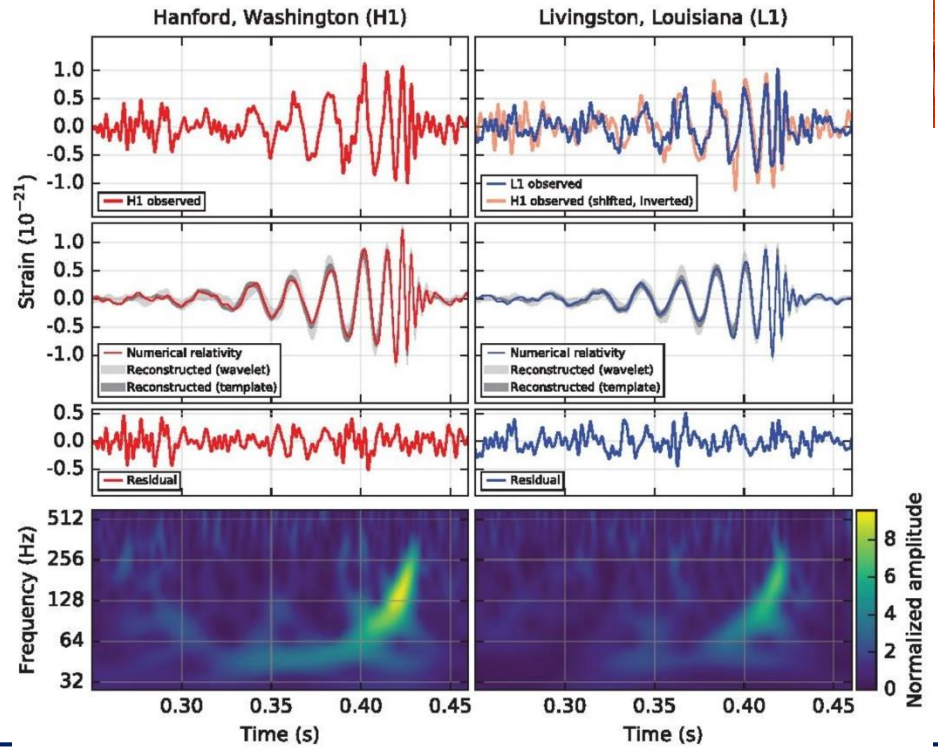
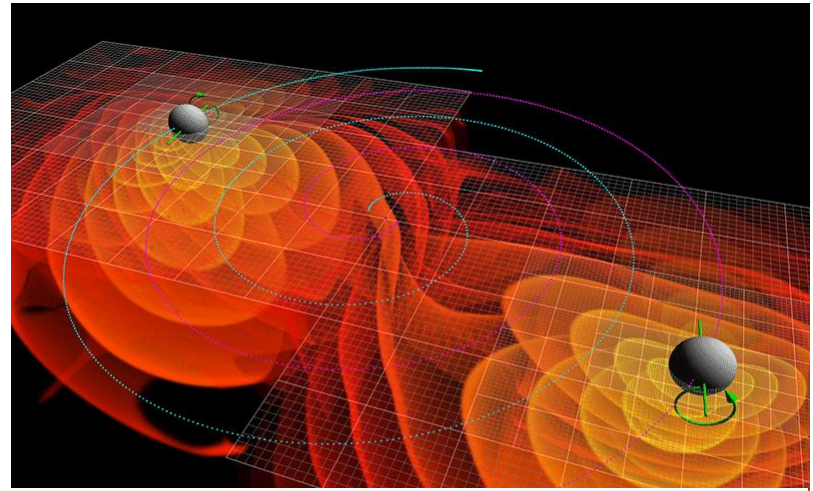
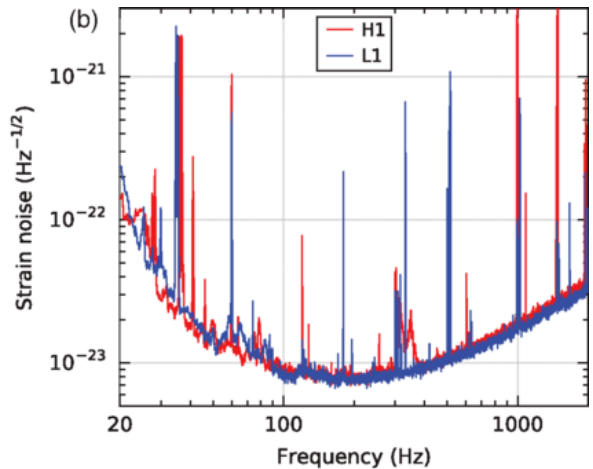
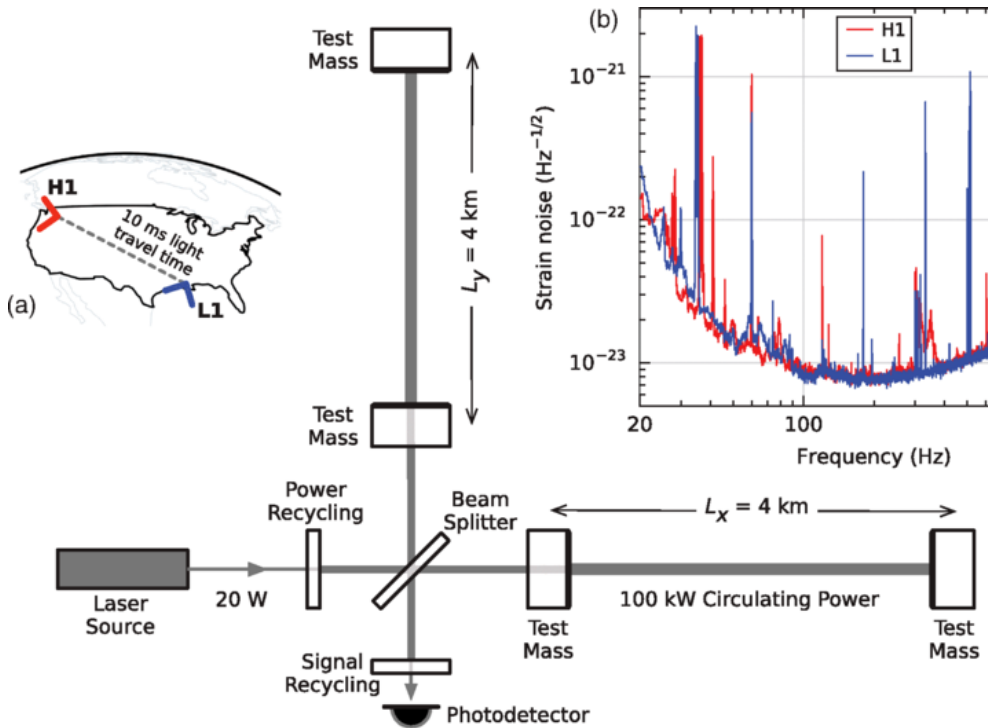
SM predictions



Model-independent limits on cross section (in narrow width approximation, NWA)

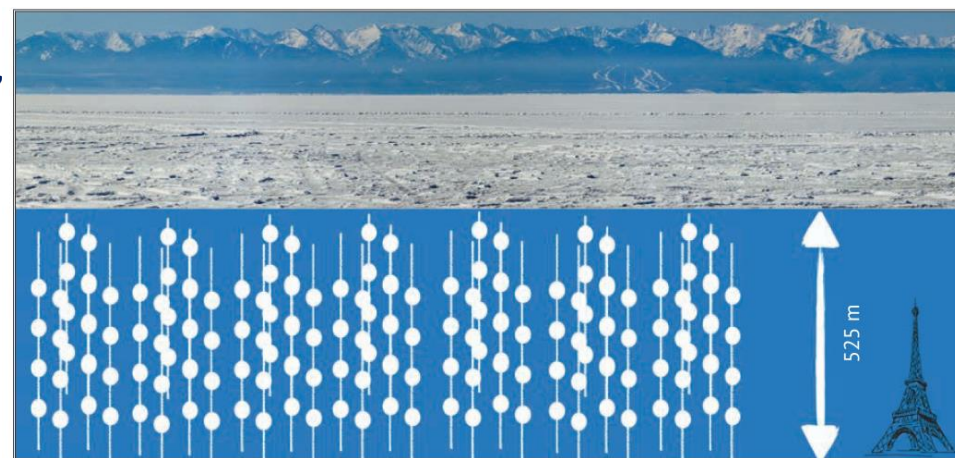
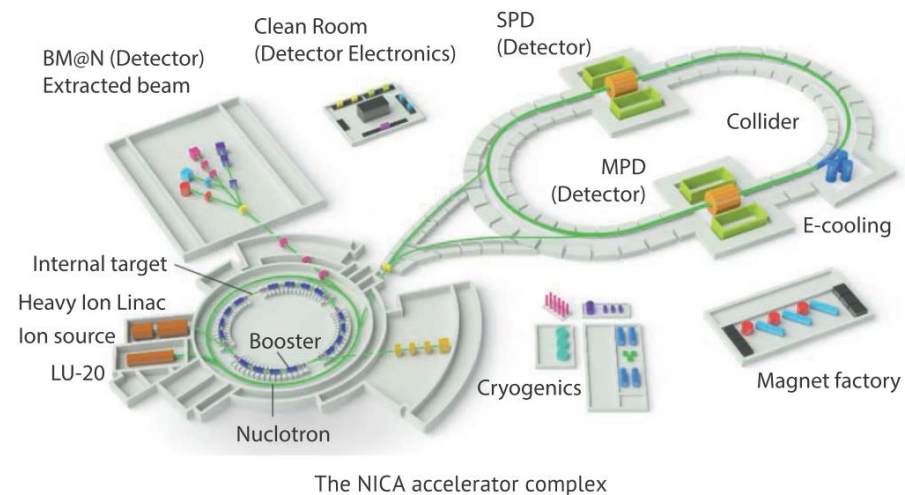
Channel	Z'_{SSM}		Z'_ψ	
	Obs. [TeV]	Exp. [TeV]	Obs. [TeV]	Exp. [TeV]
$e e$	4.72	4.72	4.11	4.13
$\mu^+ \mu^-$	4.89	4.90	4.29	4.30
$e e + \mu^+ \mu^-$	5.15	5.14	4.56	4.55

Channel	$k/\overline{M}_{Pl} = 0.01$		$k/\overline{M}_{Pl} = 0.05$		$k/\overline{M}_{Pl} = 0.1$	
	Obs. [TeV]	Exp. [TeV]	Obs. [TeV]	Exp. [TeV]	Obs. [TeV]	Exp. [TeV]
$e e$	2.16	2.29	3.70	3.83	4.42	4.43
$\mu^+ \mu^-$	2.34	2.32	3.96	3.96	4.59	4.59
$e e + \mu^+ \mu^-$	2.47	2.53	4.16	4.19	4.78	4.81

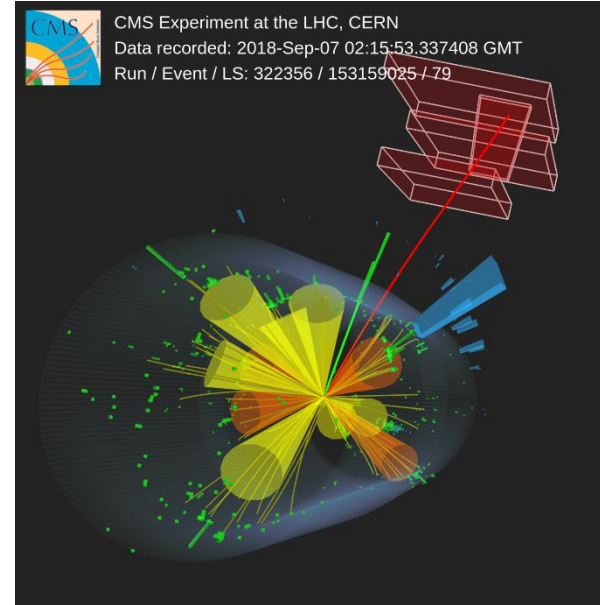


JINR LONG-TERM DEVELOPMENT STRATEGIC PLAN UP TO 2030 AND BEYOND

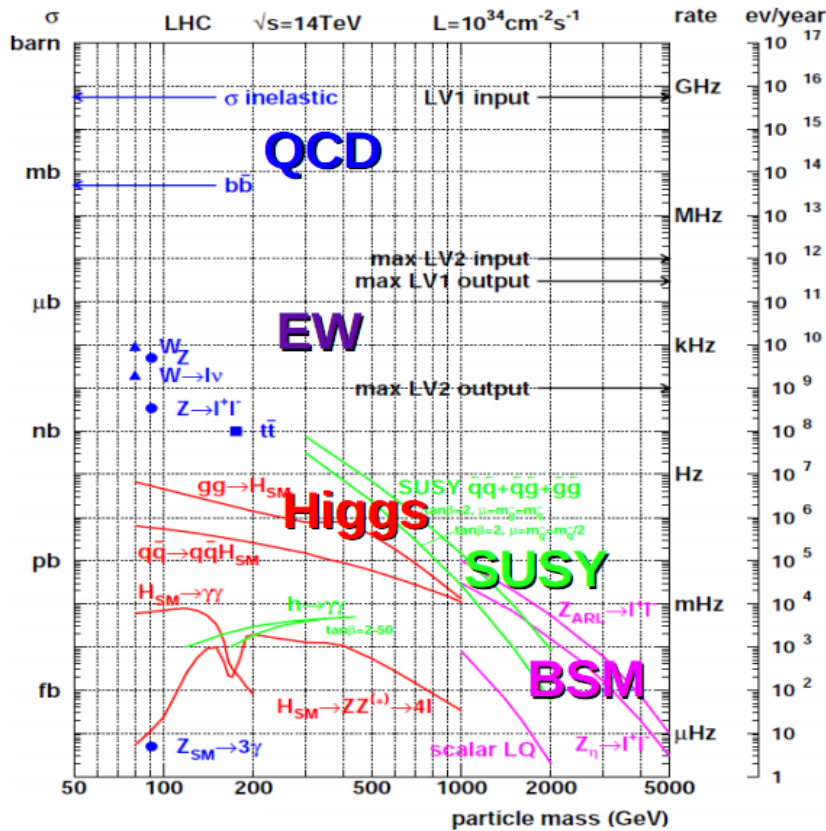
- RELATIVISTIC HEAVY-ION PHYSICS AT NICA
- JINR PARTICIPATION IN FOREFRONT EXTERNAL EXPERIMENTS OFF-SITE
 - LHC, SPS, RHIC, and at facilities under construction, as for example the FAIR facility
- NICA SPIN PHYSICS
- PARTICLE PHYSICS AT THE LHC AND BEYOND
 - Accelerator-based research and frontier accelerator technologies (LHC, SPS, NICA, FAIR, etc)
 - Neutrino physics and astroparticle physics (Baikal-GVD, JUNO, NOvA, DUNE, etc)
 - Multi-messenger astronomy including gravitational wave detection (Baikal-GVD, TAIGA, VIRGO, etc)



Baikal-GVD (Gigaton Volume Detector)



What do we know today about the Standard Model from LHC?



■ SM processes:

$$\sigma \sim 1/(100 \text{ MeV})^2$$

$10^{-8} !$

■ New Physics:

$$\sigma \sim 1/(1 \text{ TeV})^2$$

During Run 2 the LHC produced 10^{16} collisions

Large samples of various particles produced:

- W bosons: 12 billion
- Z bosons: 2.8 billion
- Top quarks: 300 million
- B quarks: 40 trillion
- Higgs bosons: 7.7 million