

HEP Software & Data Analysis

HSF-INDIA



Verena Martinez Outschoorn

University of Massachusetts—Amherst

University of
Massachusetts
Amherst

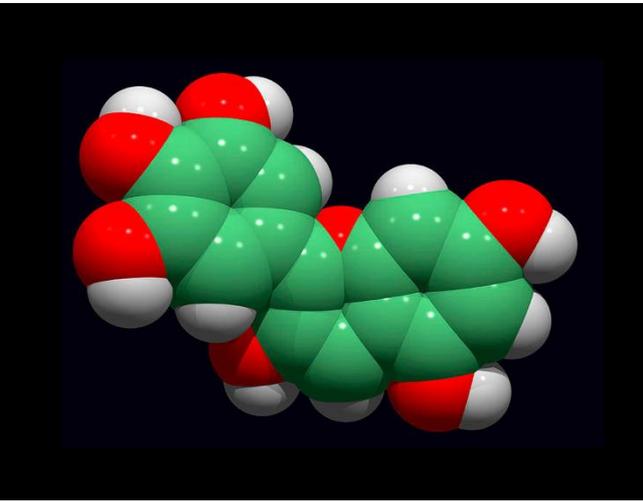
HSF-India Scientific Computing Workshop at Panjab University

15 September 2025

Particle Physics Scales

Molecule

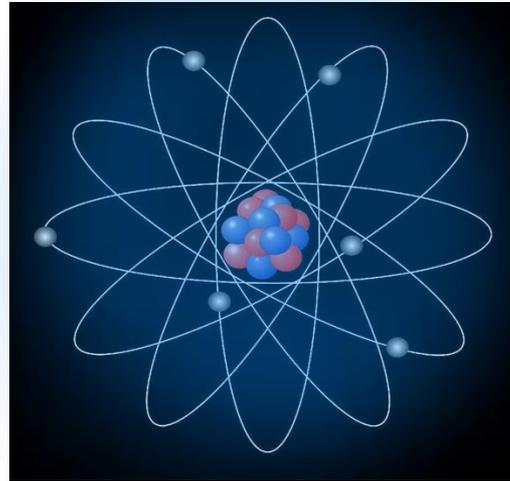
$$10^{-9} \text{ m} = 0.000\ 000\ 001 \text{ m}$$



Delphinidin Molecule
(blue pigment of flowers and grapes)

Atoms

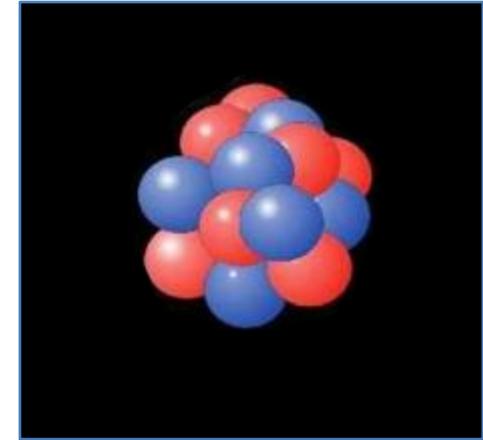
$$10^{-10} \text{ m} = 0.000\ 000\ 000\ 1 \text{ m}$$



Composed of:
Nucleus and electrons

Nucleus

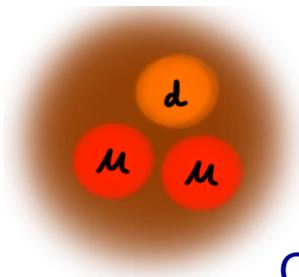
$$10^{-14} \text{ m} = 0.000\ 000\ 000\ 000\ 01 \text{ m}$$



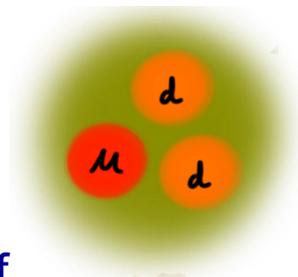
Composed of:
Protons and neutrons

Protons and Neutrons

$$10^{-15} \text{ m} = 0.000\ 000\ 000\ 000\ 001 \text{ m}$$



Composed of
quarks



Quarks

$$<10^{-18} \text{ m} = 0.000\ 000\ 000\ 000\ 000\ 001 \text{ m}$$



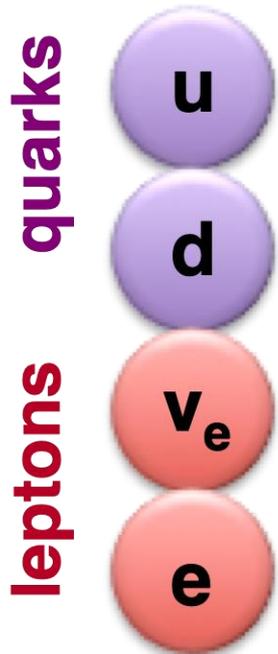
Quarks and electrons have no
dimensions

they look just like a point



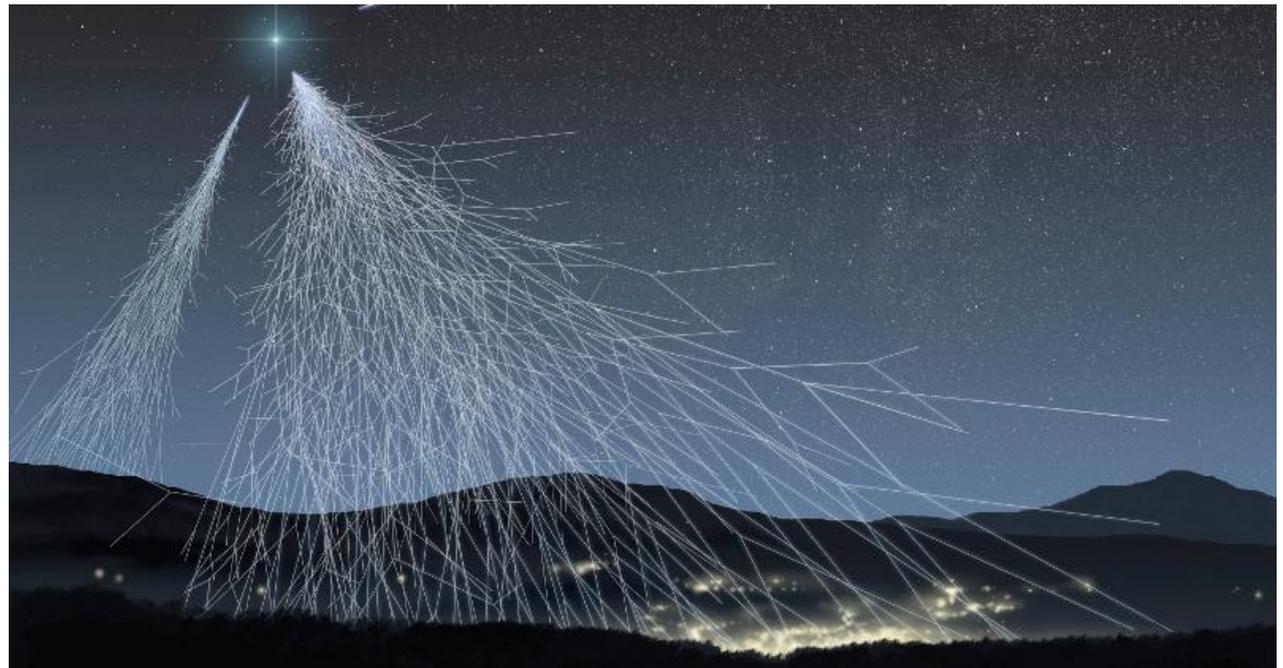
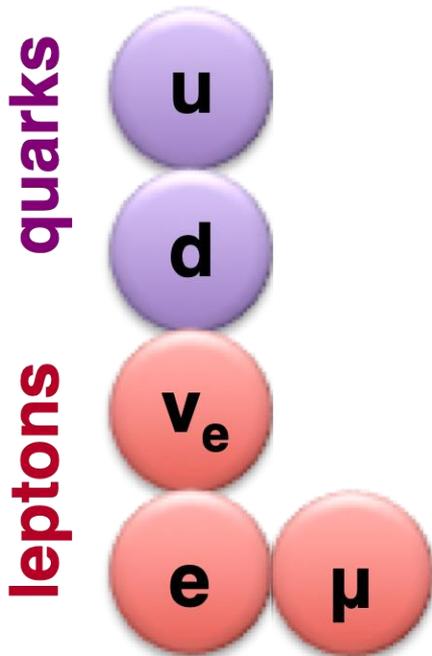
Fundamental Particles

Up and down quark, electron and electron neutrino



Fundamental Particles

1937: Discovery of the **muon** (Anderson and Neddermeyer)
a copy of the electron but with 200 times the mass ($m_{\mu} = 200 \times m_e$)



"A first surprise"

Fundamental Particles

Three complete families of fermions

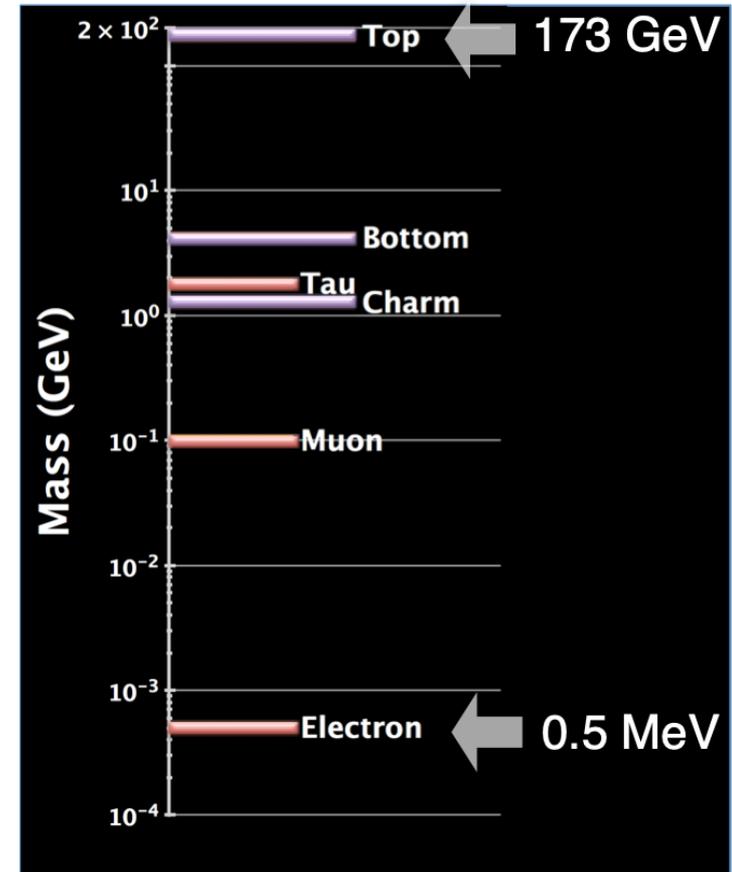
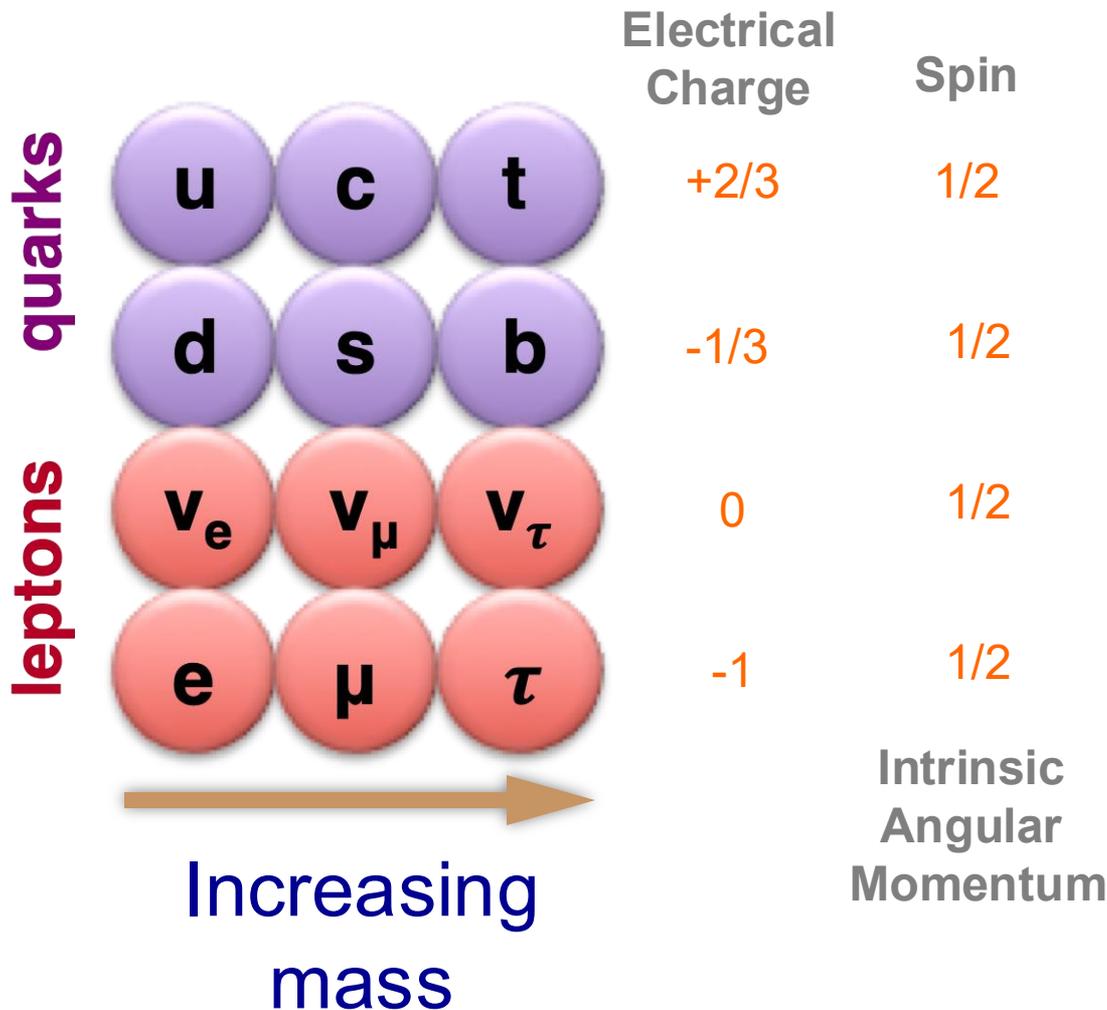
		Electrical Charge	Spin
quarks	u c t	+2/3	1/2
	d s b	-1/3	1/2
leptons	ν_e ν_μ ν_τ	0	1/2
	e μ τ	-1	1/2

Fermions matter particles

Intrinsic Angular Momentum

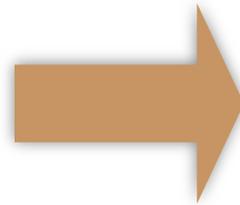
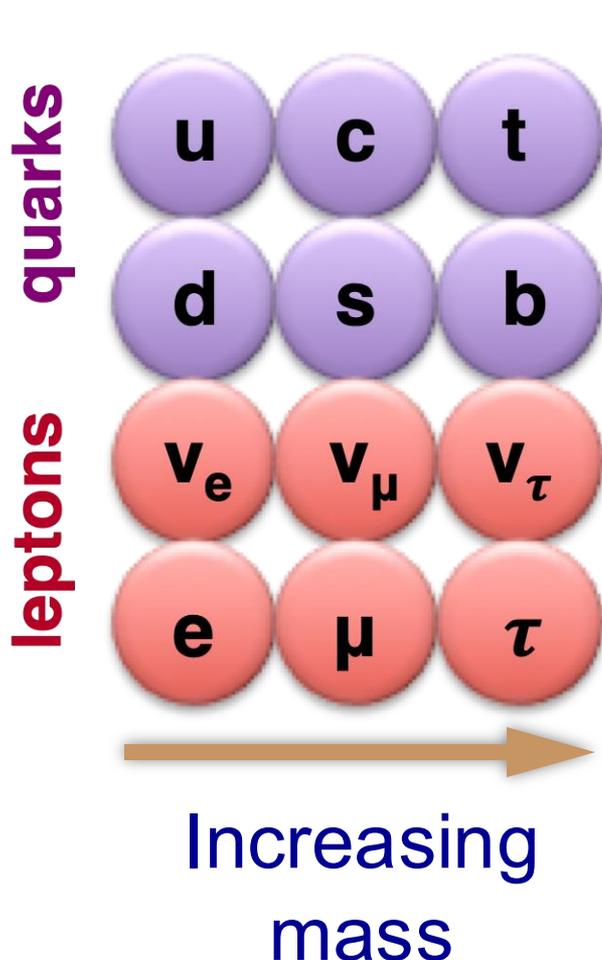
Fundamental Particles

Three complete families of fermions



Fundamental Particles

Three complete families of fermions



The Top quark

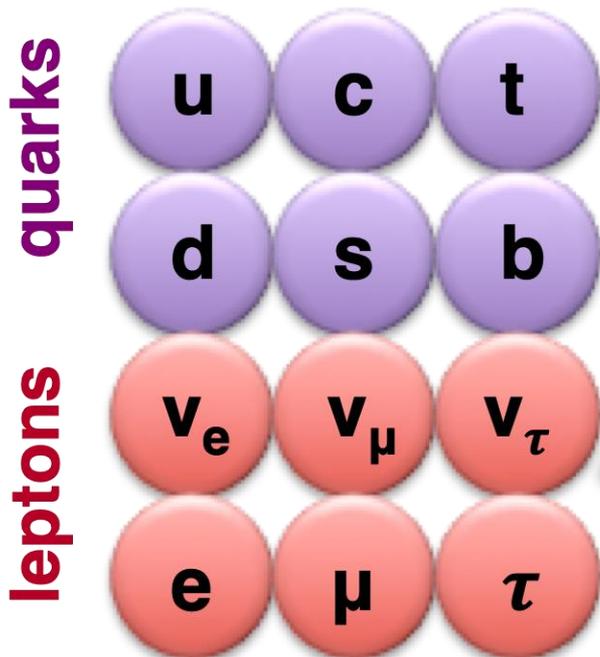
1995: Discovered by CDF and D0 experiments at Fermilab, Chicago

$m_{\text{top}} = 175 \text{ GeV}$
Same mass as a
Tungsten atom (W)

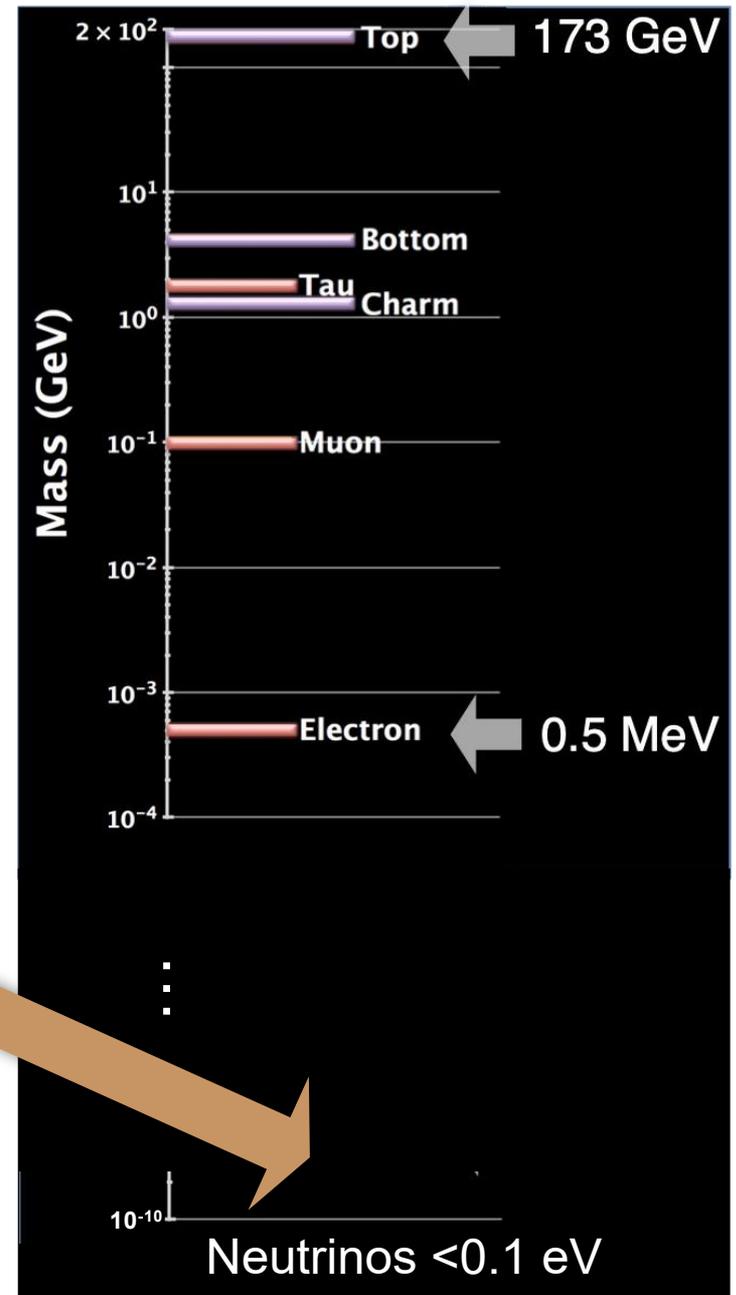
74 electrons
74 protons
108 neutrons

Fundamental Particles

Neutrinos are massive
but extremely light!



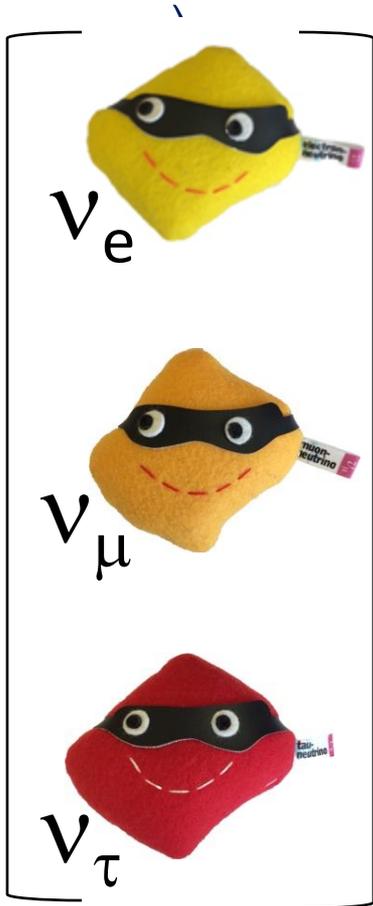
Dedicated neutrino experiments
to study these properties



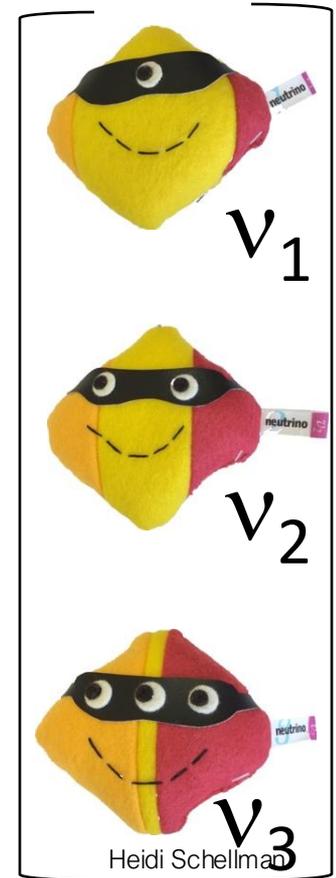
Neutrino Interactions and Mass

Two different views of the same neutrinos

Flavor Basis
(Interactions)



Mass Basis
(Motion)

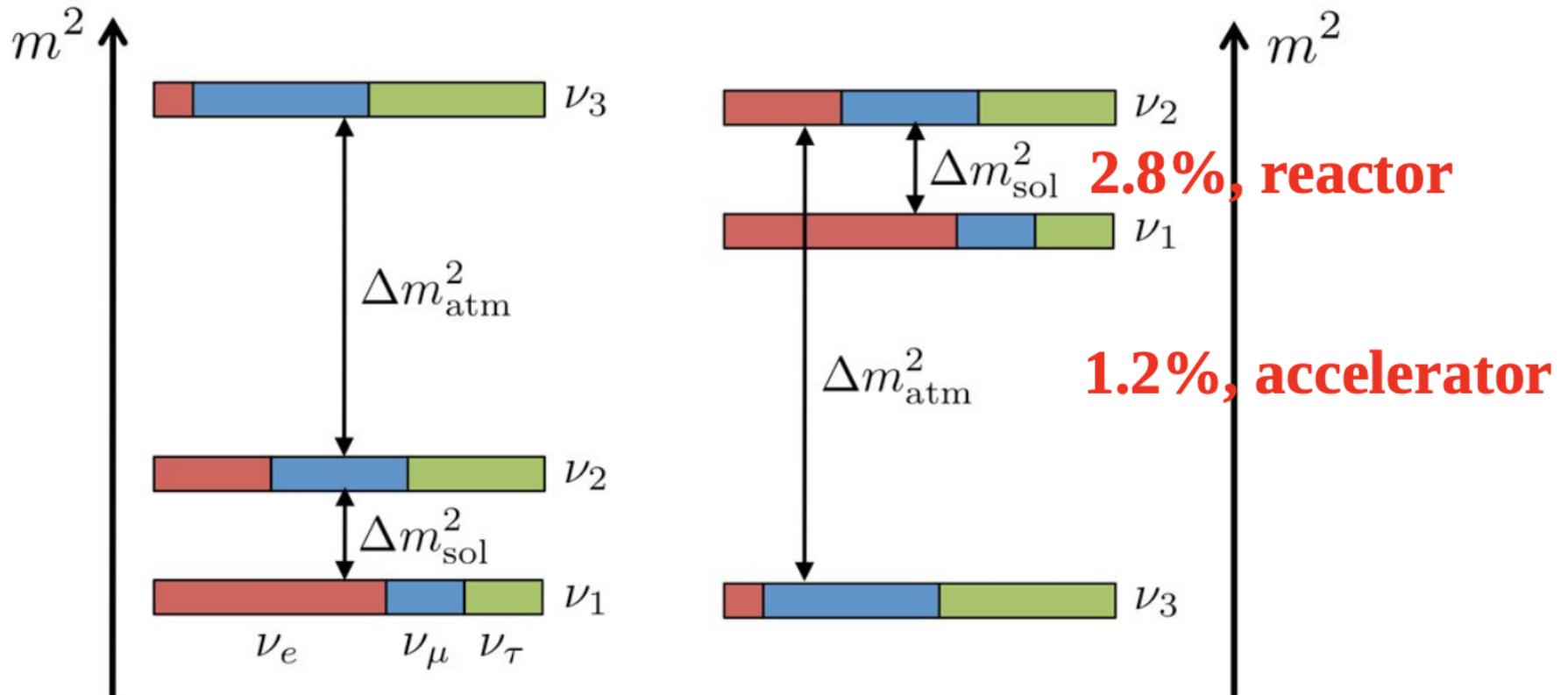


$$\begin{aligned}
 &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{bmatrix} \begin{bmatrix} c_{13} & 0 & s_{13}e^{-i\delta_{CP}} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta_{CP}} & 0 & c_{13} \end{bmatrix} \begin{bmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
 &= \begin{bmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{CP}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta_{CP}} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta_{CP}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta_{CP}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta_{CP}} & c_{23}c_{13} \end{bmatrix}
 \end{aligned}$$

PMNS (Pontecorvo–Maki–Nakagawa–Sakata) matrix

A major goal for experiments such as DUNE is the study of neutrino interactions

Neutrino Interactions and Mass



Mass ordering unknown

Heidi Schellman

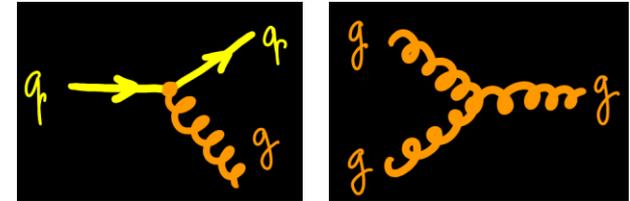
A major goal for experiments such as DUNE
is the study of neutrino interactions

Fundamental Particles

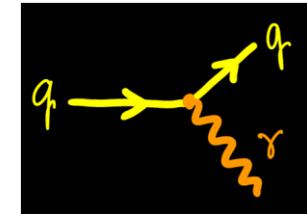
Last particles discovered by 1983

	quarks				Electrical Charge	Spin
quarks	u	c	t	g	0	1
	d	s	b	γ	0	1
leptons	ν_e	ν_μ	ν_τ	Z^0	0	1
	e	μ	τ	W^\pm	± 1	1
	gauge bosons					

Strong force (gluons)



Electromagnetic force (photon)



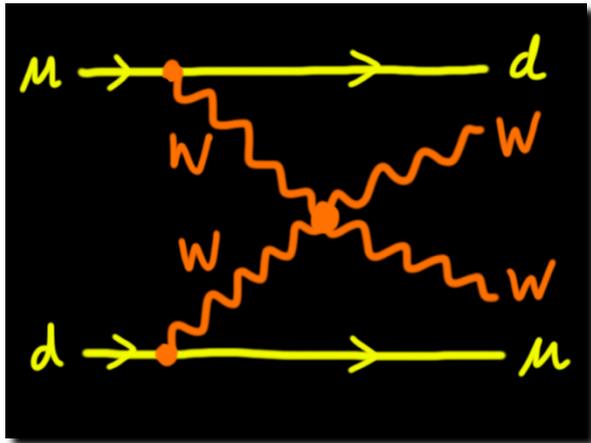
Weak force (W and Z bosons)

Beta decay: $n \rightarrow p e^- \nu_e$

Intrinsic Angular Momentum

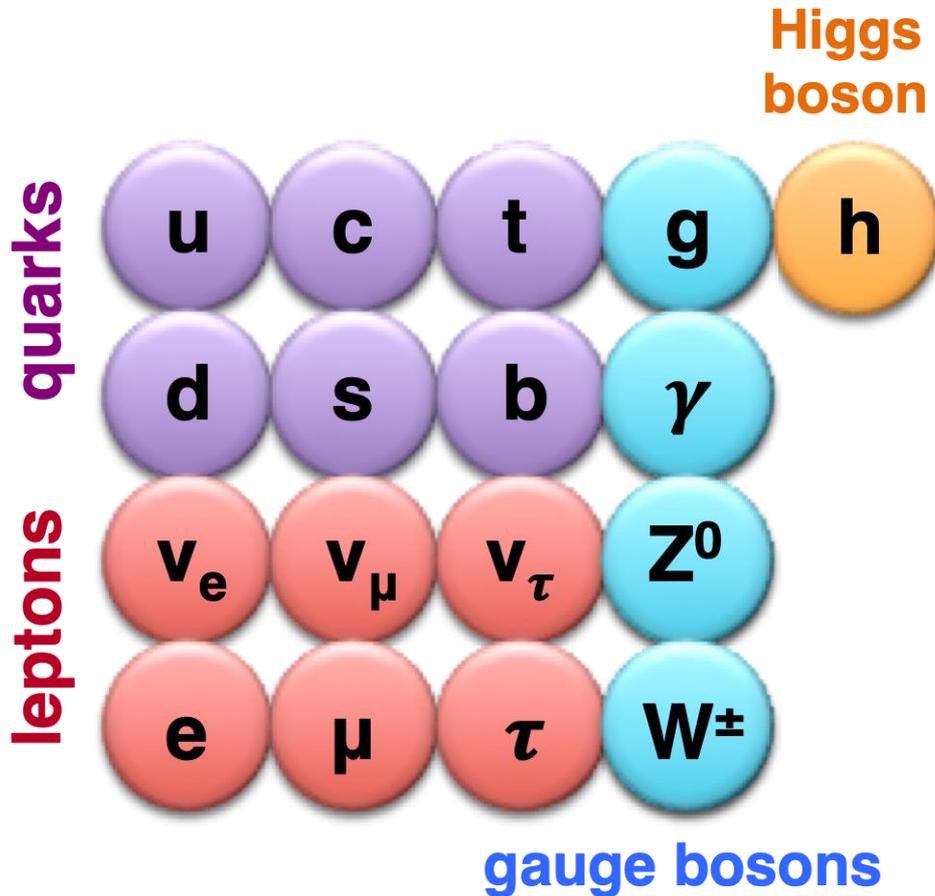
The Weak Force

The weak nuclear force has a very small range (10^{-18} m)
→ force carriers (W and Z boson) have to be massive



It is impossible to build a consistent theory for massive bosons like the W and Z without an additional particle.

Fundamental Particles



Electrical
Charge

0

Spin

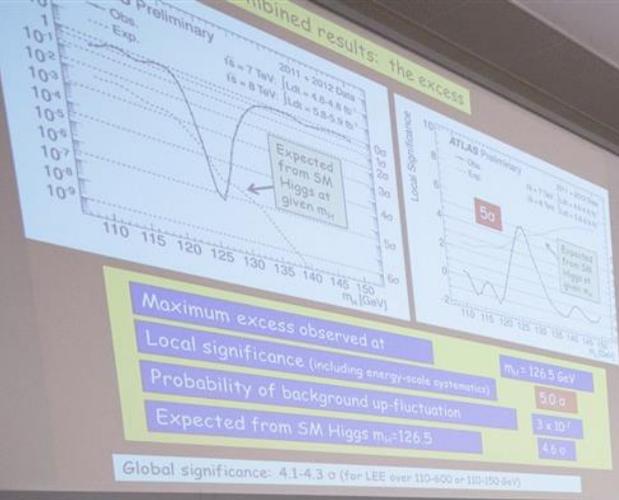
0

Intrinsic
Angular
Momentum

Discovered in 2012
after 40 years!

Higgs Particle Discovery Announcement July 4th, 2012

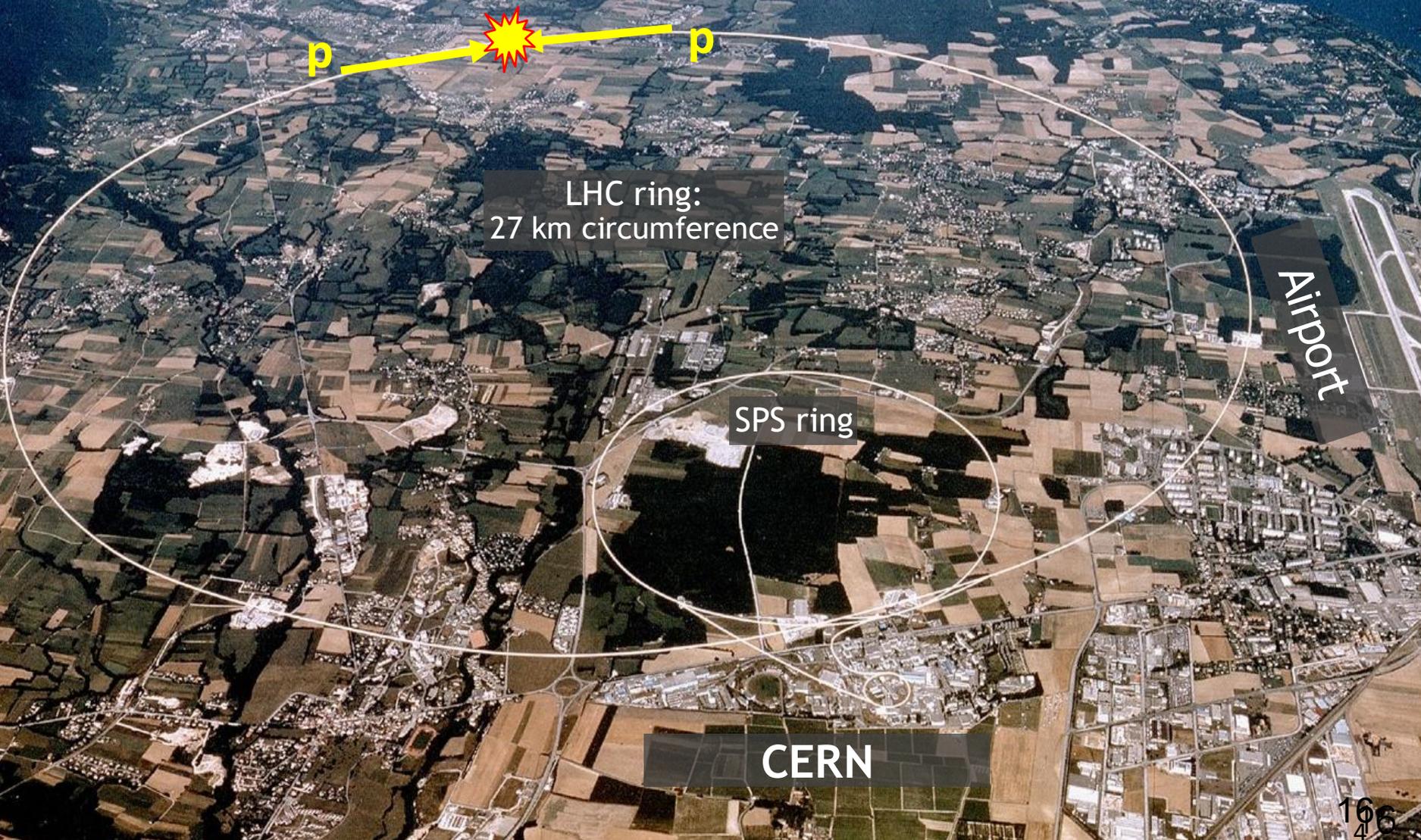
ICHEP, Melbourne



CERN, Geneva



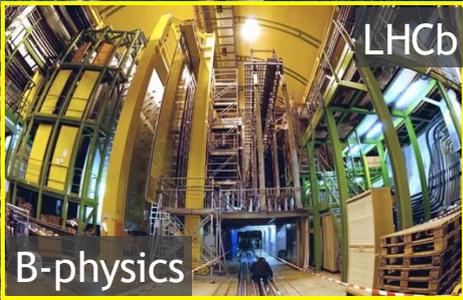
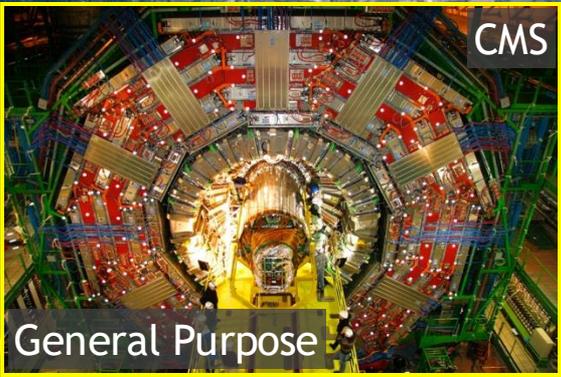
Large Hadron Collider
proton-proton collisions
Center of mass energy: 7-8-13-13.6-14 TeV



Airport

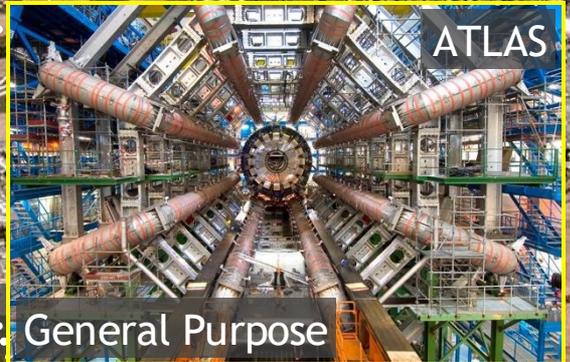
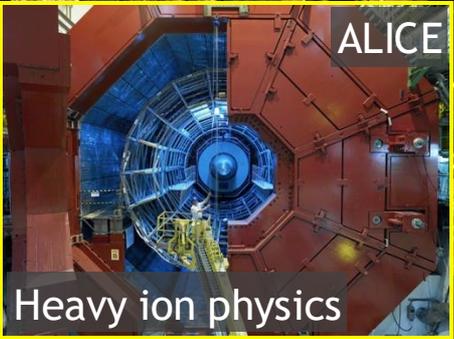
CERN

Large Hadron Collider
proton-proton collisions
Center of mass energy: 7-8-13-13.6-14 TeV



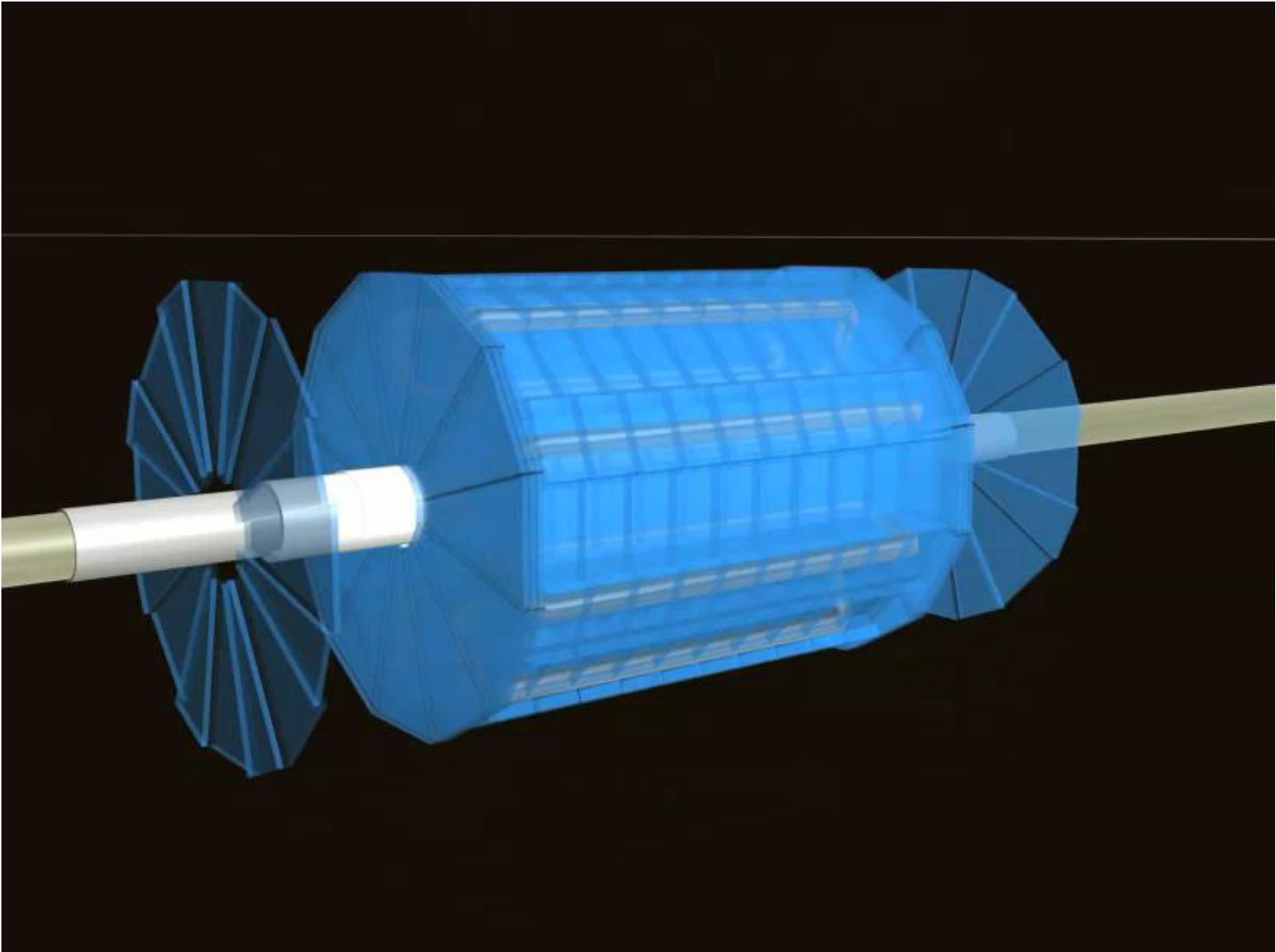
LHC ring:
27 km circumference

SPS ring



CERN

Particle Detection in ATLAS



Direction → Tracking

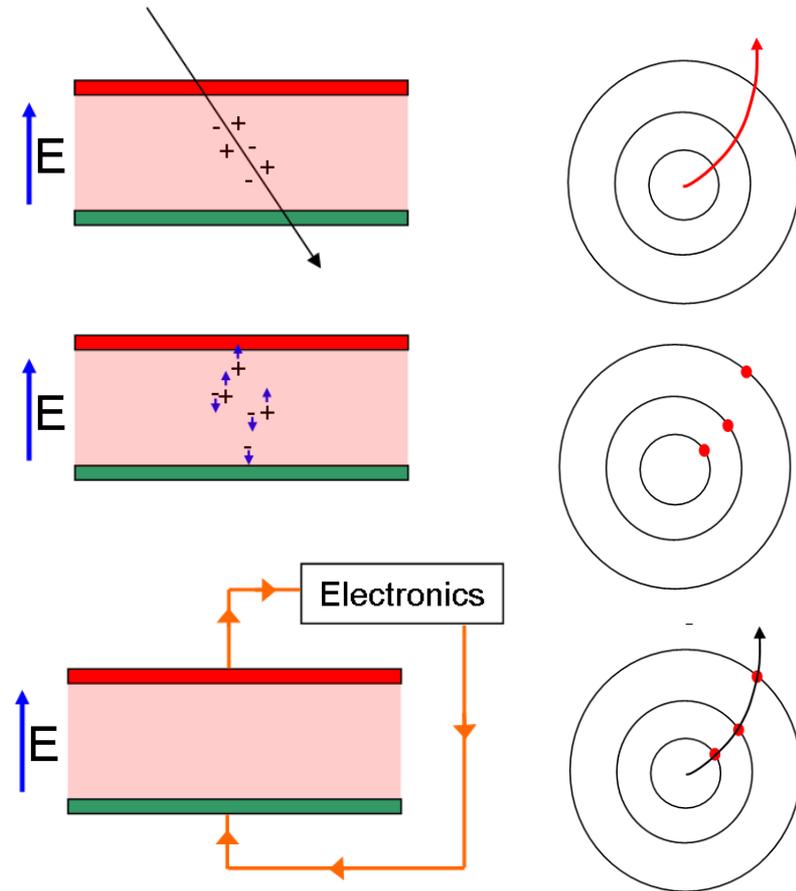
- Charged particles pass through detecting medium and knock out electrons

 - Gas, Silicon

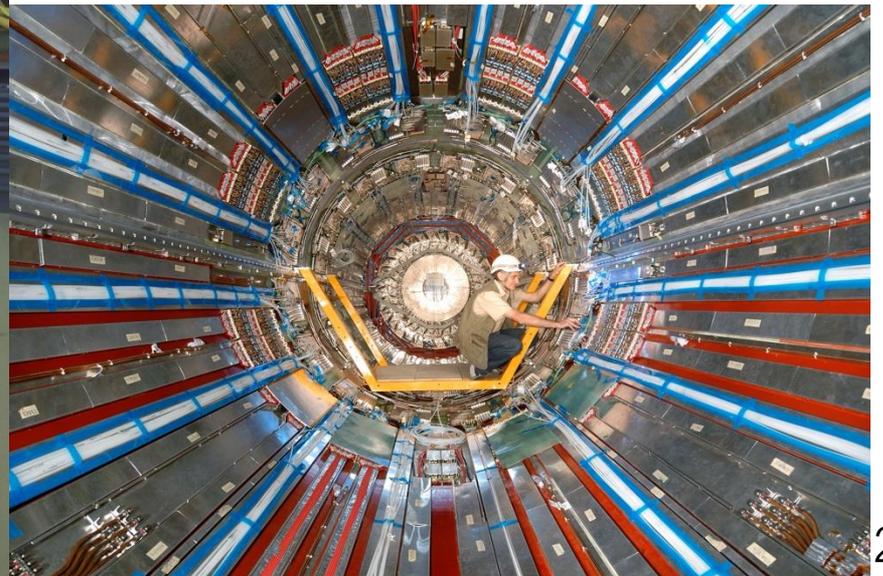
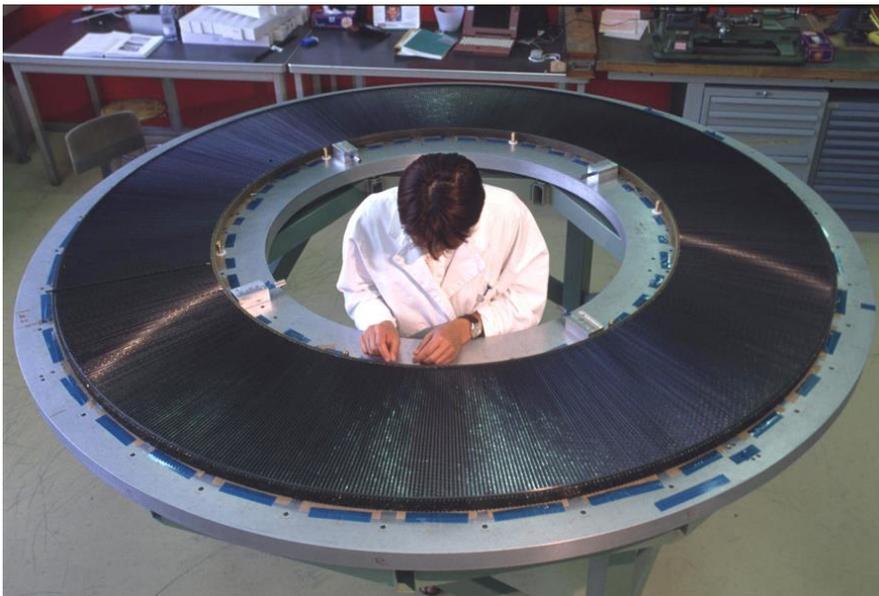
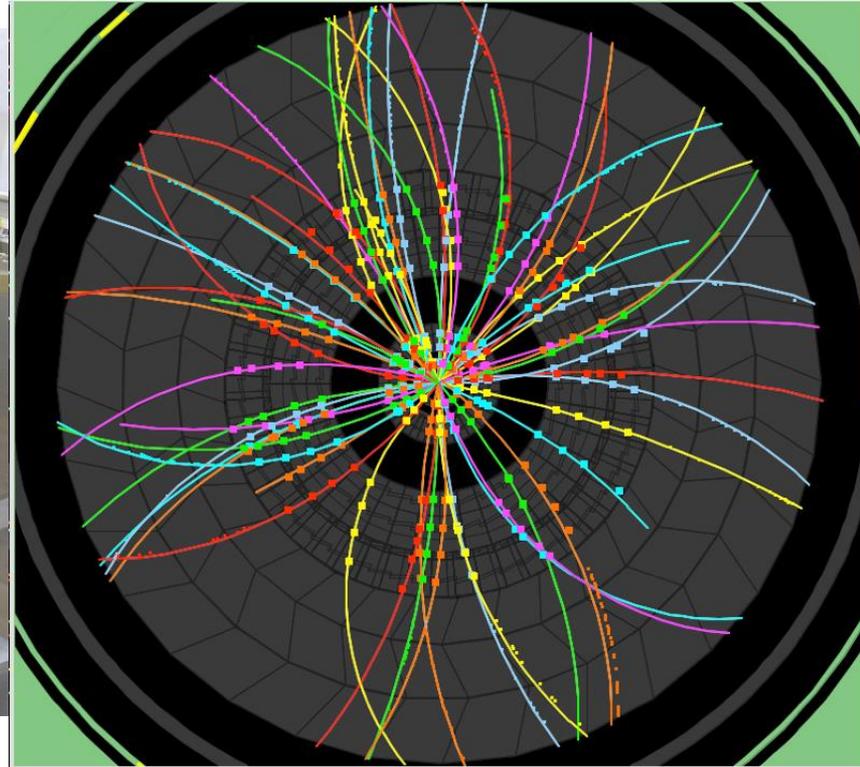
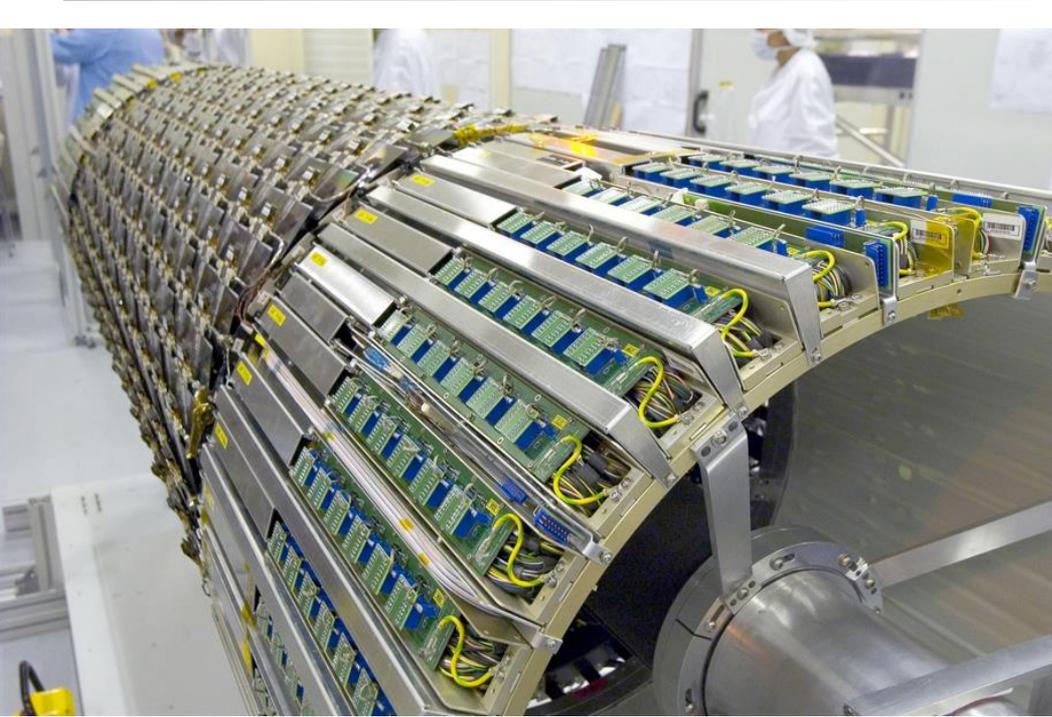
- Released electrons are collected and read out as hits

- Reconstruct trajectory out of hits

- Usually in a magnetic field so momentum can be determined by curvature



Trackers in ATLAS



Energy → Calorimetry

- Calorimeters measure total energy of particles

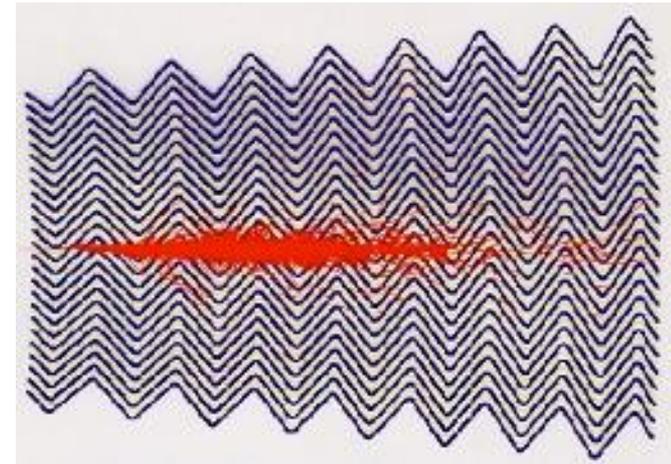
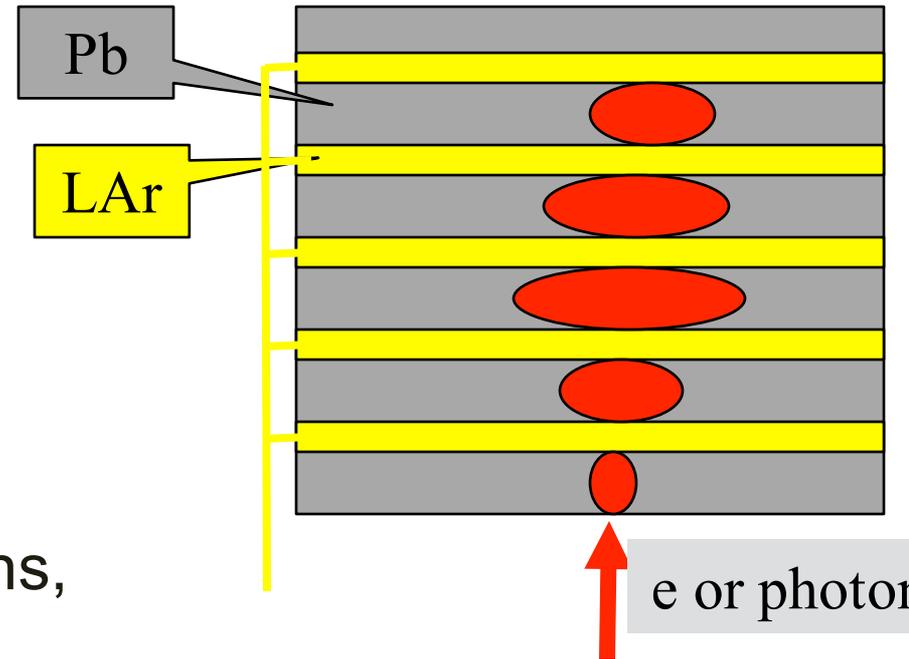
- electrons, photons, jets

- Dense material causes particles to interact

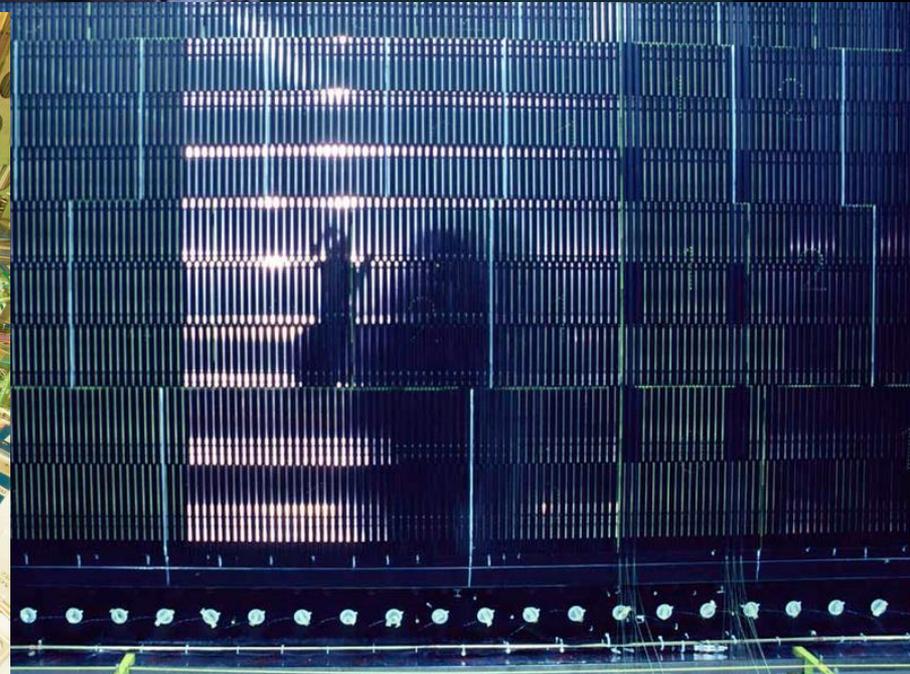
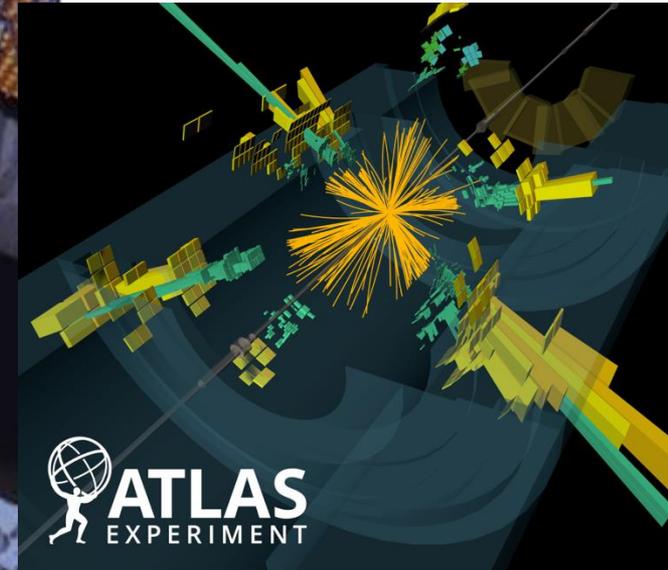
- Lose energy to ionization and nuclear interactions
 - Create cascade of electrons, photons

- Sensitive or active material

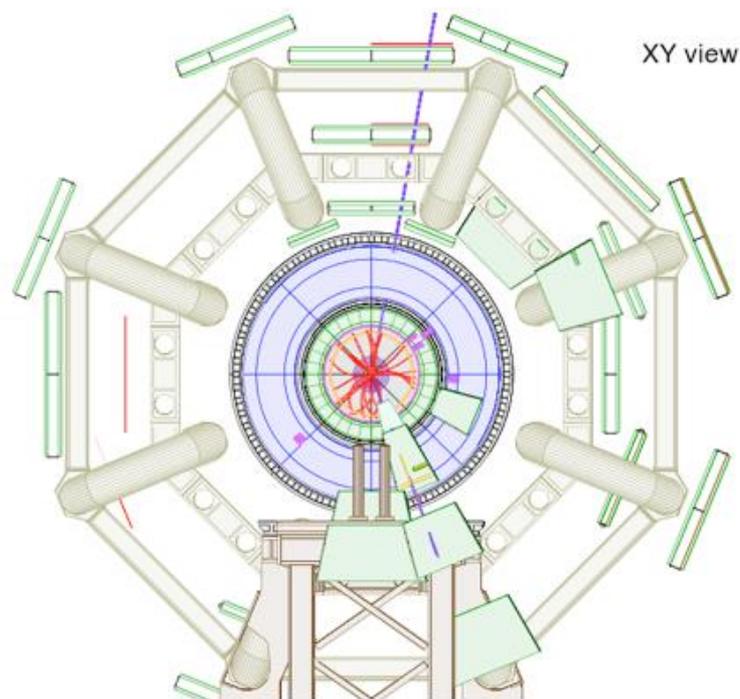
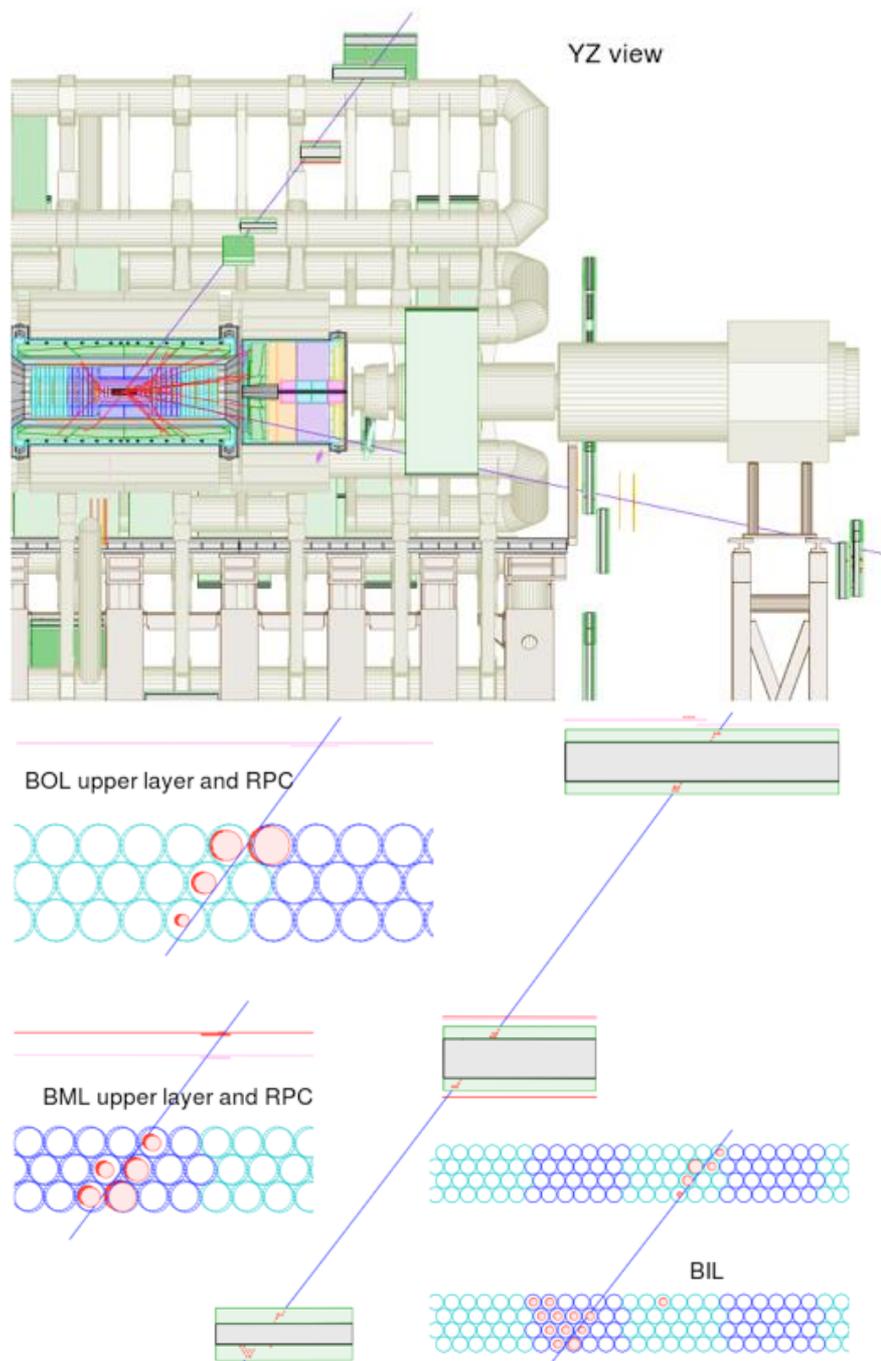
- Ionizes the material and charge is collected (e.g. LAr)
 - Excitation & scintillation processes can also be used



Calorimeters in ATLAS

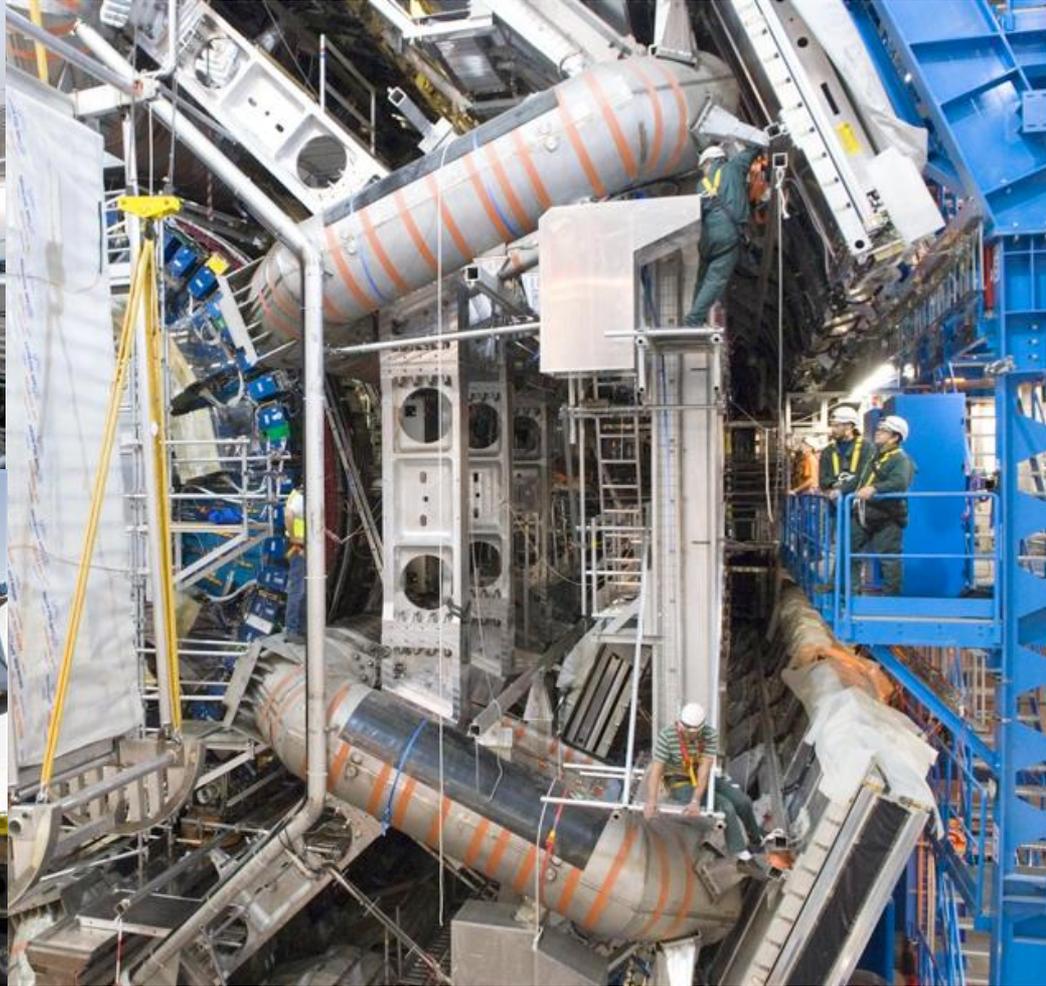
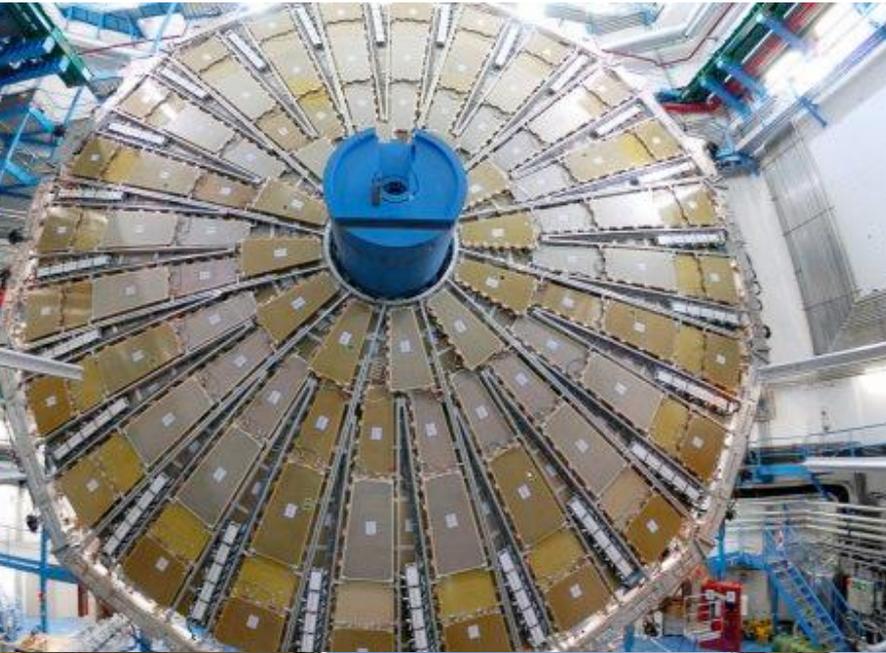


Muon Tracking



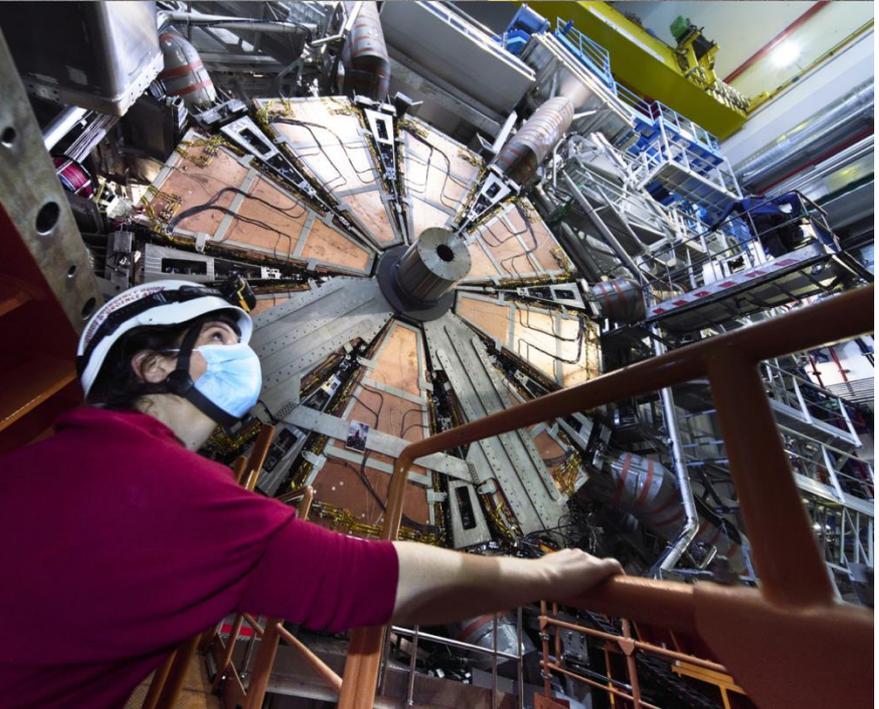
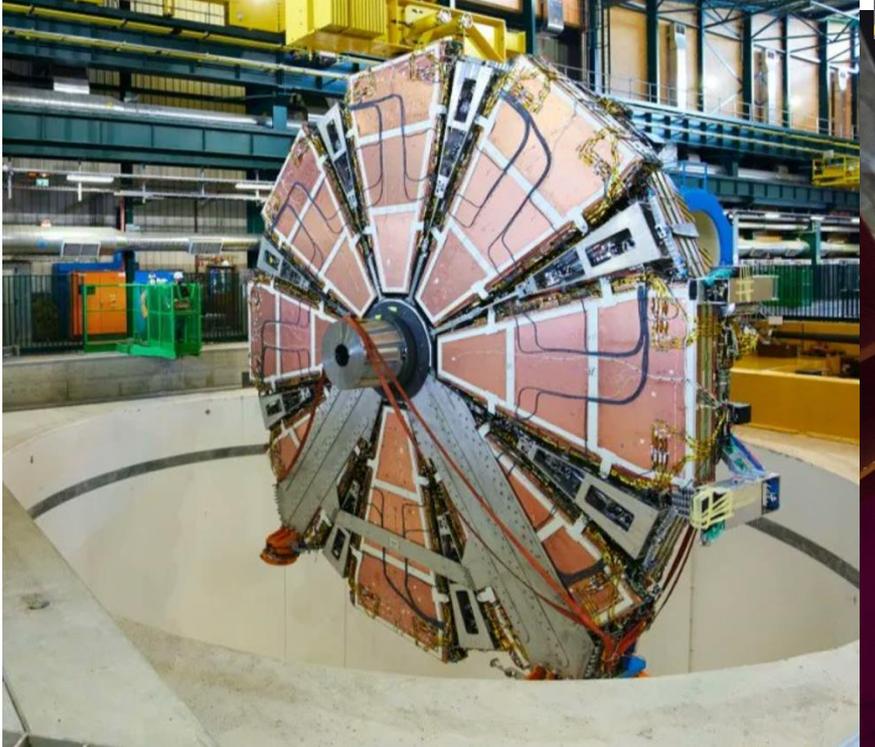
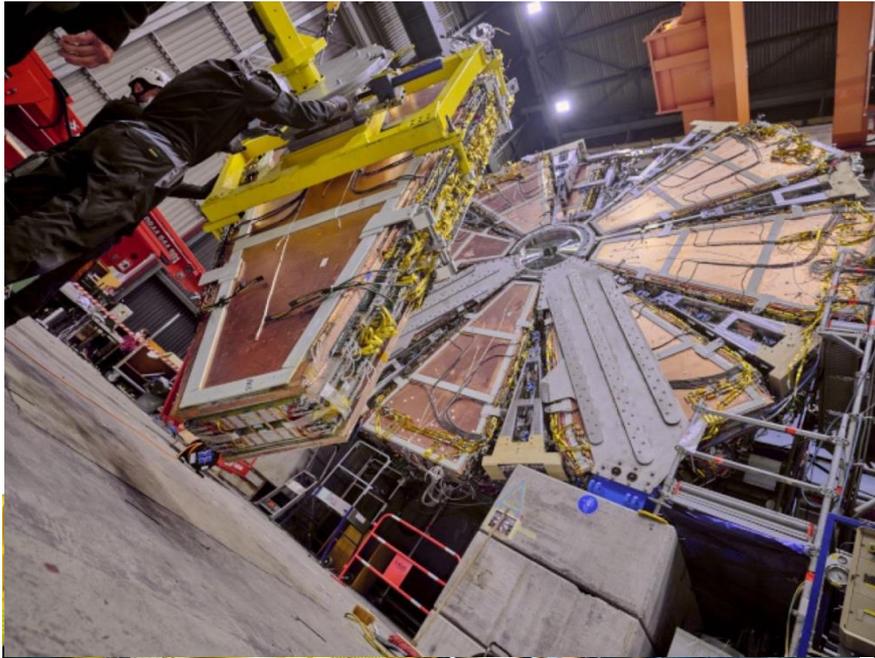
- Muons escape full detector → only other particle is neutrino
- Use tracking detectors that cover large areas far away from collision region to identify muons

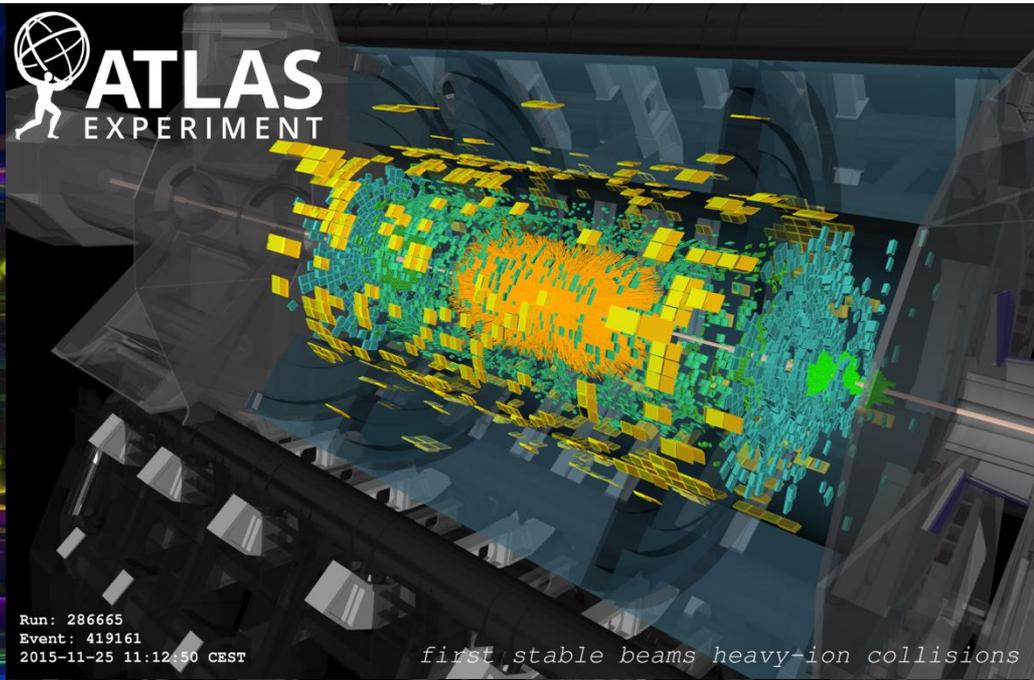
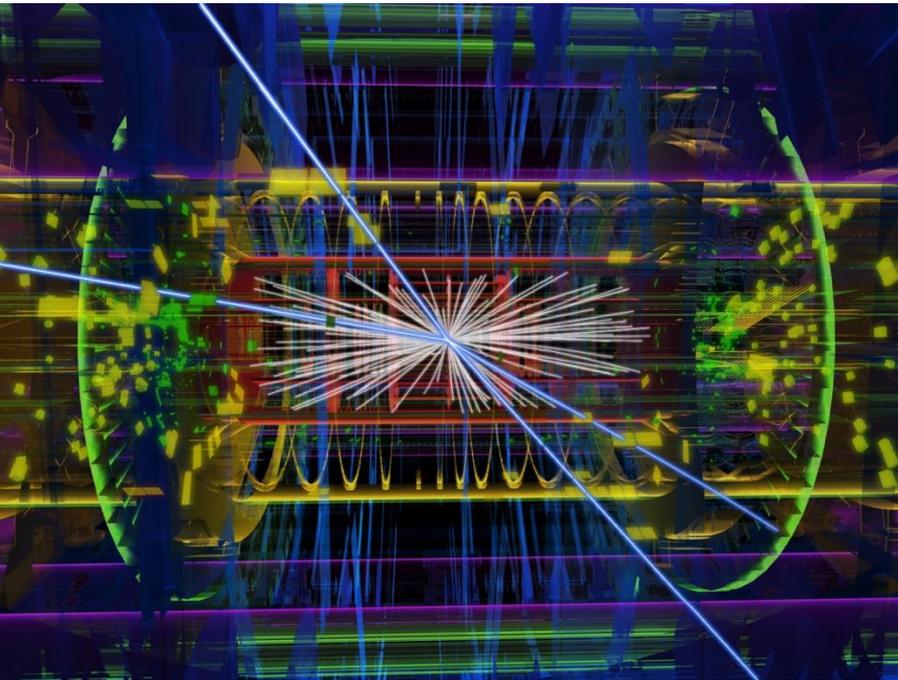
Muon System in ATLAS



Installation completed in 2008

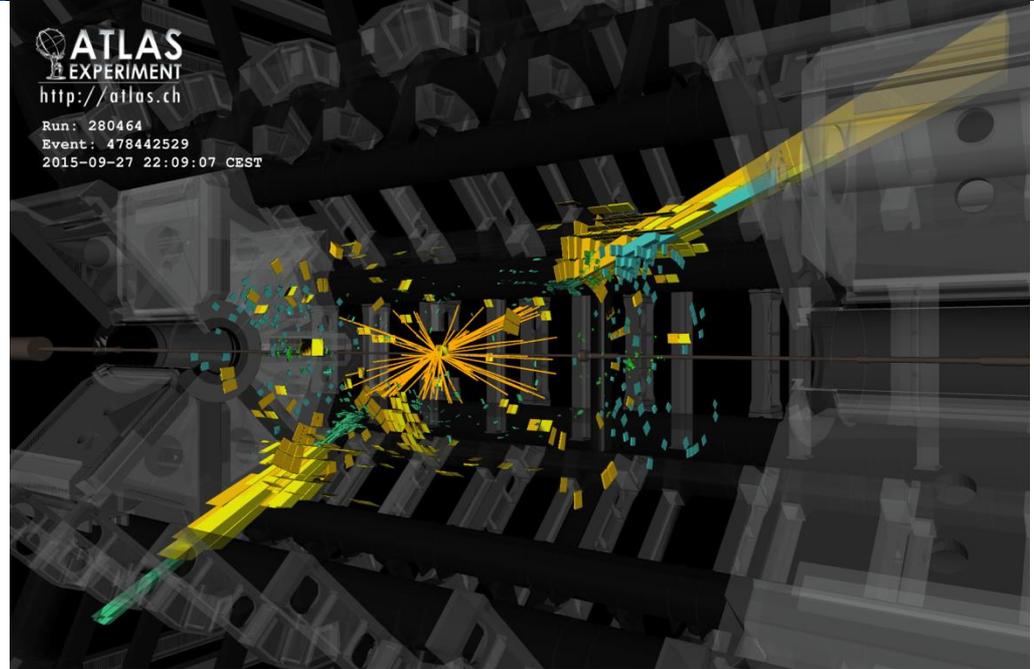
Muon System in ATLAS – Upgrade in 2021





40 million per second

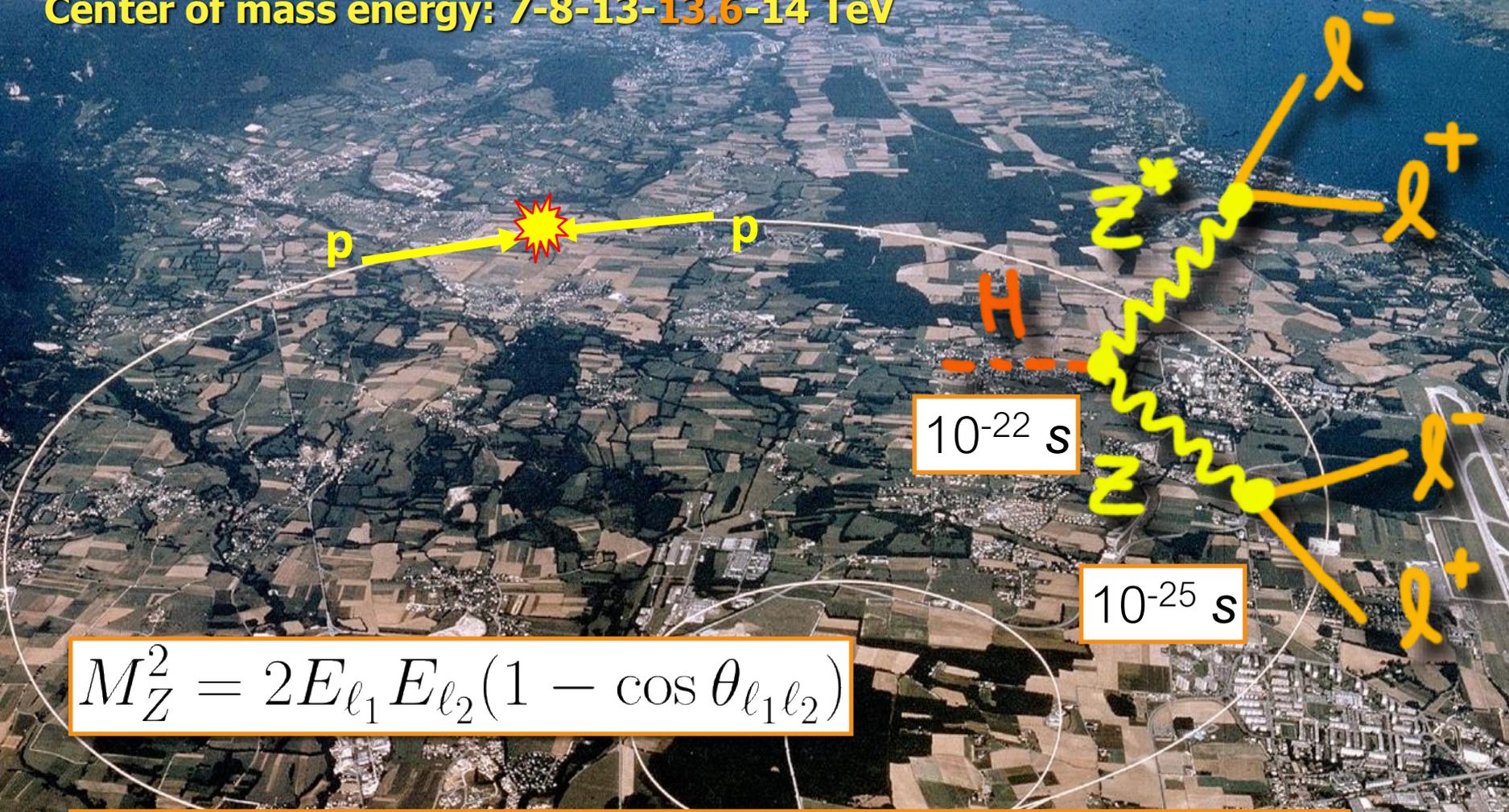
~1000 per second
stored for
analysis



Large Hadron Collider

proton-proton collisions

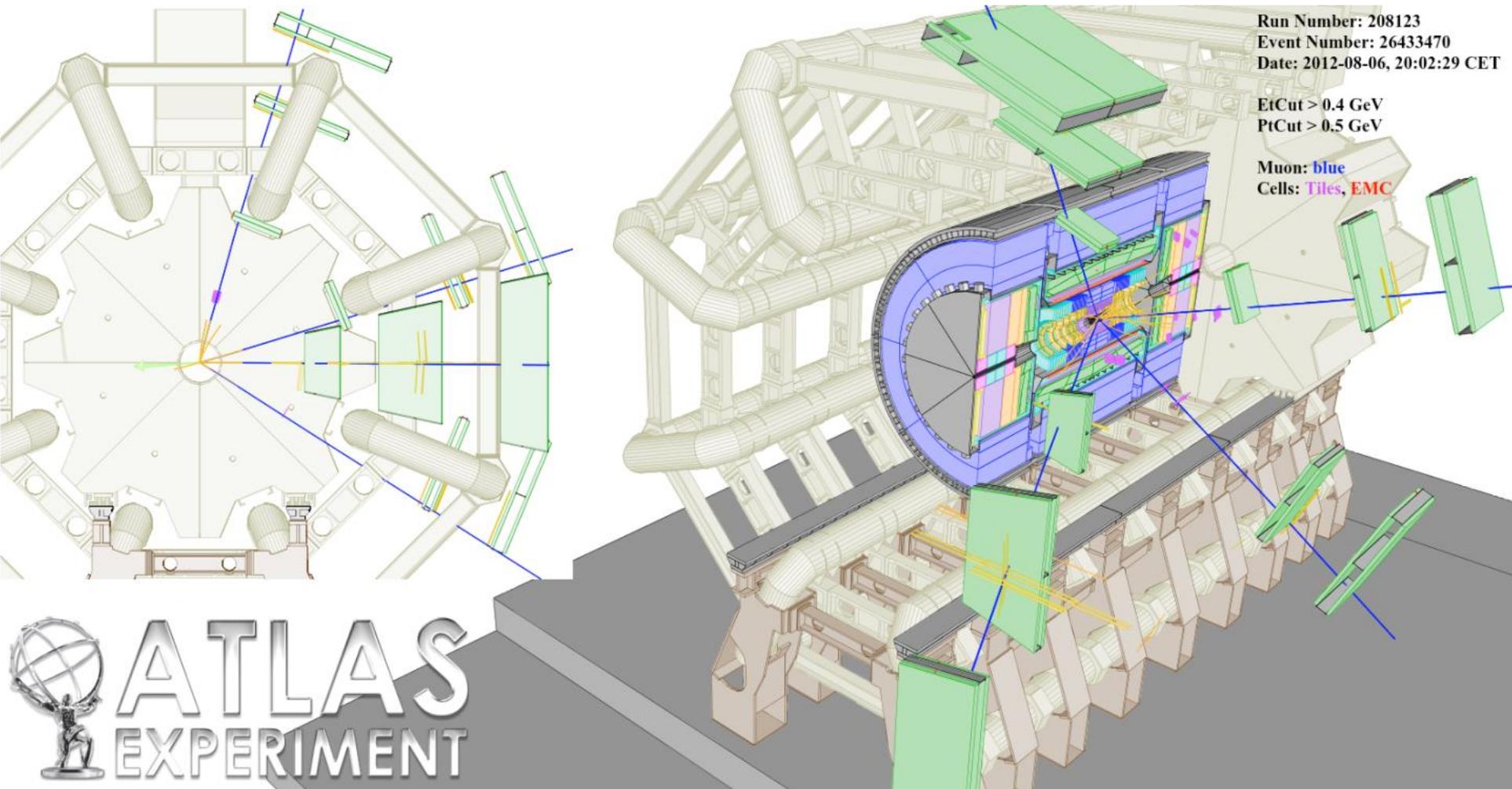
Center of mass energy: 7-8-13-13.6-14 TeV



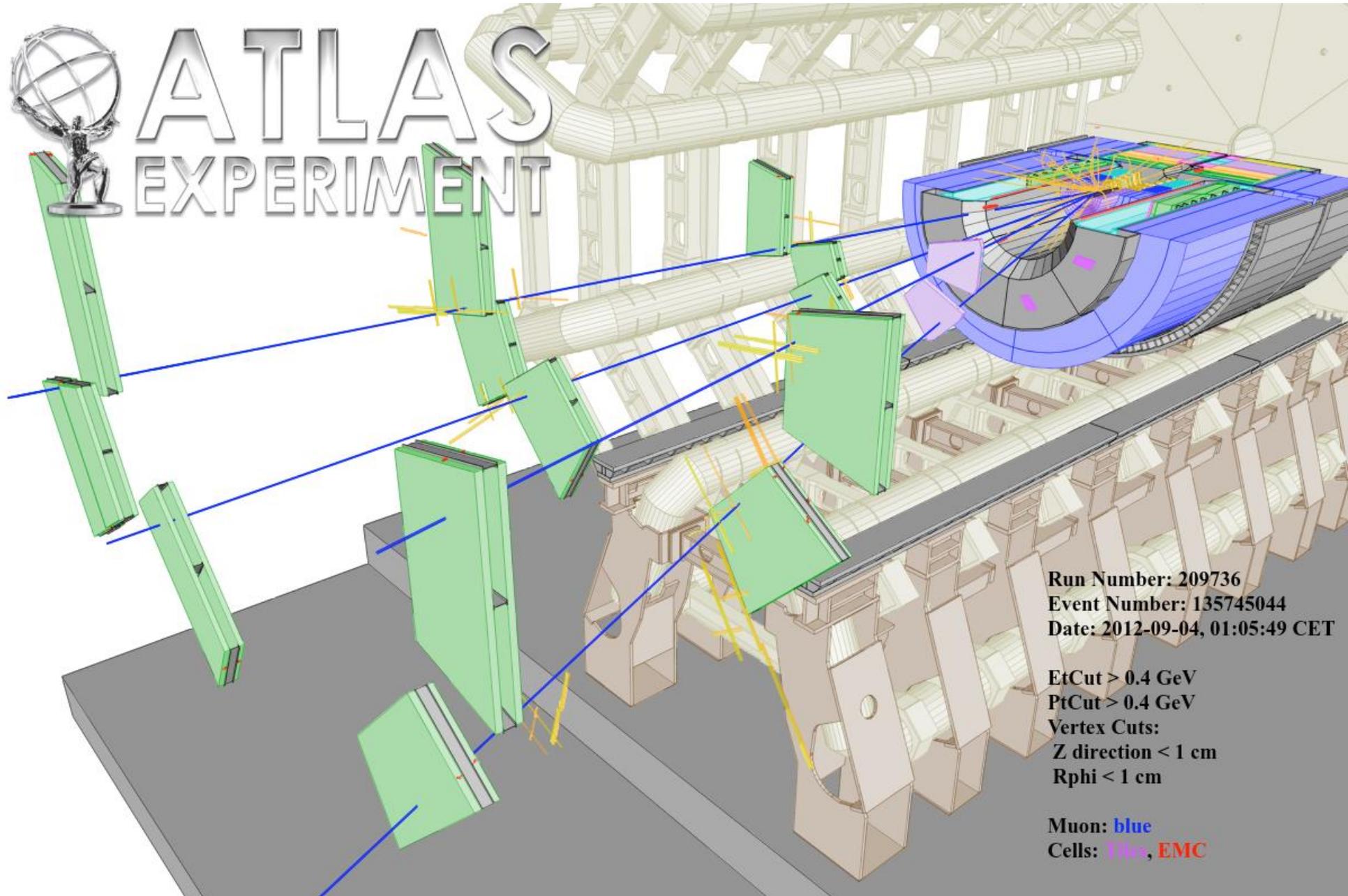
$$M_Z^2 = 2E_{l_1} E_{l_2} (1 - \cos \theta_{l_1 l_2})$$

$$M_H^2 = M_{Z_1}^2 + M_{Z_2}^2 + 2E_{Z_1} E_{Z_2} - 2p_{Z_1} p_{Z_2} \cos \theta_{Z_1 Z_2}$$

Higgs Boson in 4 Muons

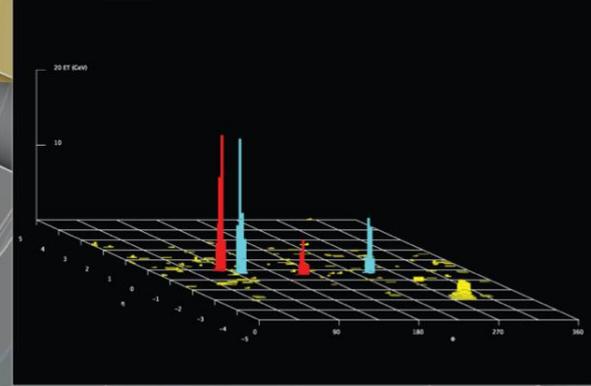
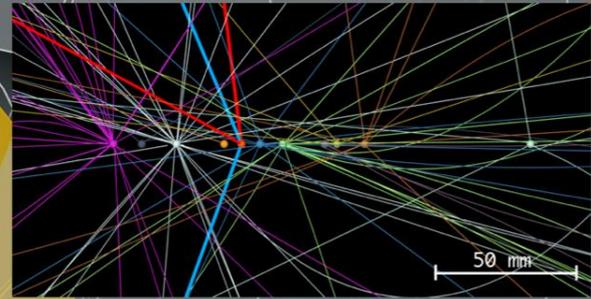
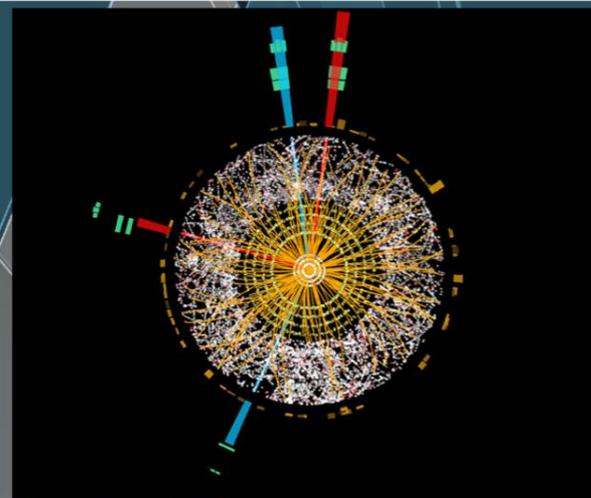
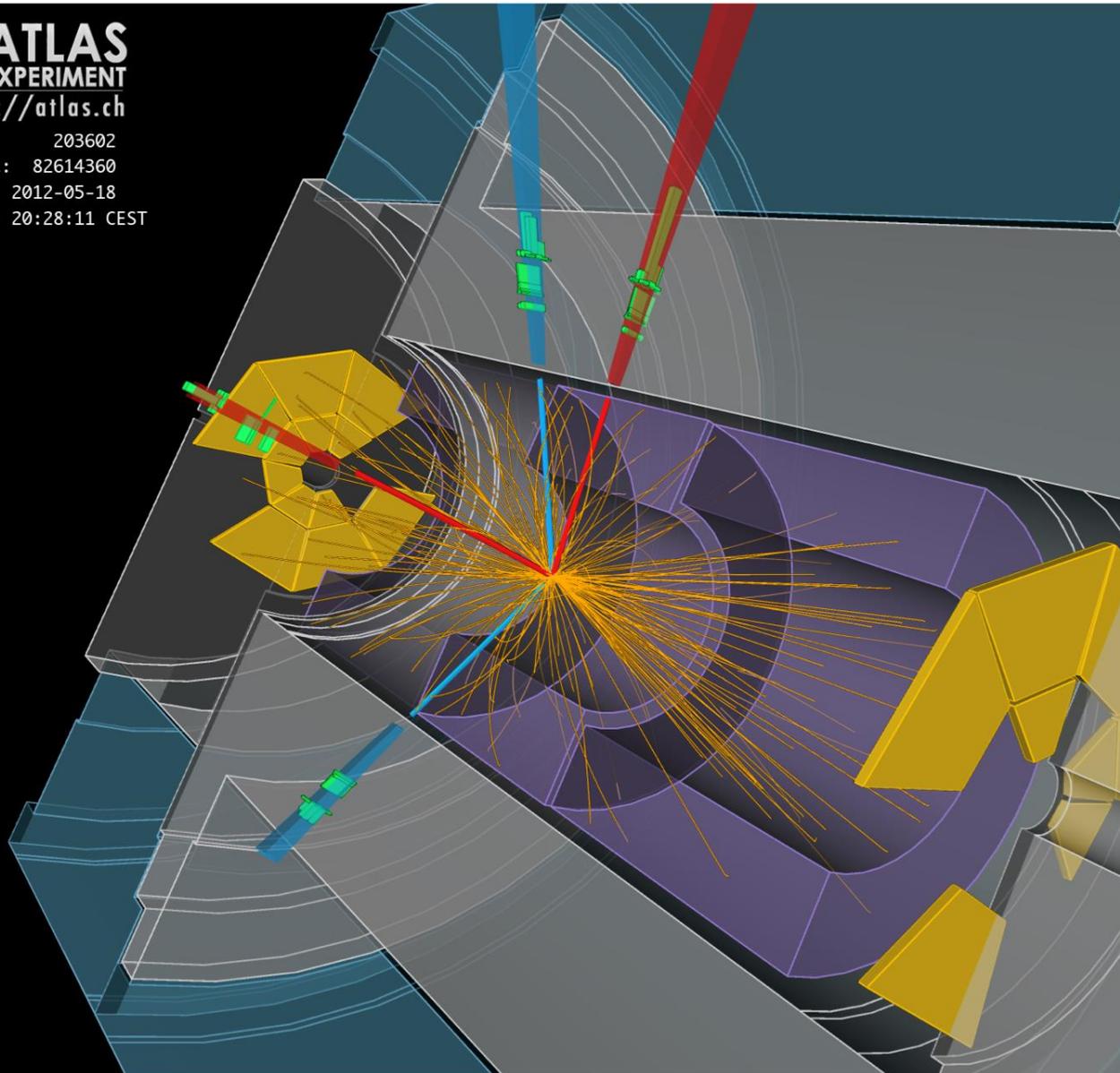


Higgs Boson in 4 Muons



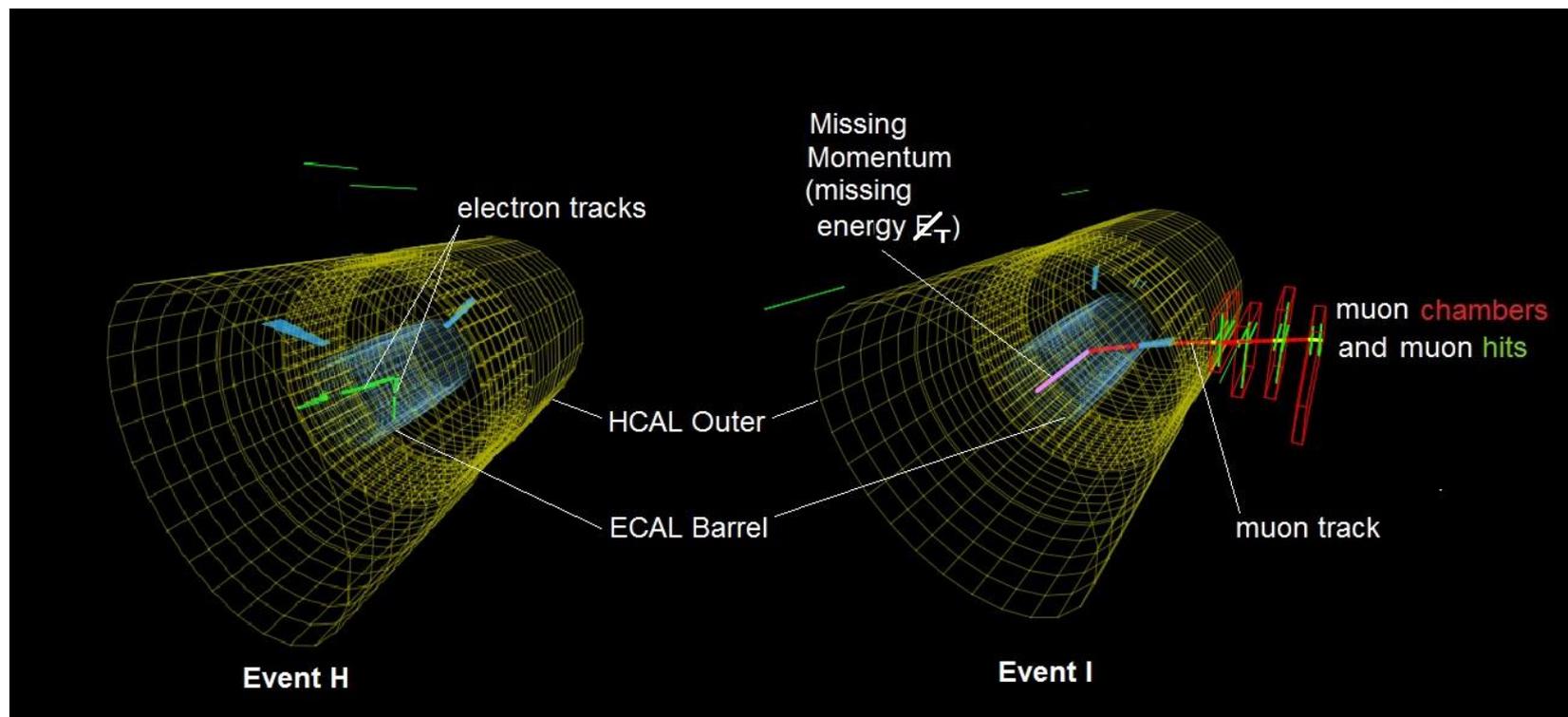
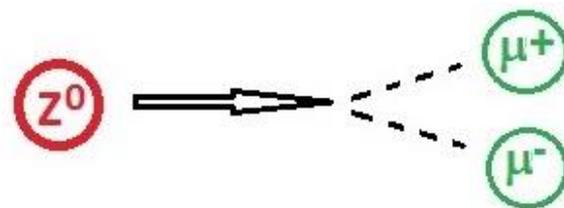
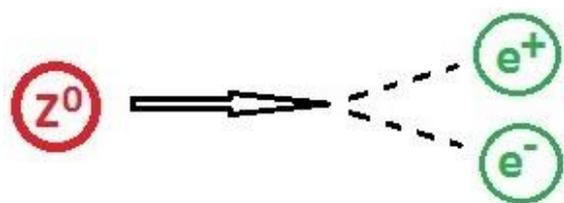
Higgs Boson in 4 Electrons

ATLAS
EXPERIMENT
<http://atlas.ch>
Run: 203602
Event: 82614360
Date: 2012-05-18
Time: 20:28:11 CEST



Activity with Event Displays

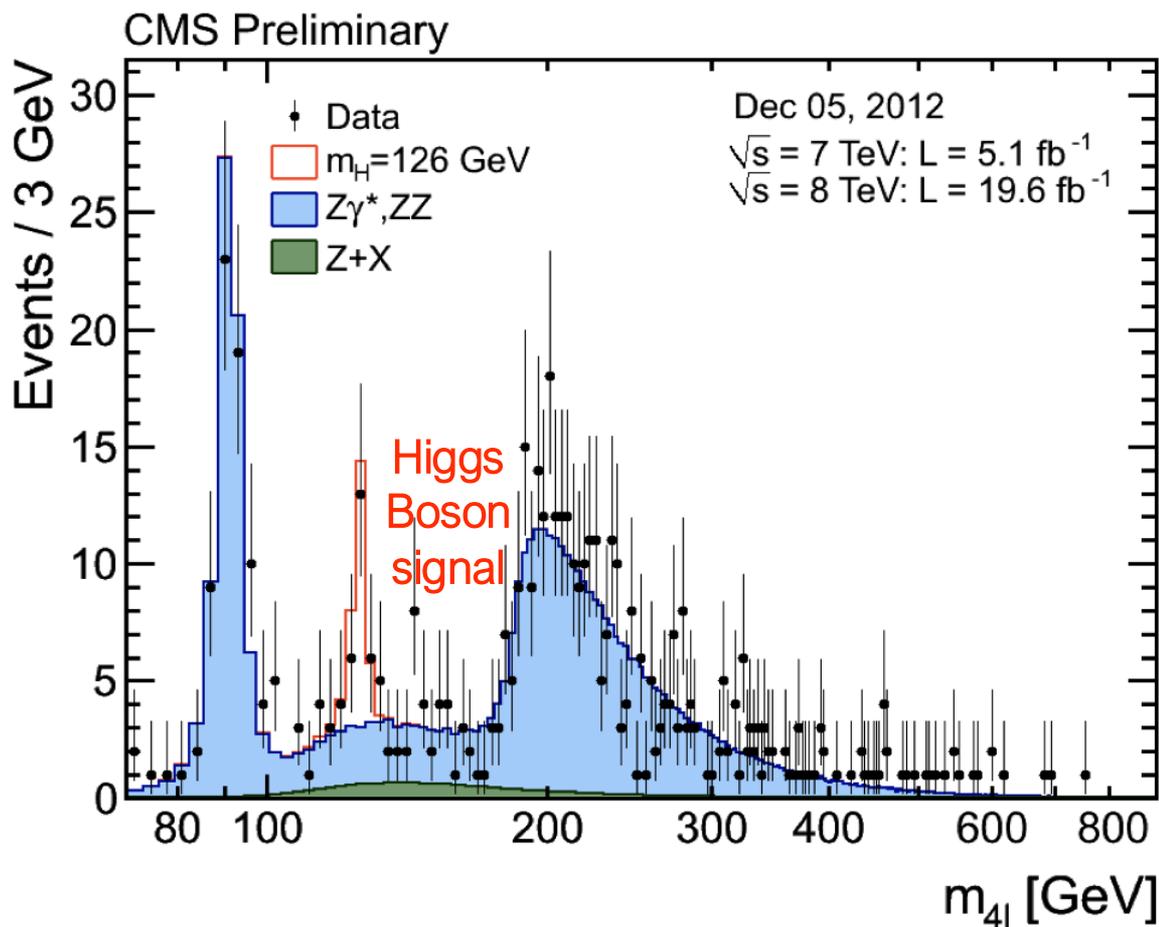
- Search for a Higgs boson event decaying to 4 leptons
- <http://ispy-webgl-masterclass.web.cern.ch/ispy-webgl-masterclass/>
- dataset: N5 file: masterclass_1.ig



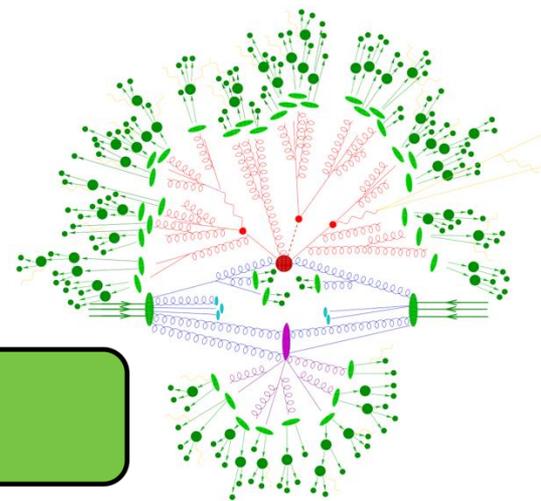
Activity with Event Displays

- Search for a Higgs boson event decaying to 4 leptons
- <http://ispy-webgl-masterclass.web.cern.ch/ispy-webgl-masterclass/>
- dataset: N5 file: masterclass_1.ig

$m_Z \sim 91 \text{ GeV}$
 $m_H \sim 125 \text{ GeV}$



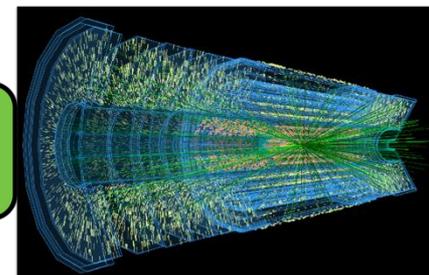
Analysis in the Data Processing Chain



TRIGGER

GENERATION

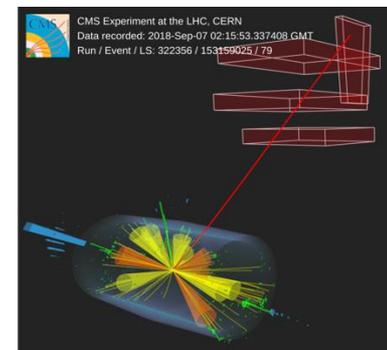
SIMULATION



RECONSTRUCTION

RECONSTRUCTION

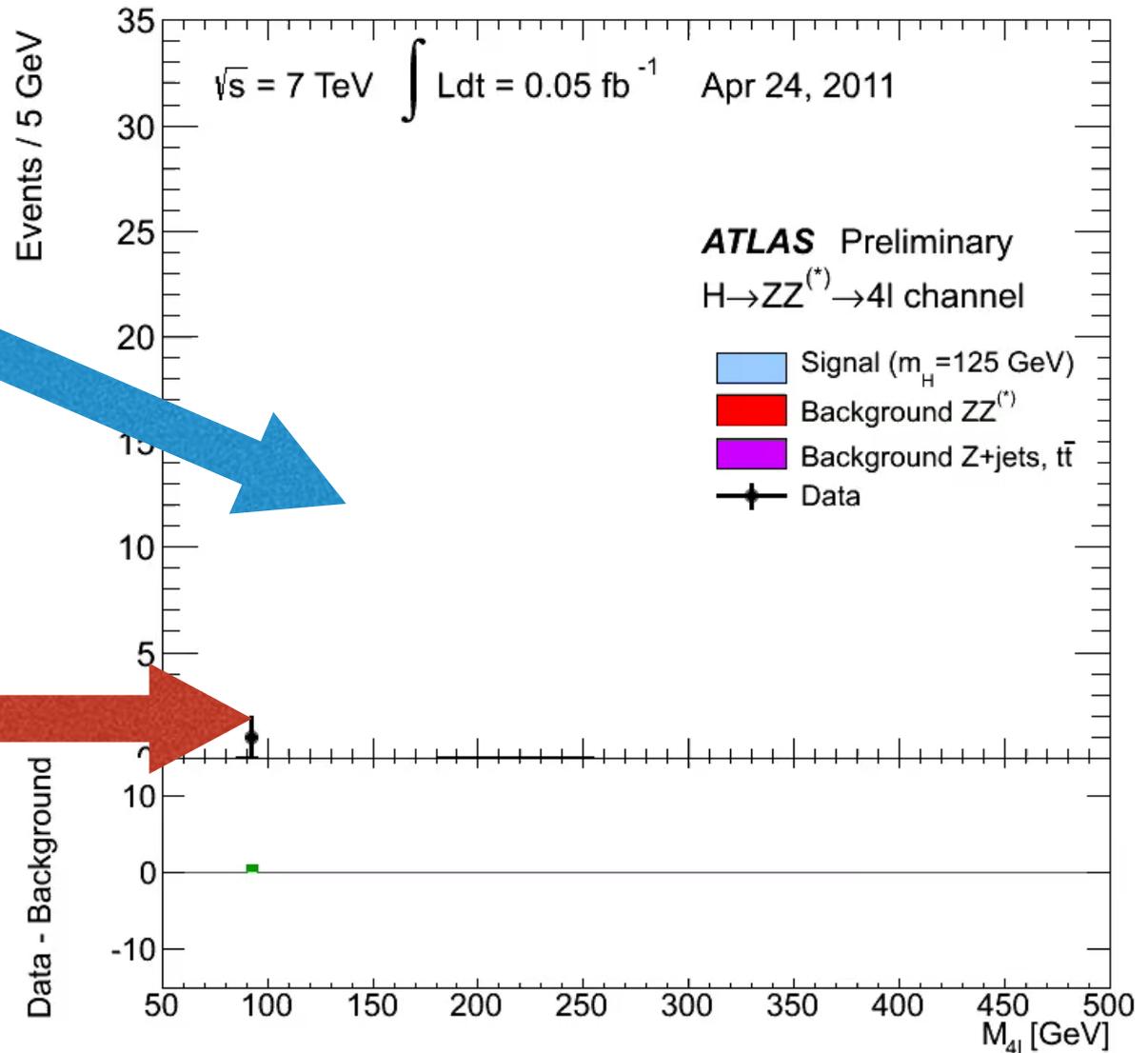
ANALYSIS



Higgs in 4 Electrons/Muons

Higgs
Boson
Signal

Other
processes



Events/2.5 GeV

ATLAS Preliminary

$H \rightarrow ZZ^* \rightarrow 4l$

13 TeV, 139 fb⁻¹

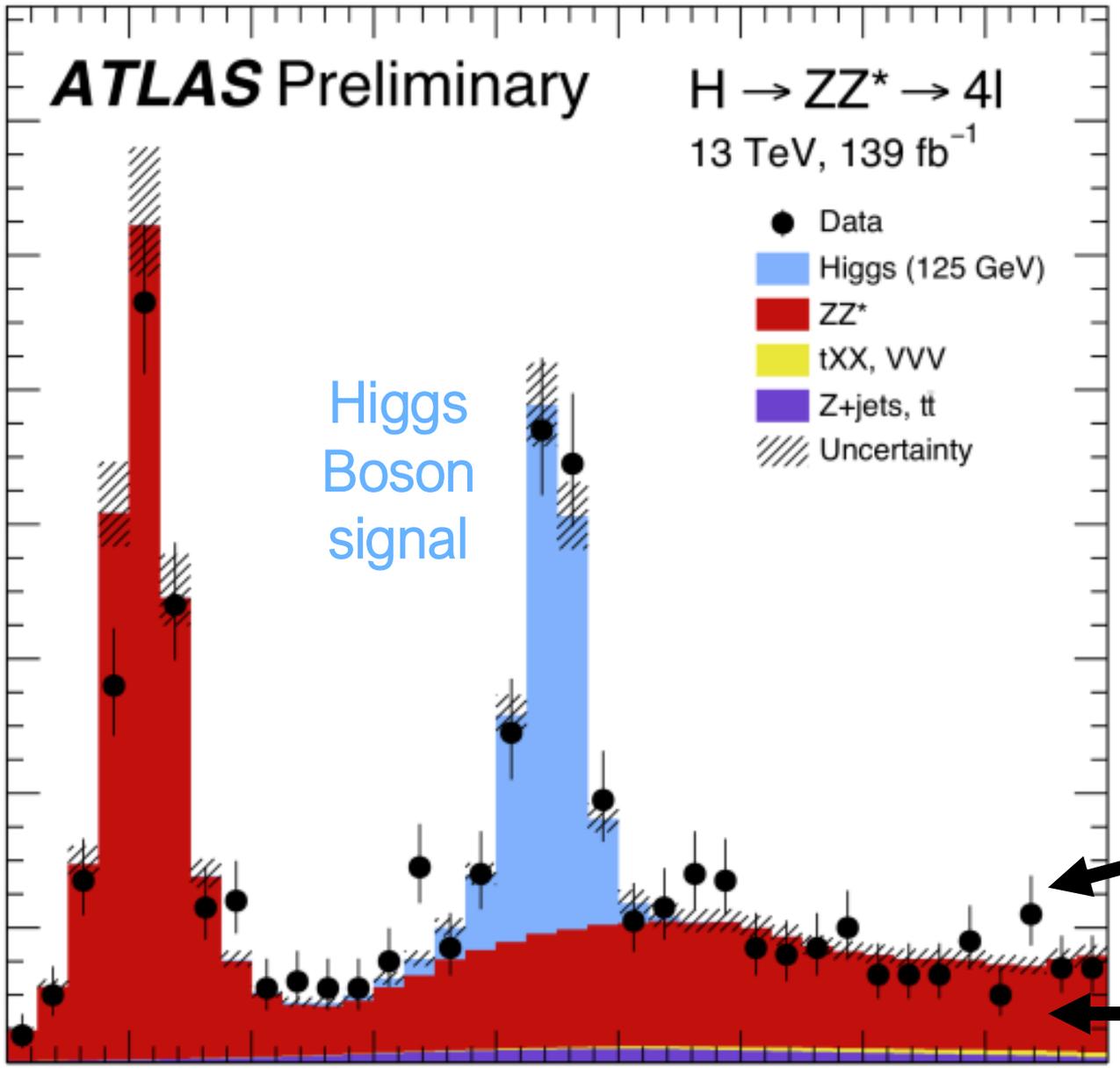
- Data
- Higgs (125 GeV)
- ZZ*
- tXX, VVV
- Z+jets, tt
- ▨ Uncertainty

Higgs
Boson
signal

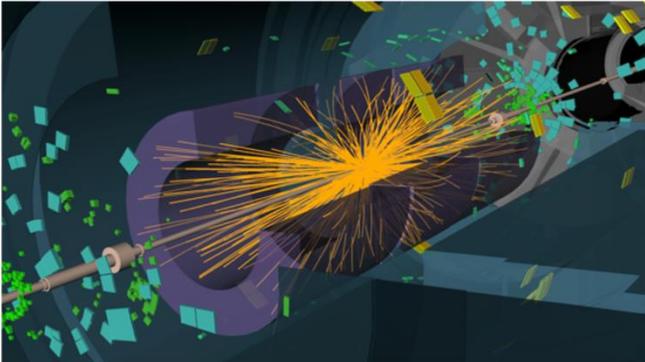
← Data

← Simulated data

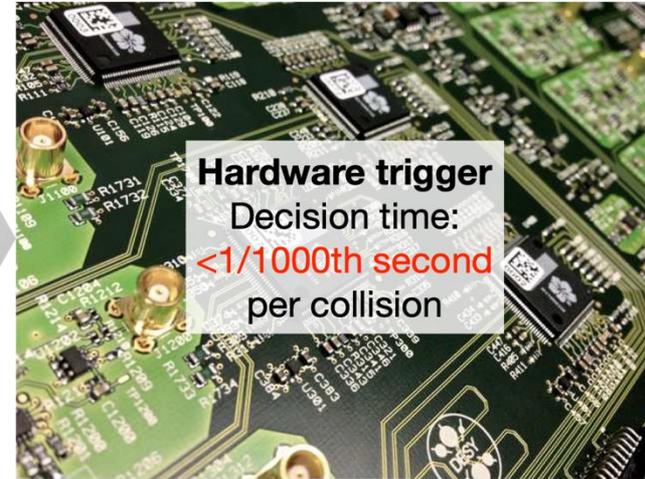
80 90 100 110 120 130 140 150 160 170
 m_{4l} [GeV]



Collecting Data from the Detector - Trigger



**40 MILLION
COLLISIONS PER
SECOND**
= **60 TB/second**
= **24 million** 30 Mbps
broadband
connections



Hardware trigger
Decision time:
<1/1000th second
per collision



High level trigger
Decision time:
0.5 seconds
per collision

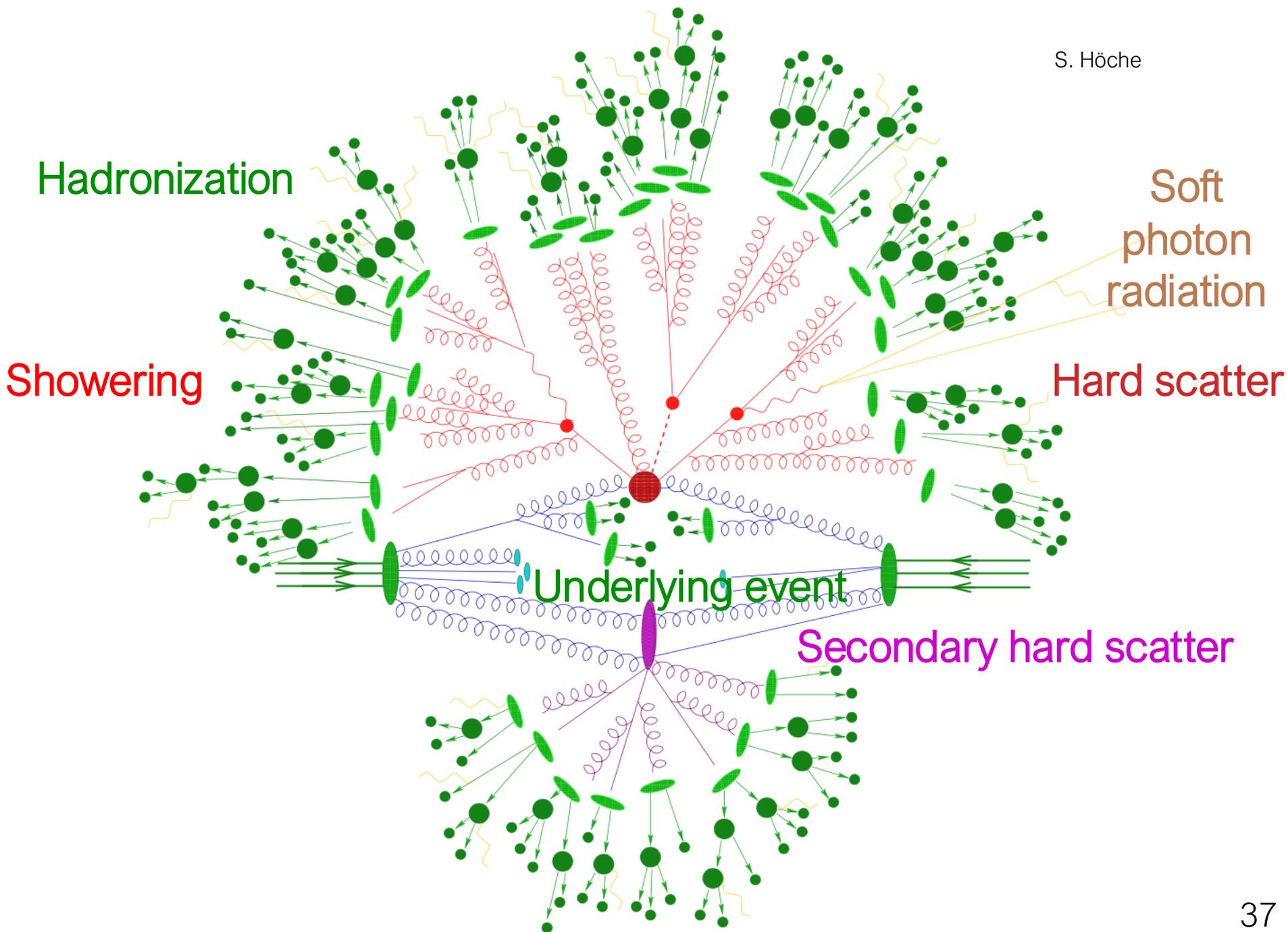
**100 000
COLLISIONS PER
SECOND**
= **160 GB/second**
= **43 000** broadband
connections

**1 000 COLLISIONS PER
SECOND**
= **1.5 GB/second**
= **400** broadband
connections

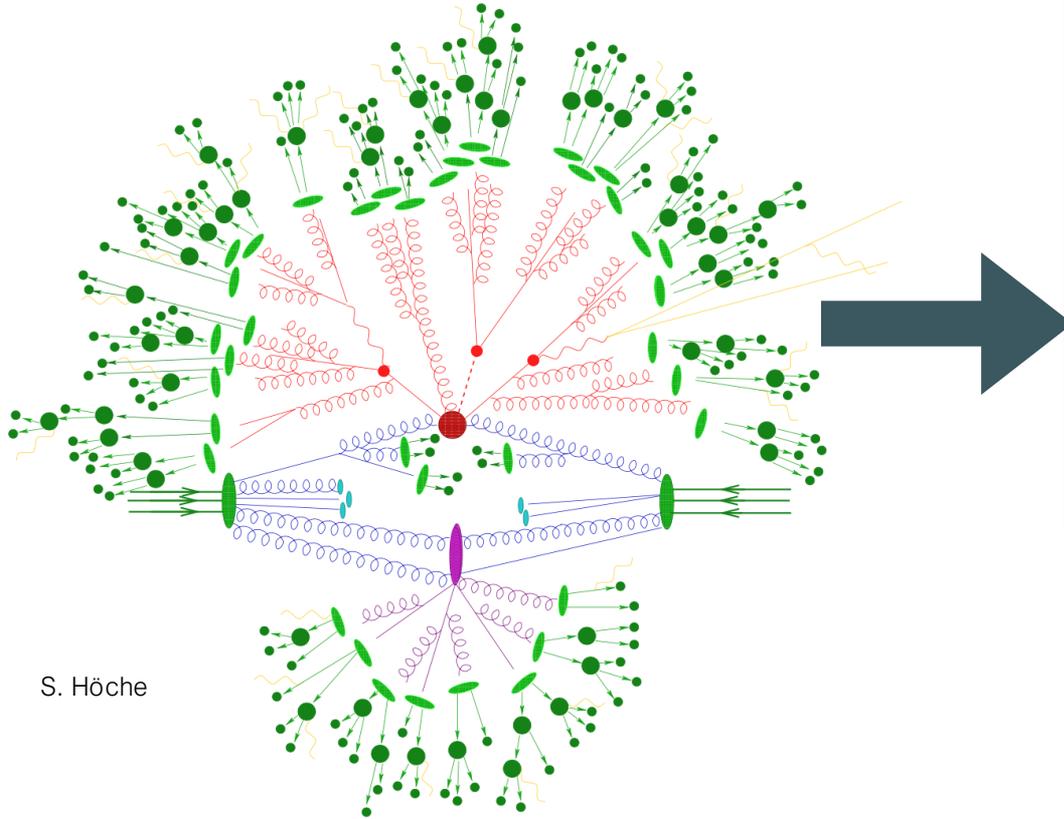
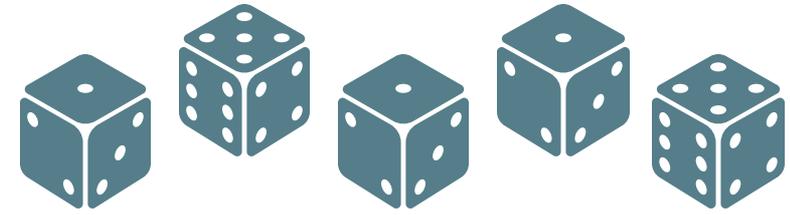


Simulated Data - Event Generation

S. Höche

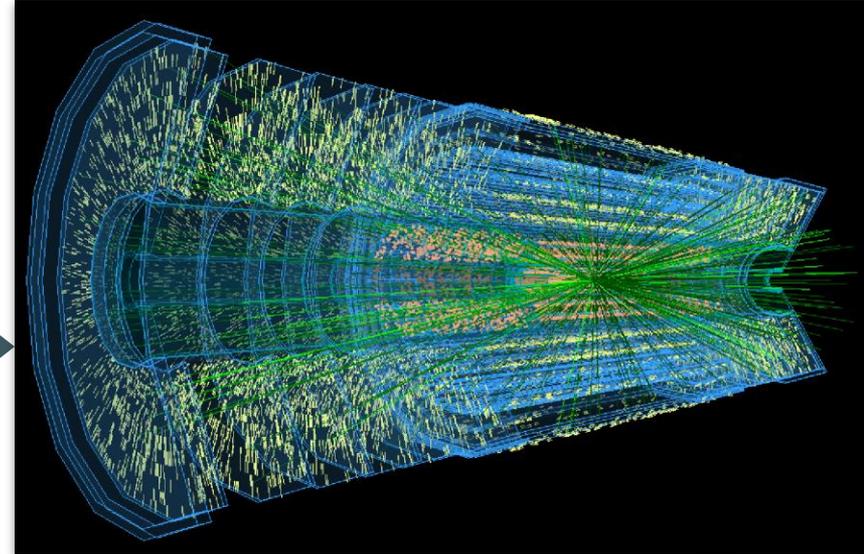


“Monte Carlo” methods



S. Höche

Event generation



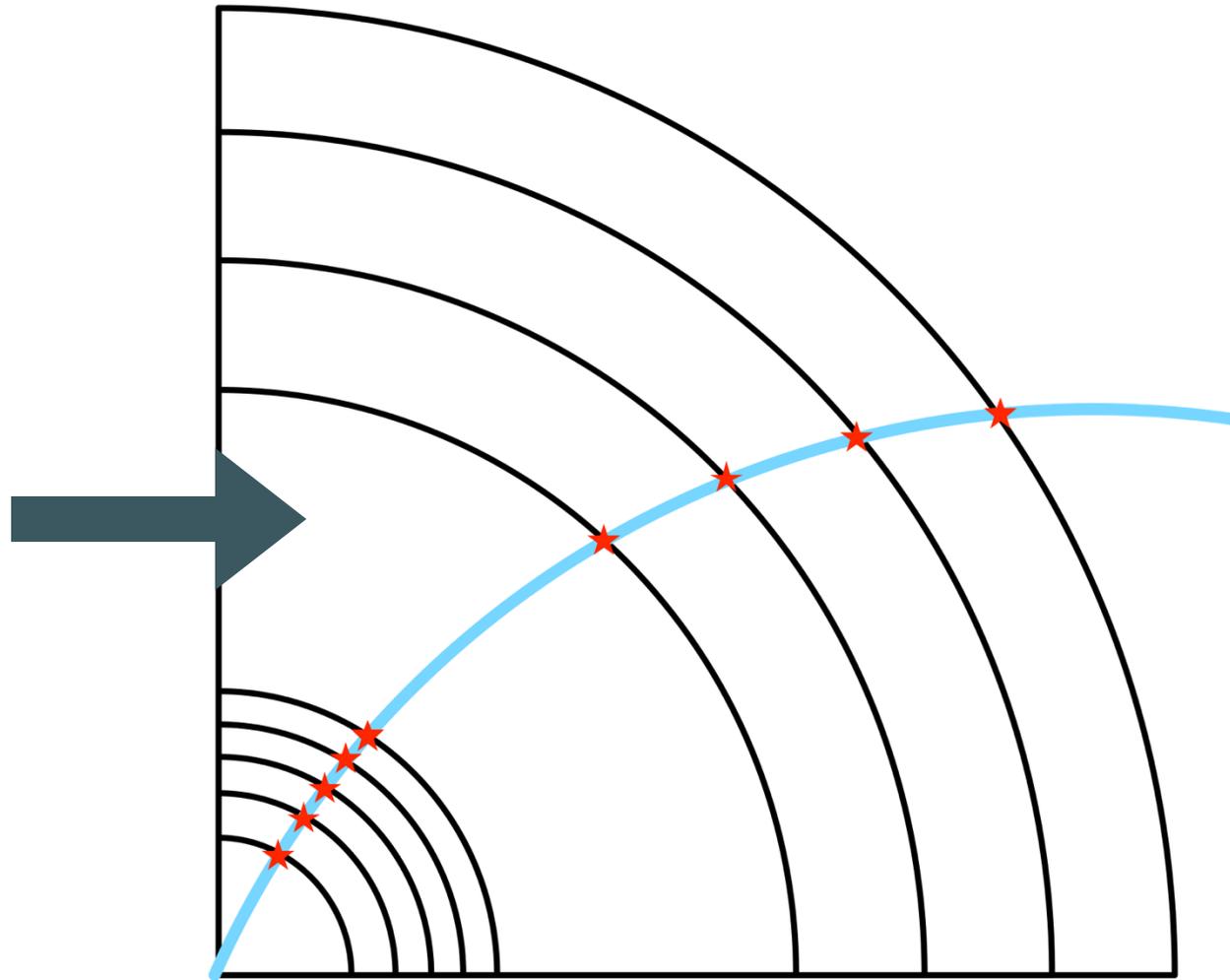
Detector simulation & digitisation

```
01100011010010010100010
10101000101010100111010
10101010001001010100010
00101010101010010101010
01010111010111101010101
11101011011010101110101
01011110000010001010001
00010010100101101010101
01101011110100101111010
11011010101110101010111
10000010001010001000100
10100101101010101011010
```



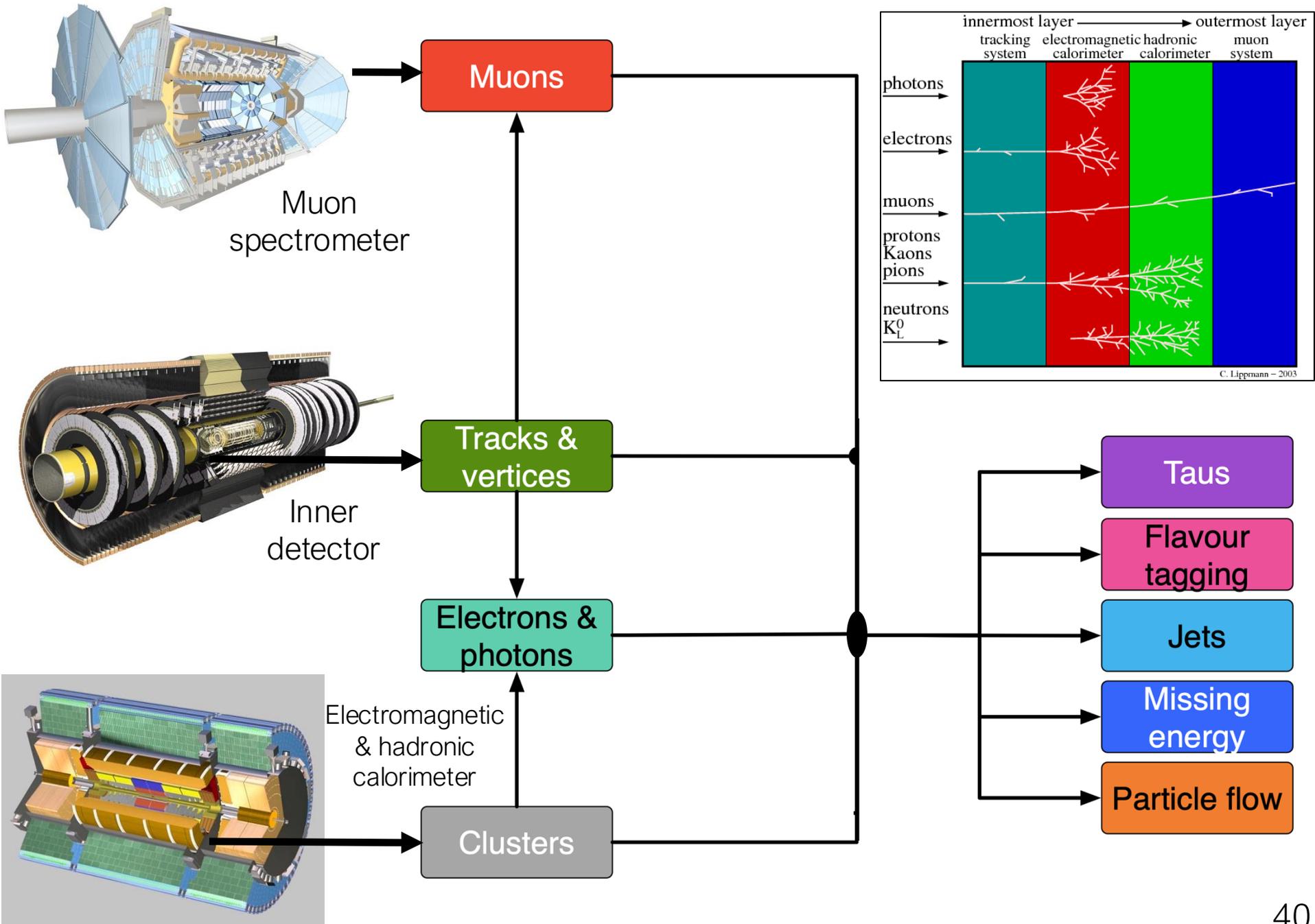
```
0110001101001001010001010
1010001010101001110101010
1010001001010100010001010
1010101001010101001010111
0101111010101011110101101
1010101110101010111100000
1000101000100010010100101
1010101010110101111010010
1111010110110101011101010
1011110000010001010001000
1001010010110101010101101
0101101010101101010110110
```

Raw data



Reconstruction

Reconstruction



Tracking variables

Electron variables

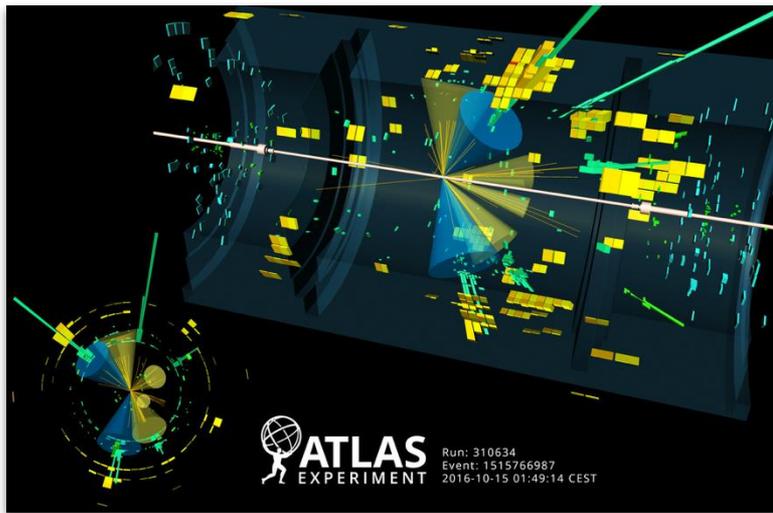
Photon variables

Jet variables

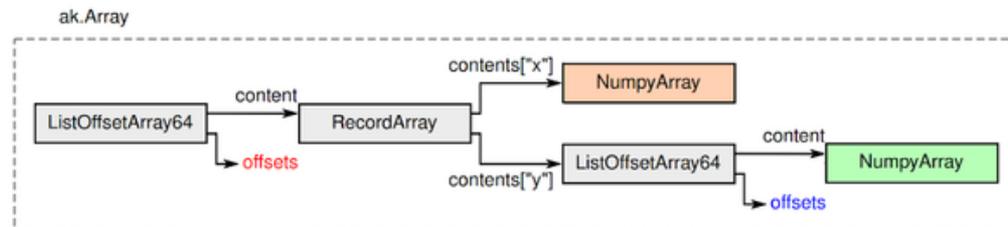
Tau variables

Muon variables...

	tr	tr	tr	e	e	e	γ	γ	γ	j	j	j	τ	τ	τ	μ	μ	μ						
1	#	#	...	#	#	...	#	#	...	#	#	...	#	#	...	#	#	...	#	#	...	#	#	...



```
array = ak.Array([
  [{"x": 1.1, "y": [1]}, {"x": 2.2, "y": [1, 2]}, {"x": 3.3, "y": [1, 2, 3]},
  []],
  [{"x": 4.4, "y": [1, 2, 3, 4]}, {"x": 5.5, "y": [1, 2, 3, 4, 5]}]
])
```



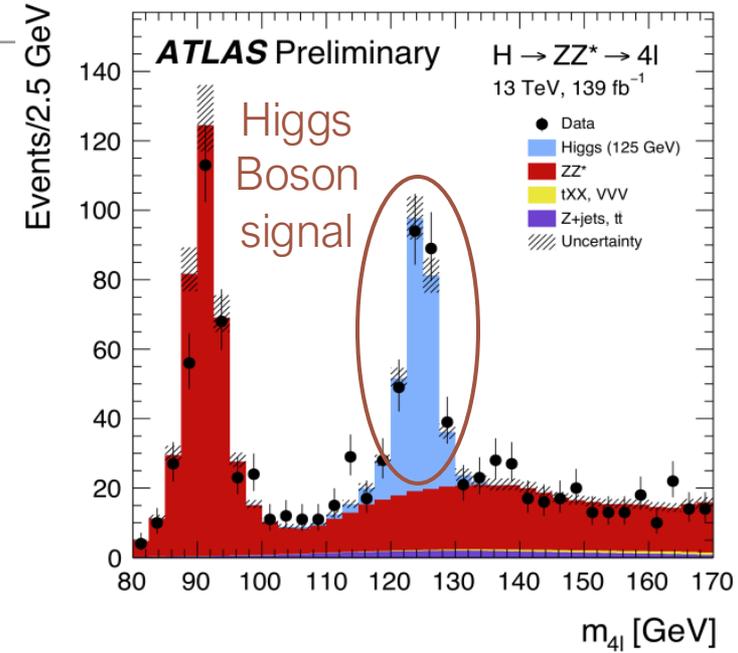
- Record structures with differently typed fields
- Array of variable-length lists ("ragged" or "jagged" arrays)
- Nested variable-length lists
- Missing data
- Heterogenous data (union/variant types)

Tracking variables Electron variables Photon variables Jet variables Tau variables Muon variables...

	tr	tr	tr	e	e	e	γ	γ	γ	j	j	j	τ	τ	τ	μ	μ	μ				
1	#	#	..	#	#	...	#	#	...	#	#	...	#	#	...	#	#	...	#	#	...	#
2	#	#	..	#	#	...	#	#	...	#	#	...	#	#	...	#	#	...	#	#	...	#
3	#	#	..	#	#	...	#	#	...	#	#	...	#	#	...	#	#	...	#	#	...	#
4	#	#	..	#	#	...	#	#	...	#	#	...	#	#	...	#	#	...	#	#	...	#
5	#	#	..	#	#	...	#	#	...	#	#	...	#	#	...	#	#	...	#	#	...	#
6	#	#	..	#	#	...	#	#	...	#	#	...	#	#	...	#	#	...	#	#	...	#
7	#	#	..	#	#	...	#	#	...	#	#	...	#	#	...	#	#	...	#	#	...	#
...
N	#	#	..	#	#	...	#	#	...	#	#	...	#	#	...	#	#	...	#	#	...	#

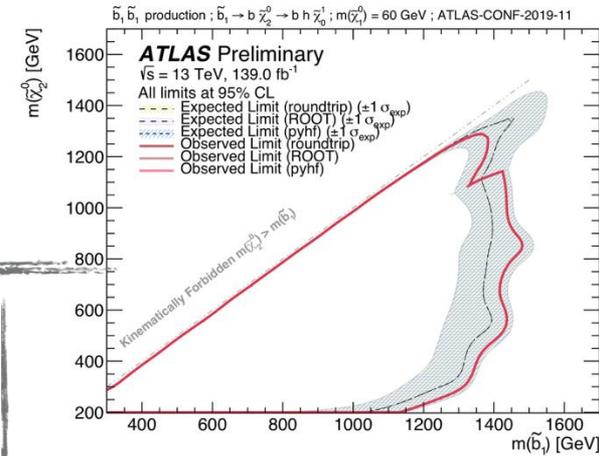
Analysis Tasks

	tr	tr	tr	tr	ee	ee	ee	ee	$\gamma\gamma$	$\gamma\gamma$	$\gamma\gamma$	$\gamma\gamma$	jj	jj	jj	jj	$\tau\tau$	$\tau\tau$	$\tau\tau$	$\tau\tau$	$\mu\mu$	$\mu\mu$	$\mu\mu$	$\mu\mu$
1	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#
2	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#
3	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#
4	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#
5	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#
6	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#
7	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#
...
N	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#	#	#	...	#



Skimming/Slimming

Build Histograms



Datasets

Add information

Build Statistical Models

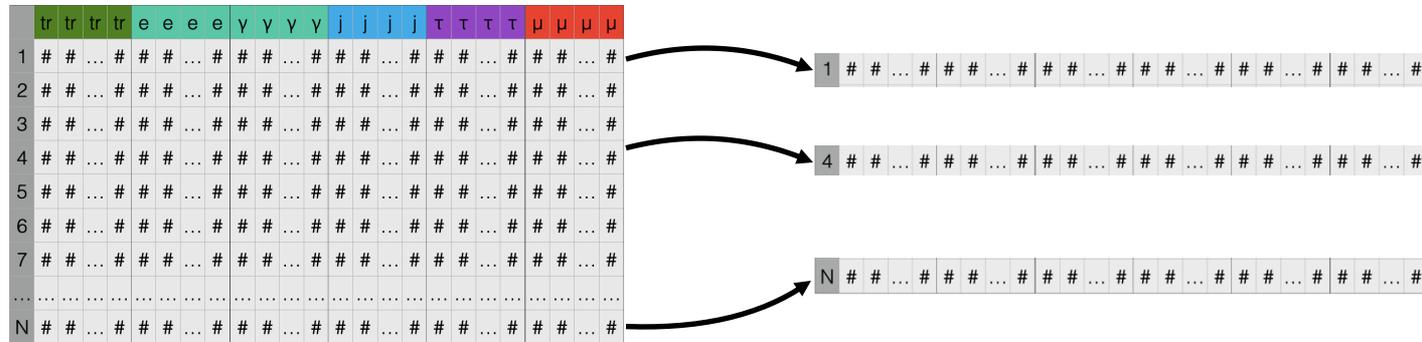
Step 1: bulk analysis

Step 2: final analysis

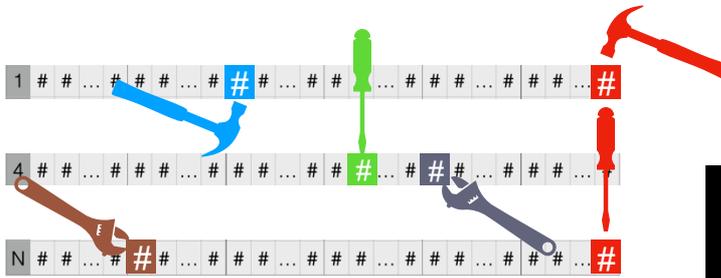
Data Processing for Analysis

- Step 1: bulk analysis – usually done on distributed computing resources – the Grid

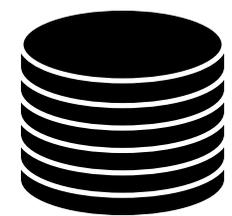
Scan over thousands of files, millions/billions events
TB-PB
Select events of interest



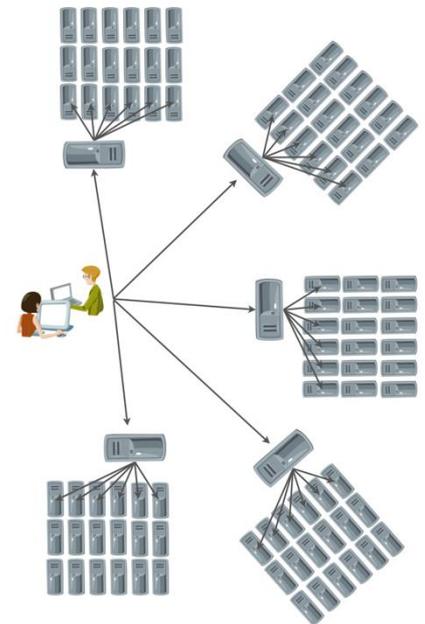
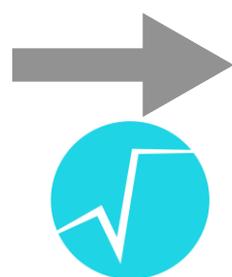
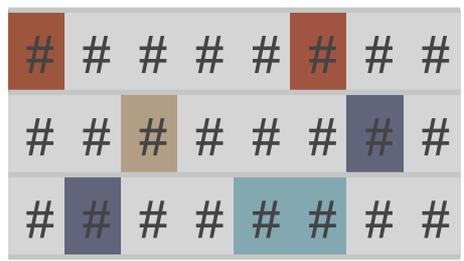
Apply calibration/corrections



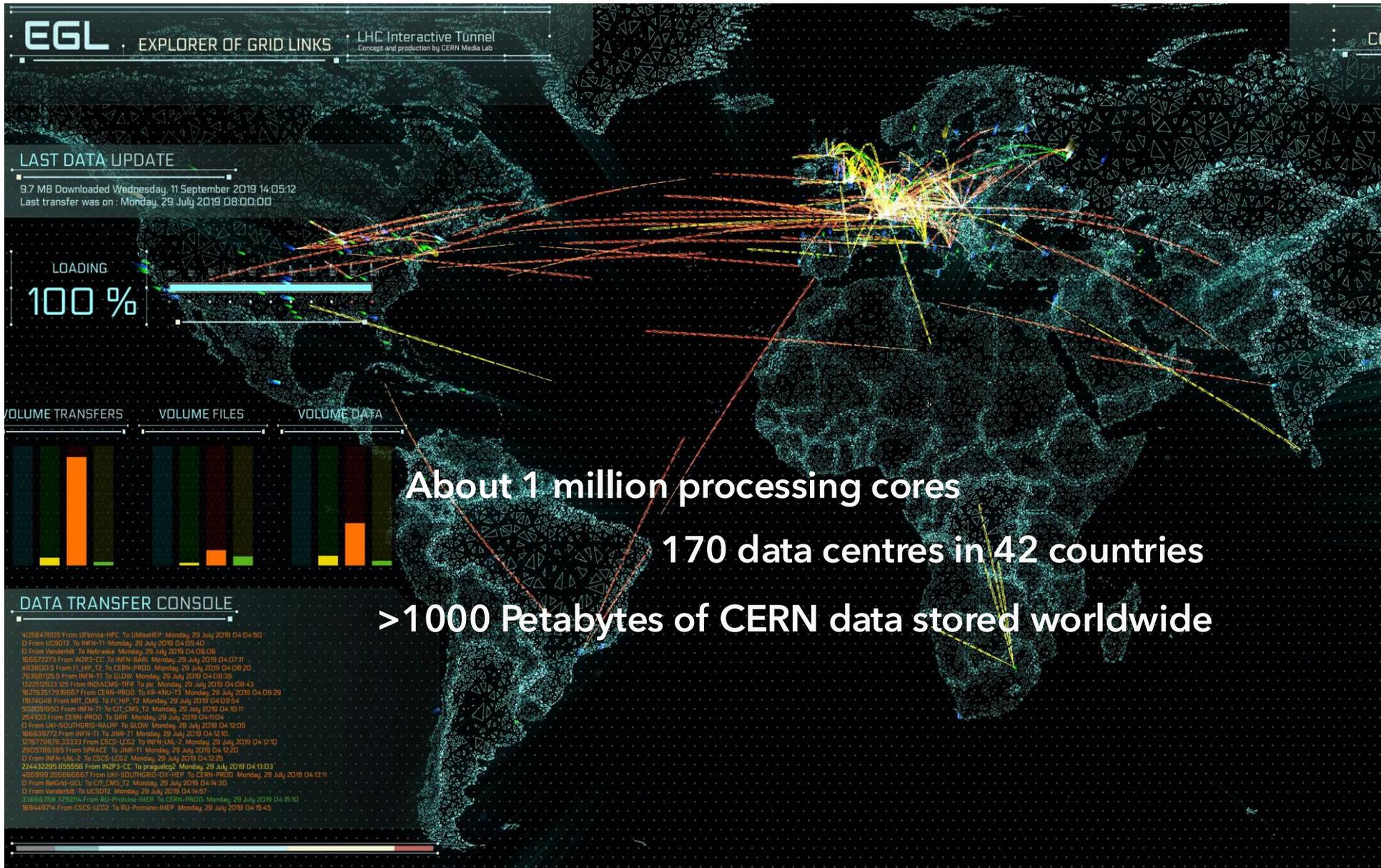
Save to disk
(GB-TB)



Retain only required variables



The Worldwide LHC Computing Grid (WLCG)



About 1 million processing cores

170 data centres in 42 countries

>1000 Petabytes of CERN data stored worldwide

Data Processing for Analysis

- Step 2: final analysis - usually done locally – small clusters or personal desktop/laptops

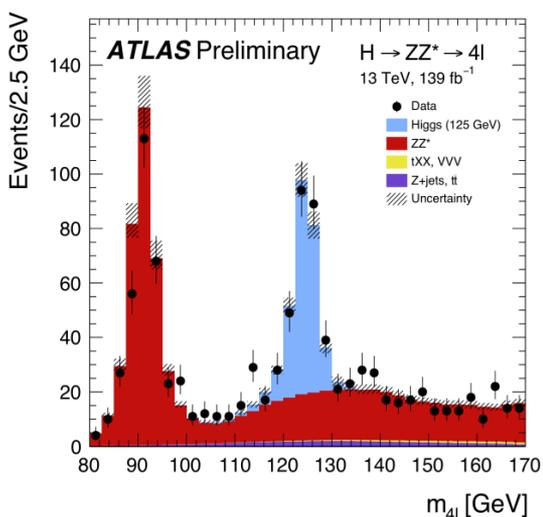
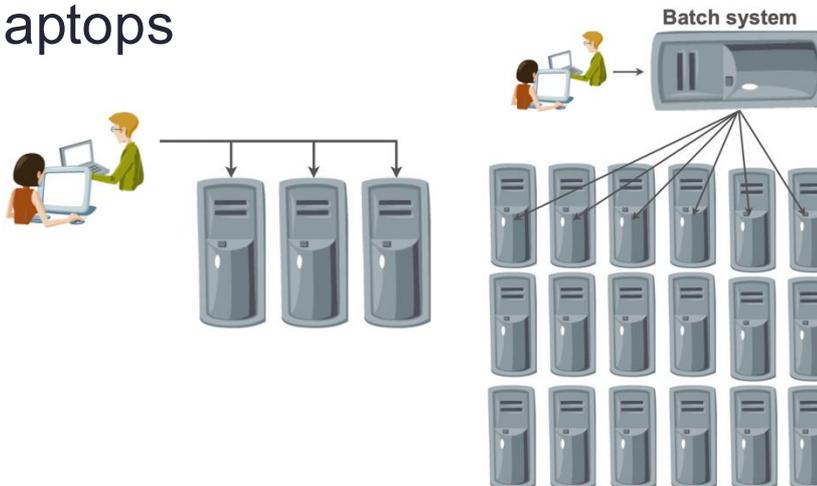
Machine learning training

Background studies

Systematics

Statistical analysis

Final plots



```
In [9]: # Number of events to process
# Note that skipping the first N events is also possible, using the second arg of the Range method in the next cell
maxEvents = 5000

In [10]: # Extract the data for the chosen branch via the AsNumpy method
branchName = 'MuonsAuxDyn.eta'
columnType = dataframe.GetColumnType(branchName)
column = dataframe.Range(0, maxEvents).AsNumpy(columns=[branchName])
# Second arg is for case where container is a vector-of-vectors (e.g. track cov matrix)
# In which case the i-th element is taken. Not used for vector
theList = AsList(column[branchName], 0)

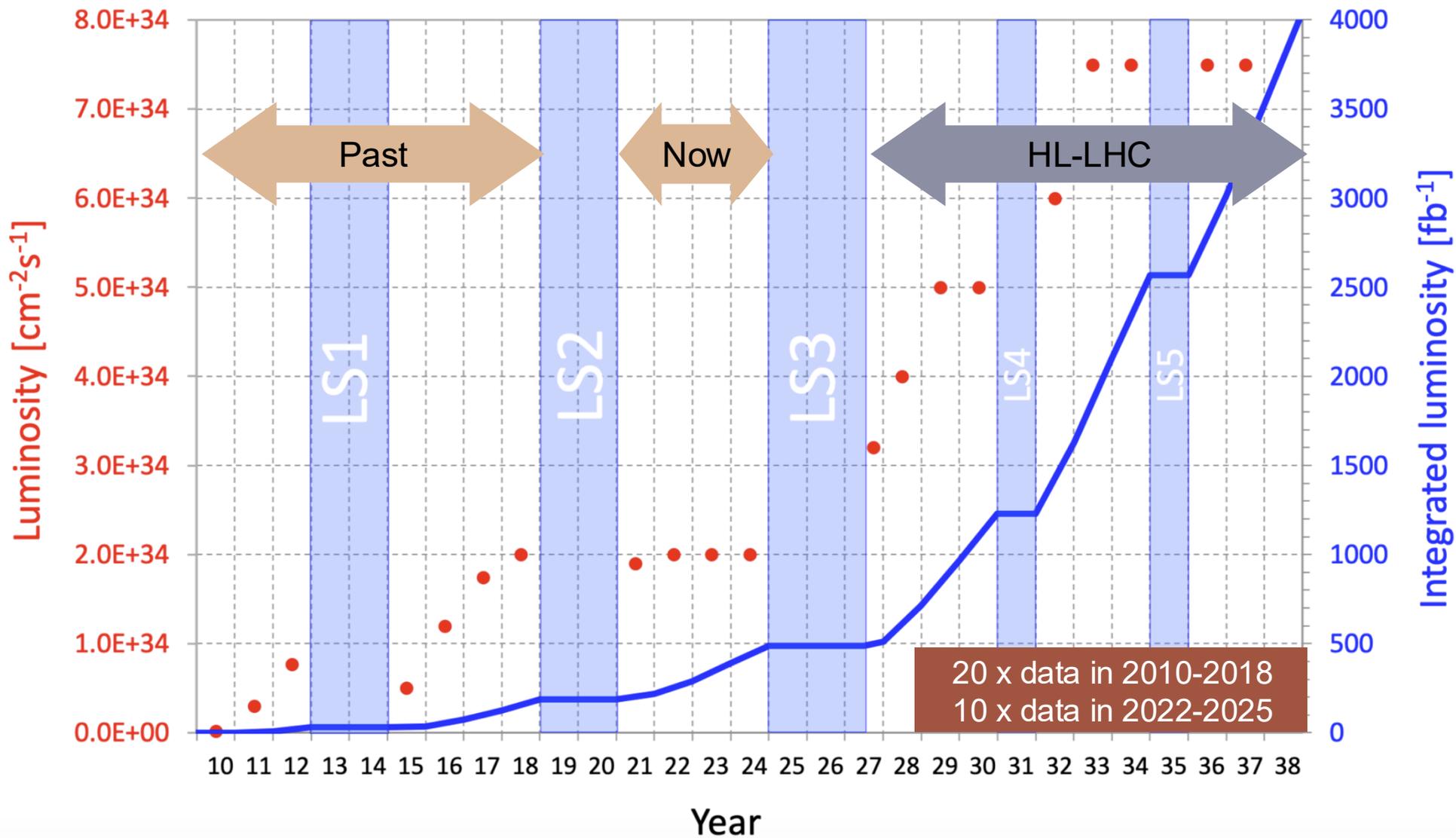
In [11]: # Make a plot
num_bins = 500
fig, ax = plt.subplots(figsize=(20,10))
ax.set_xlabel(branchName)
n, bins, patches = ax.hist(theList, num_bins, density=False, alpha=0.5, color='blue')
plt.tight_layout()
plt.show()
```

The Future at the High Luminosity LHC (HL-LHC)

Rate of collisions

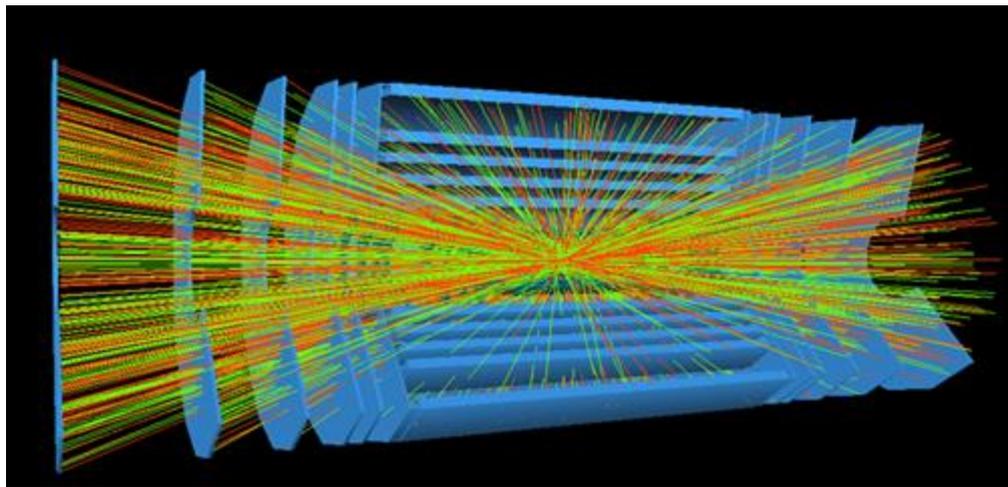
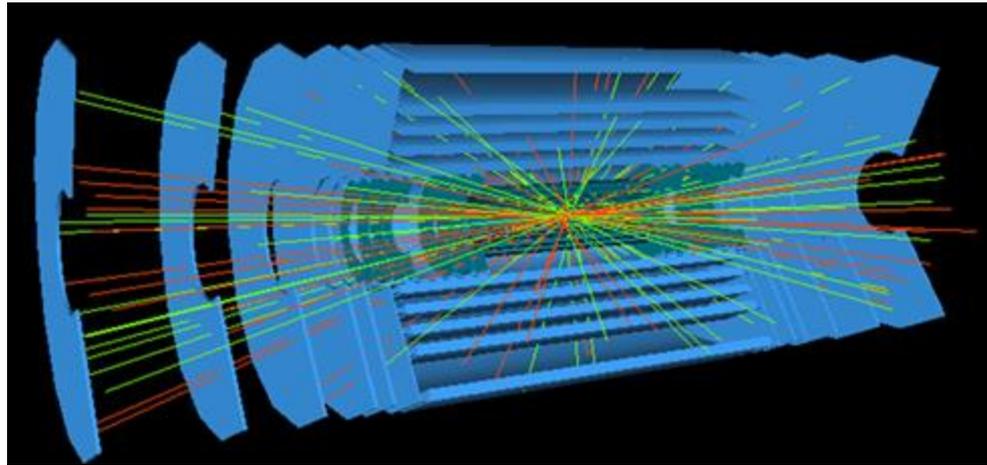
Total data recorded

● Peak luminosity — Integrated luminosity



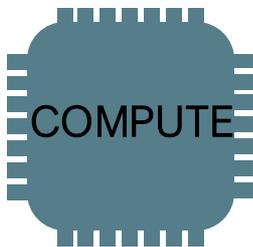
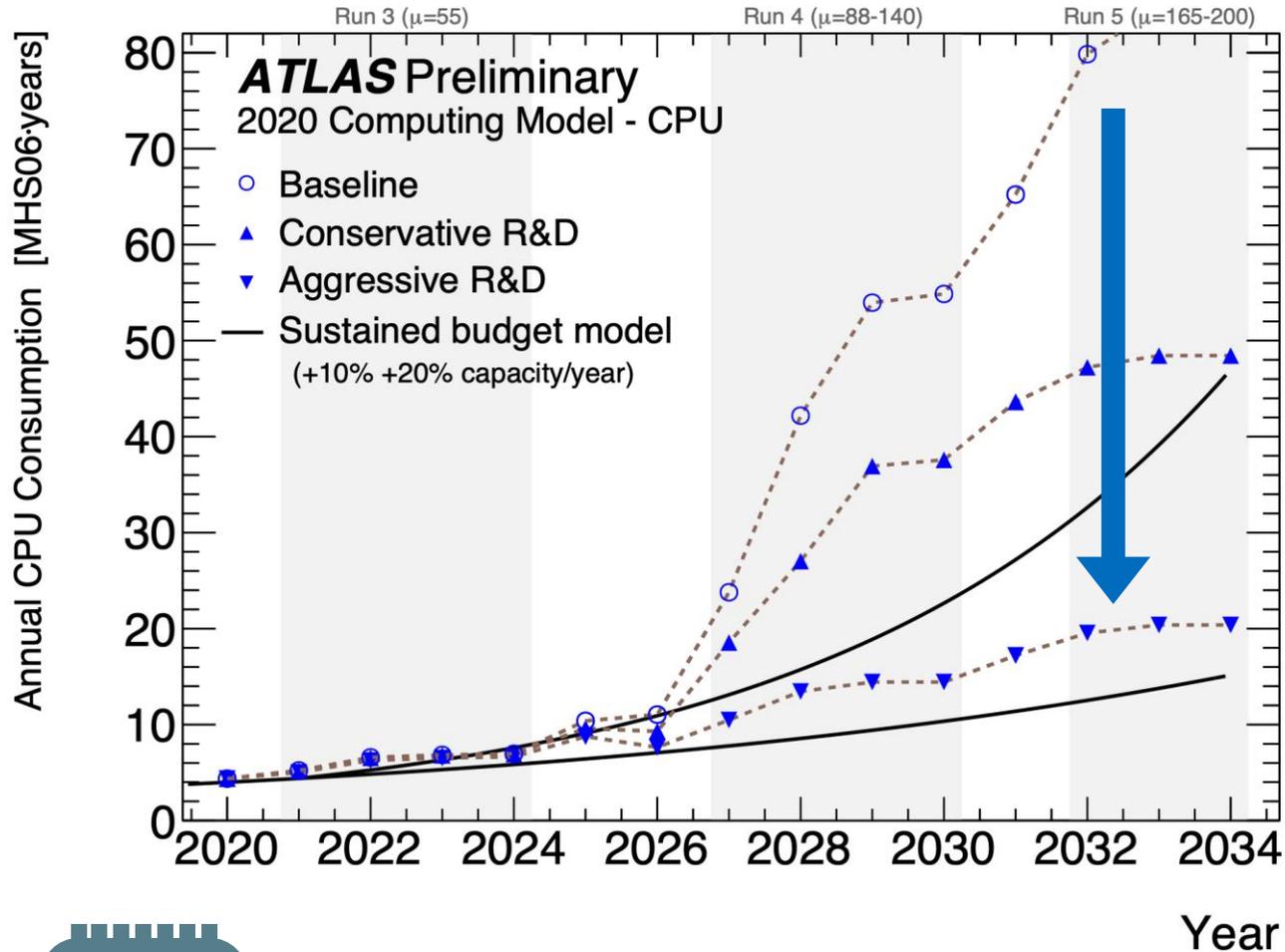
20 x data in 2010-2018
10 x data in 2022-2025

The Future at the High Luminosity LHC (HL-LHC)



200 collisions in each bunch crossing

Computing Demands of the HL-LHC



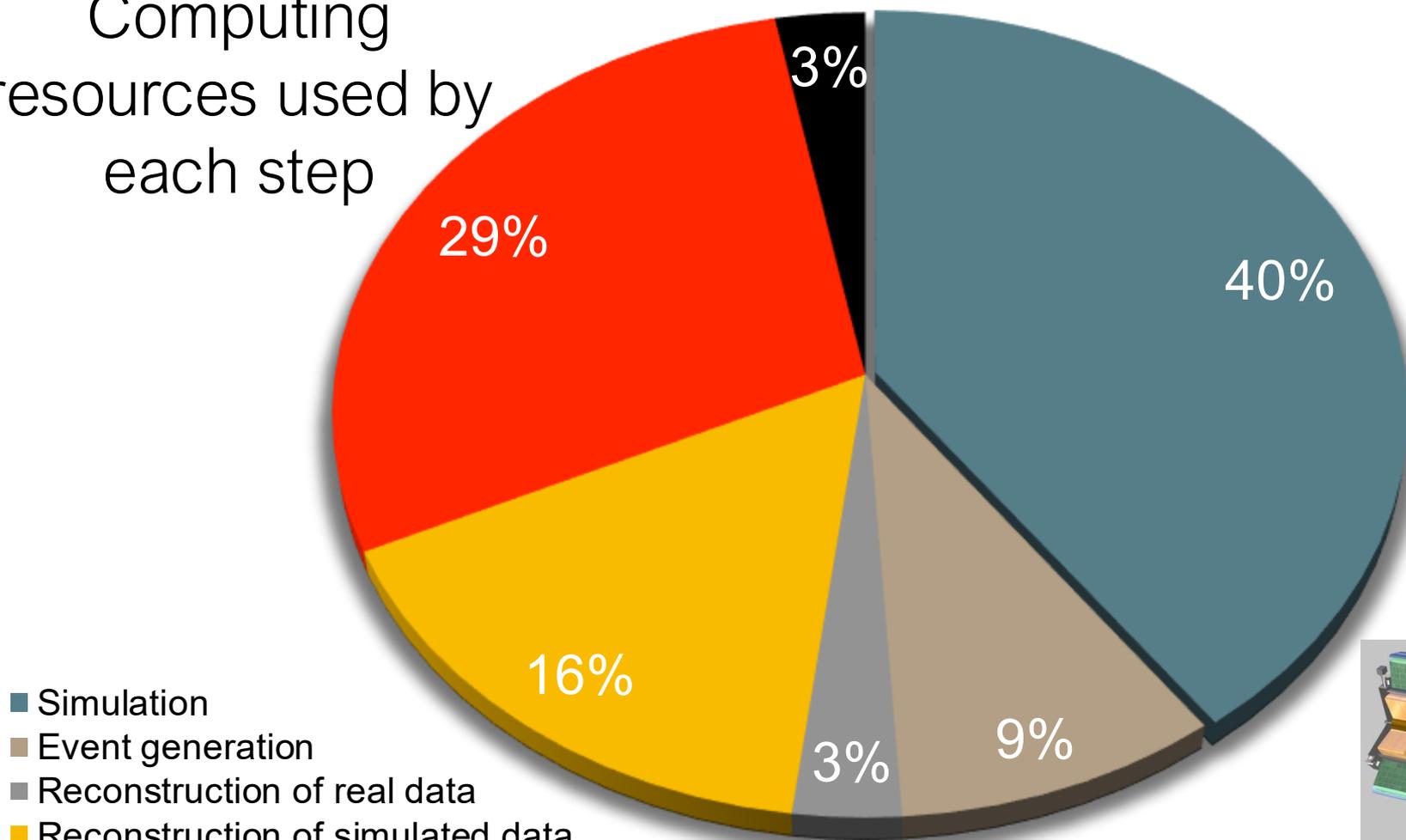
With current usage, would need more than about 5M cores by ~2035

R&D program to reduce compute needs by ~ x4

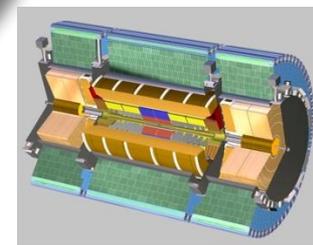
Meeting the HL-LHC Challenge

More efficient software & new methods

Computing resources used by each step



- Simulation
- Event generation
- Reconstruction of real data
- Reconstruction of simulated data
- Preparation of analysis data & analysis
- Other



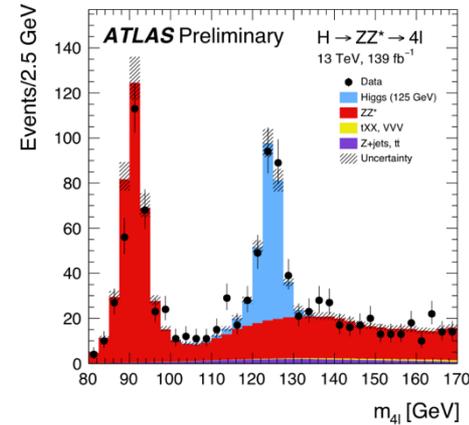
Simulating the calorimeters is very compute-intensive.

Speed up with
"FastCaloSim" 50

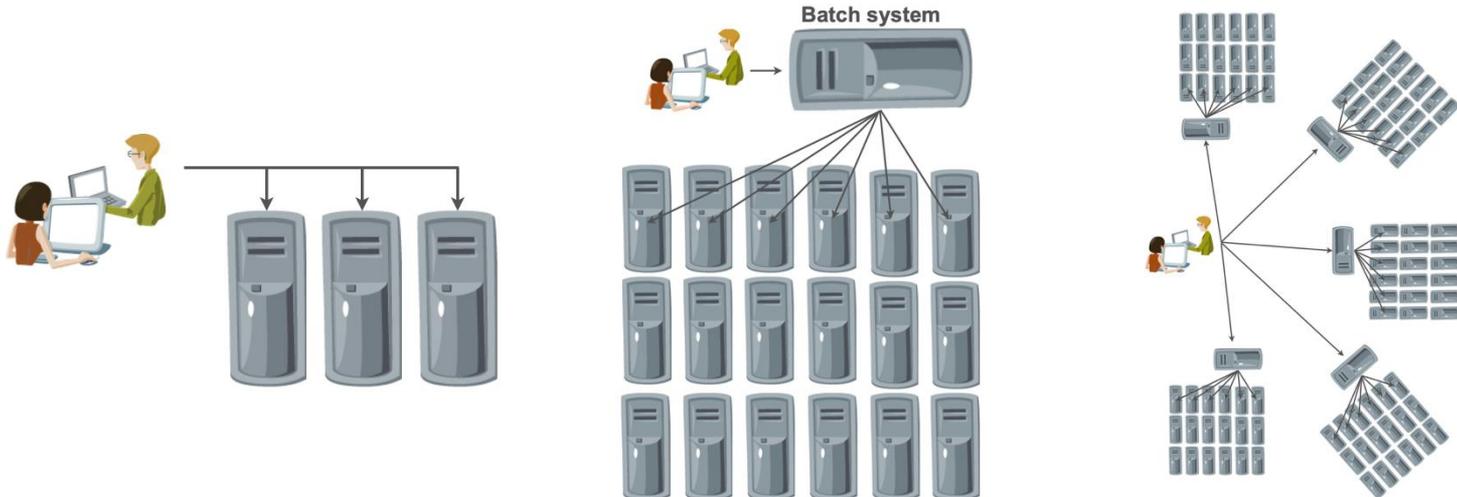
Challenges for Analysis in the Future

Already facing **several bottlenecks**, expected more challenging the future

- Processing times – need to be fast



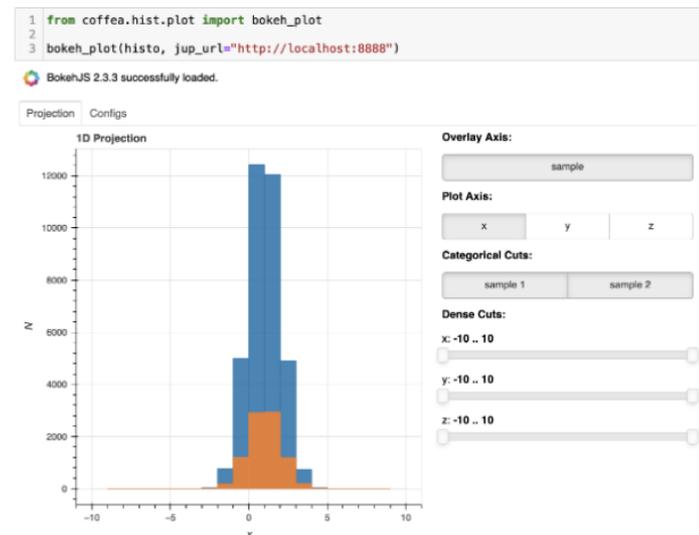
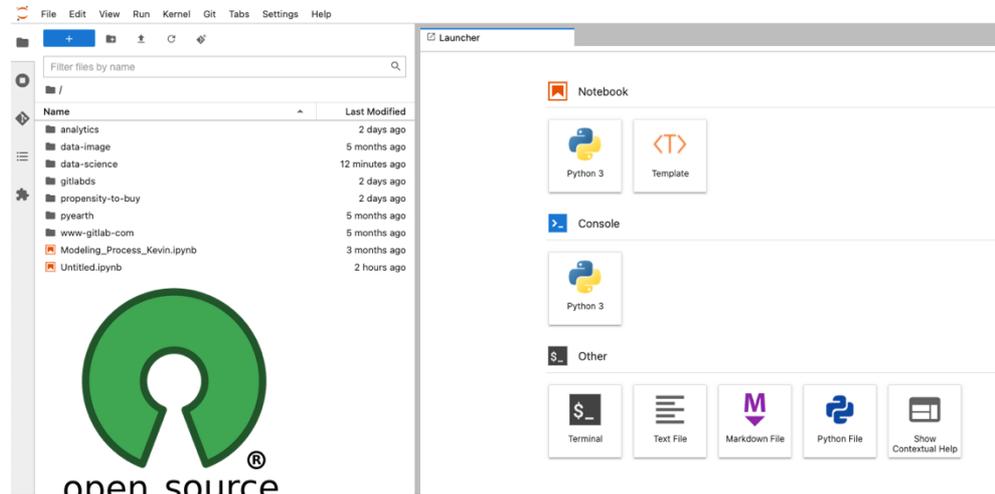
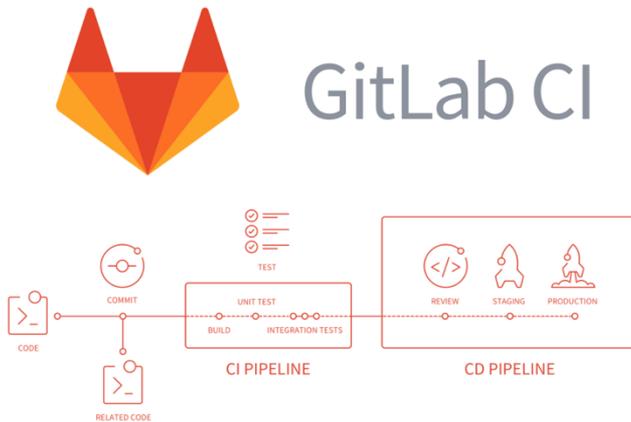
- Dataset sizes – need to be able to scale out



Challenges for Analysis in the Future

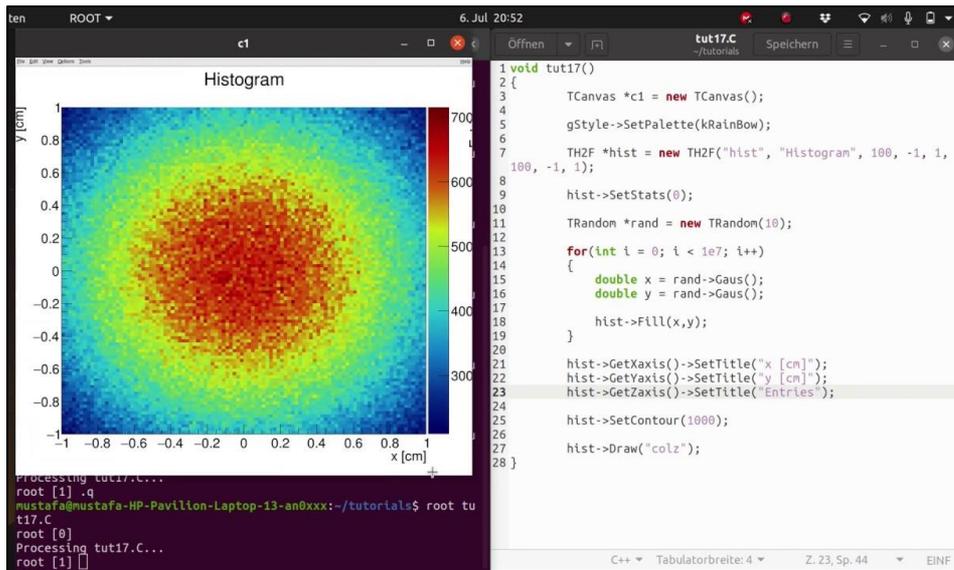
Already facing **several bottlenecks**, expected more challenging the future

- o Maintenance of code & ease of use

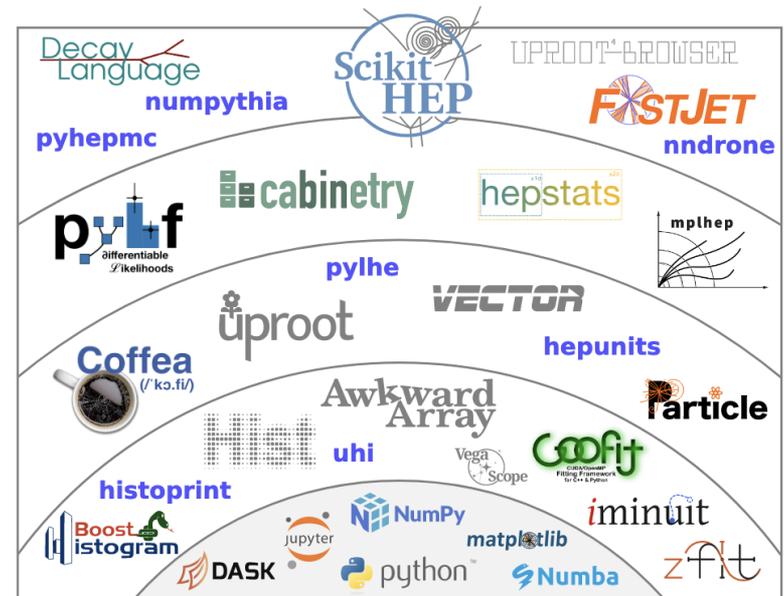


Some History of Analysis Software

- Several scientific software toolkits have been used to deal with big data processing, storage, statistical analysis and visualization
- Increasingly modular, increasingly focused on **interoperability**



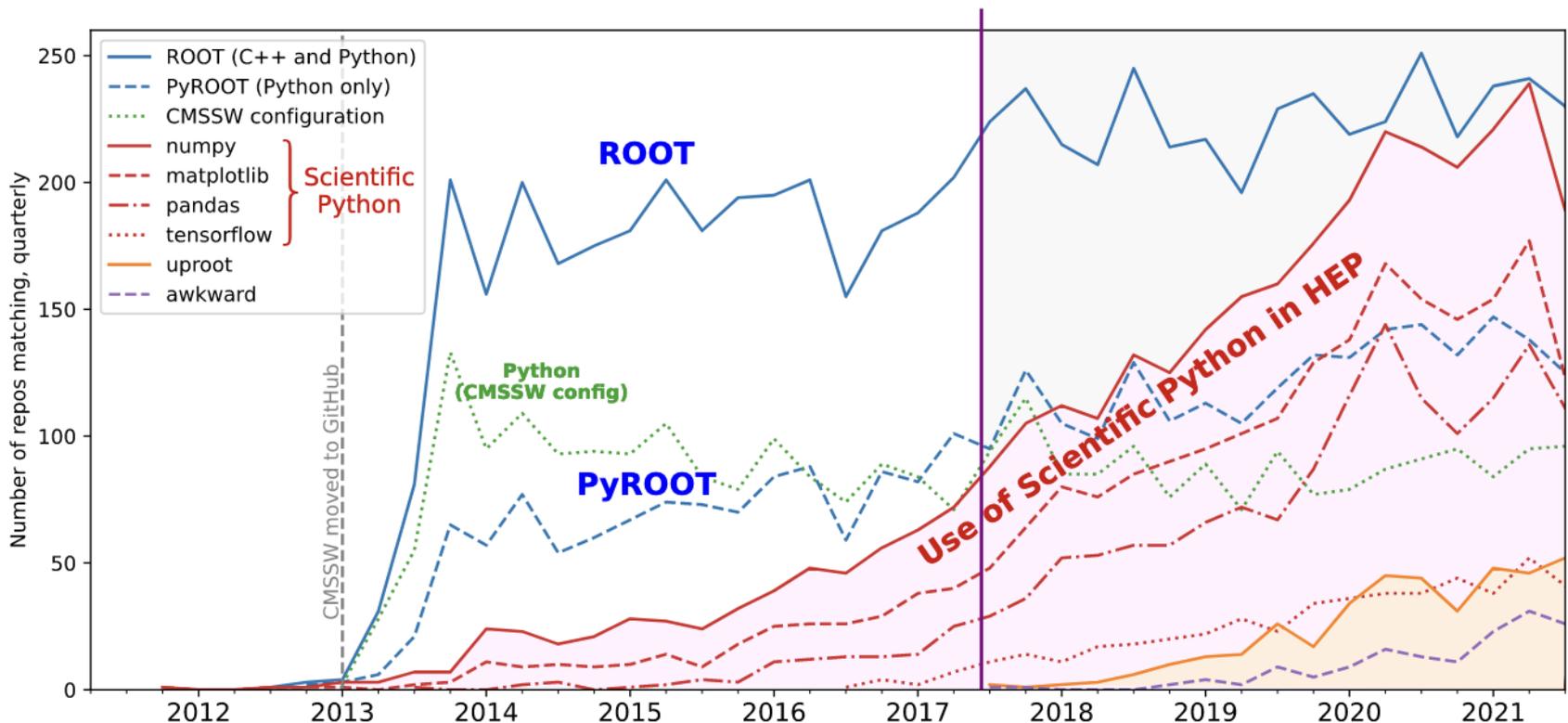
ROOT
1994-Present
C++ libraries
can interface with python, R



Python Ecosystem Tools
Python interfaces
connected to developments in AI/ML
and data science more broadly

Increasing use of Python for Analysis

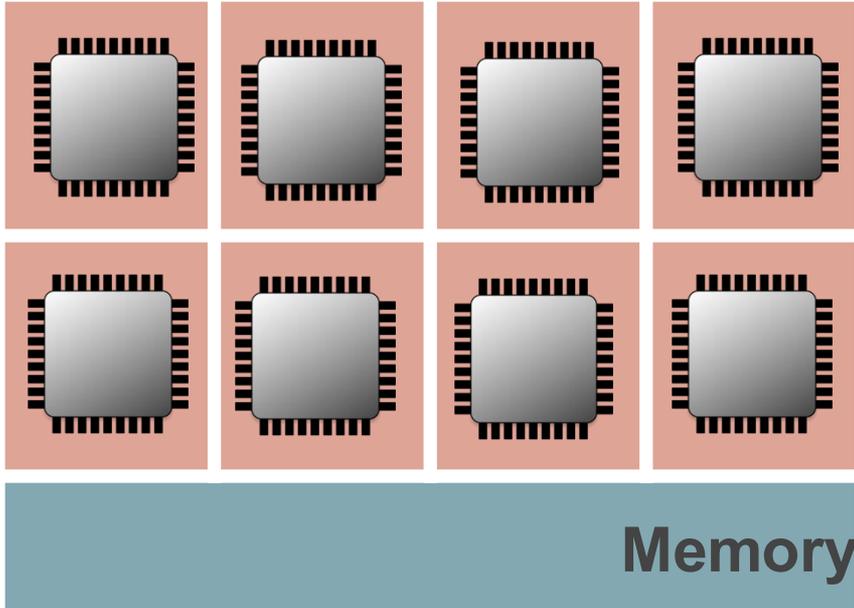
- Python has been in use for a long time for several purposes:
 - steering scripts, configuration-building, machine learning models, etc



Python is increasingly becoming a portal for analysis

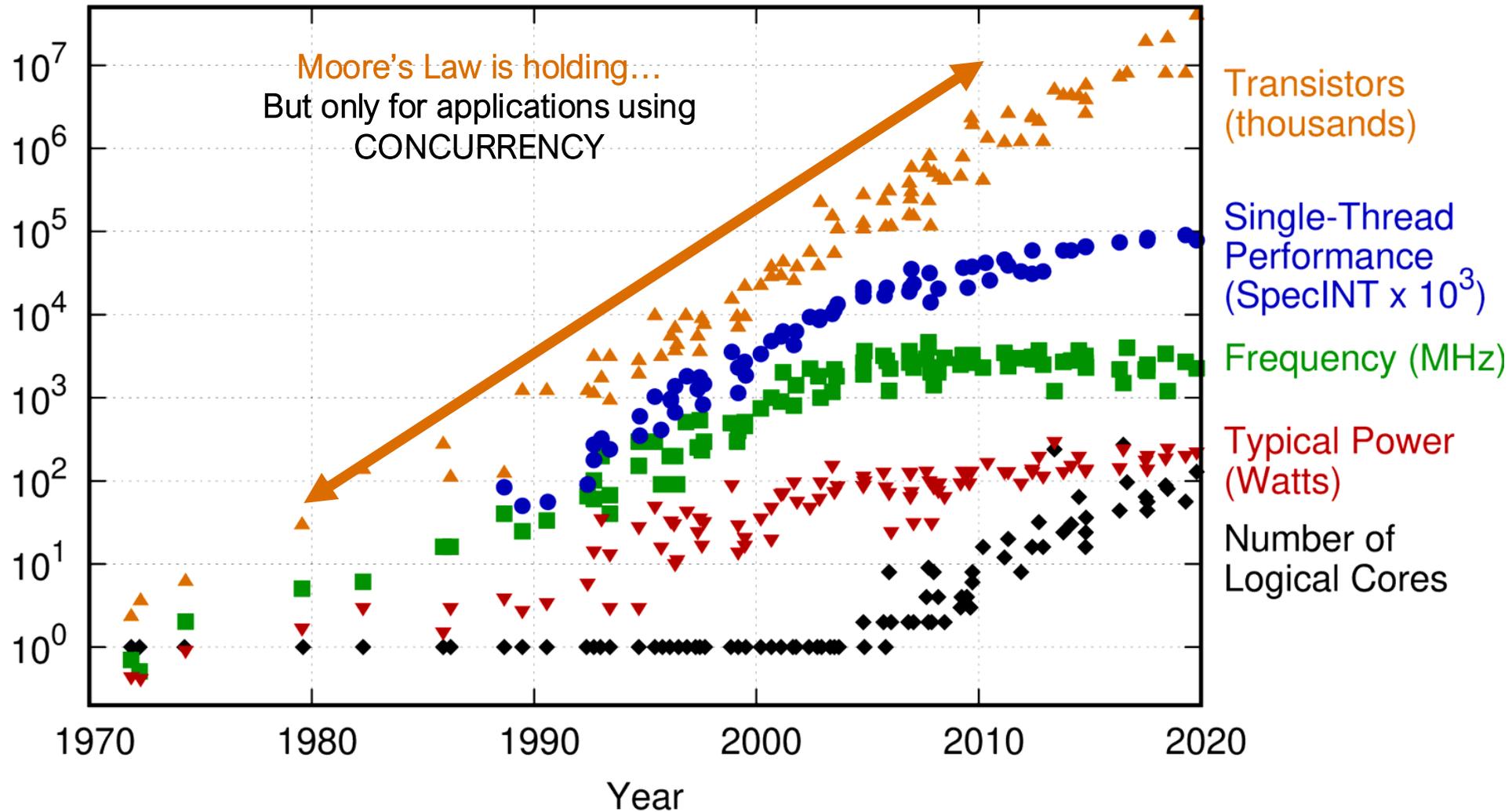
– e.g. PyROOT + Scikit-HEP

Multi-core processors



From the mid-2000s, multi-core processors became common. The cores share work between them to continue to allow increased performance.

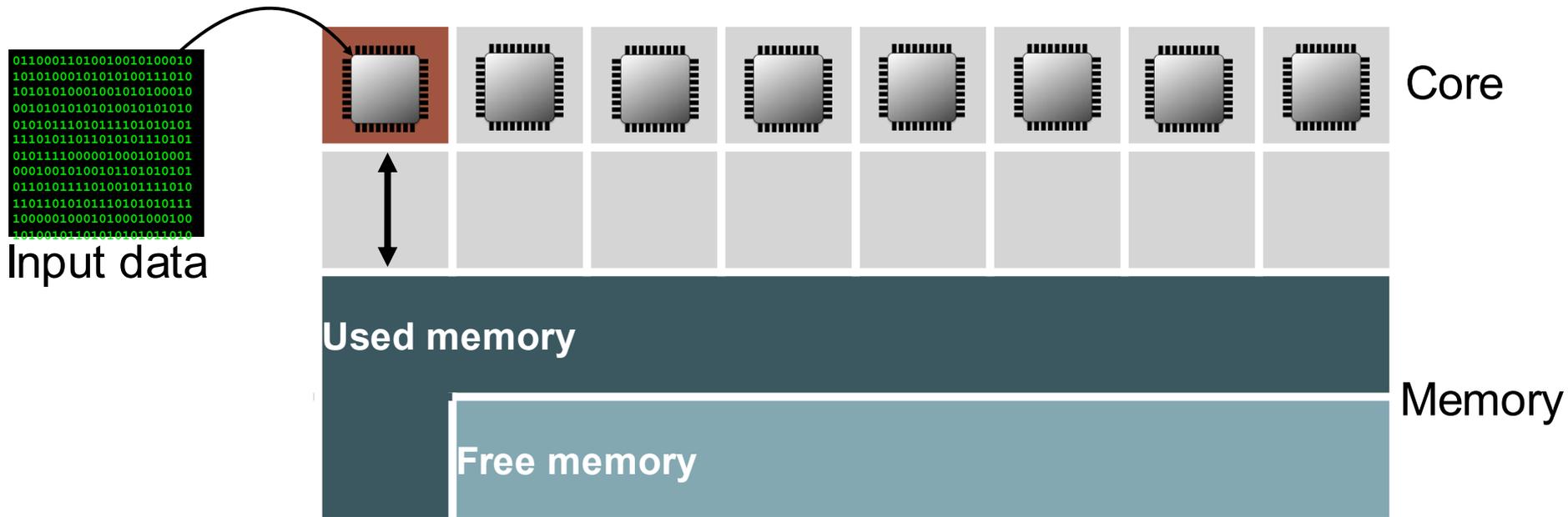
48 Years of Microprocessor Trend Data



Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2019 by K. Rupp

Types of concurrency

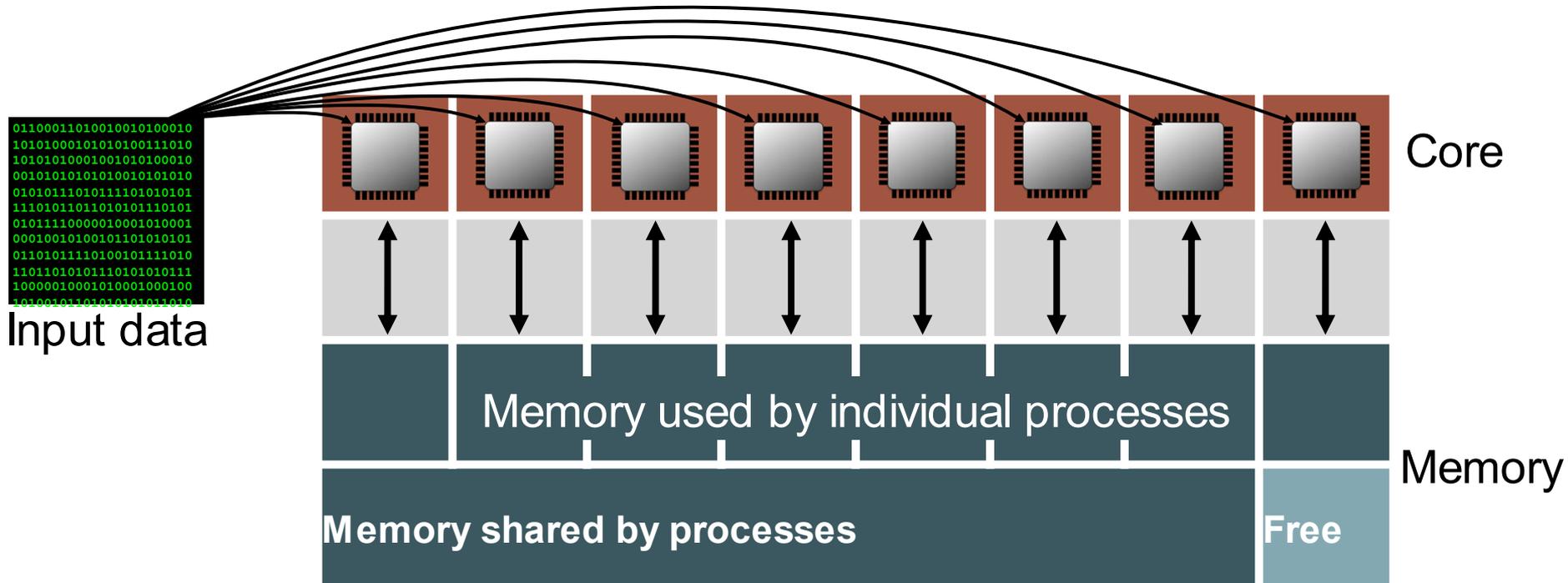
Serial (e.g. no concurrency)



If no attempt is made to share the workload, **most of the memory is used by one core and the other cores can't be used**

Types of concurrency

Multi-process

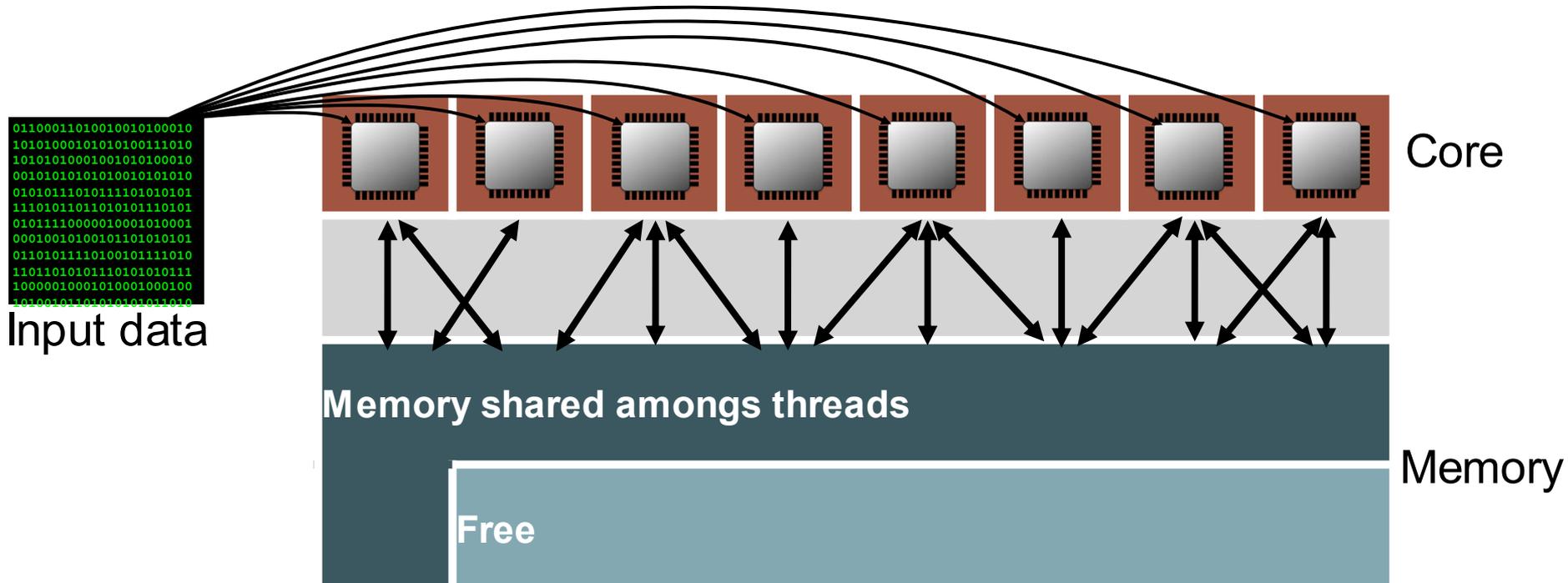


Memory needed by all processes is shared at the start of the task. Each core runs an independent process that needs its own share of memory to handle its batch of events.

Adding extra processes still adds a lot of extra memory

Types of concurrency

Multi-threaded



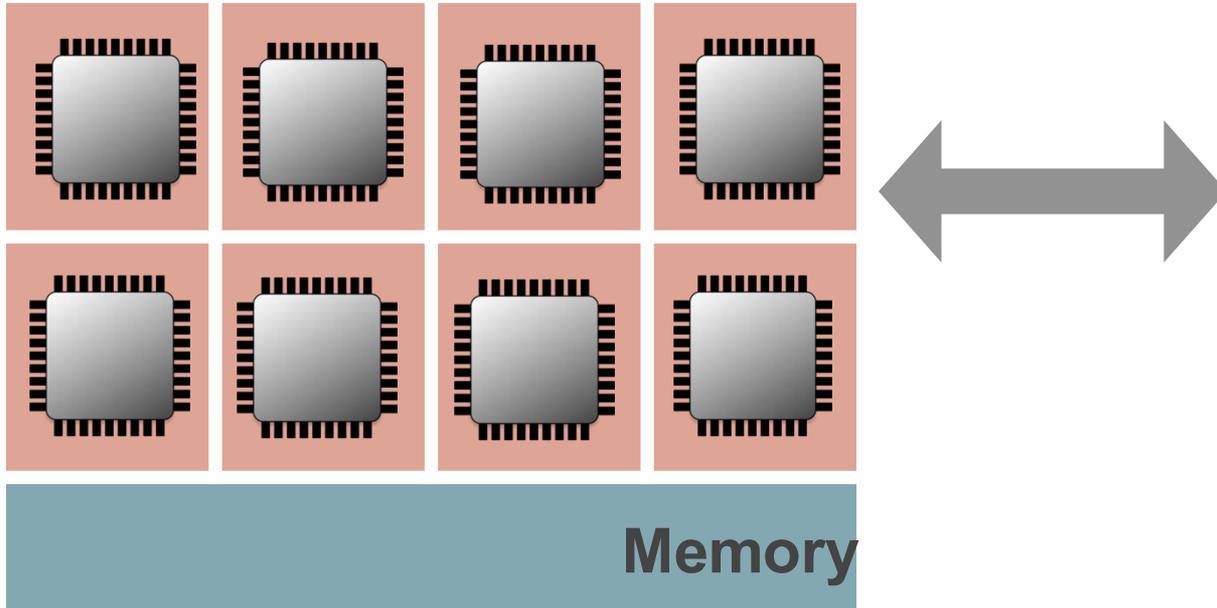
Cores can share workload & memory throughout the task processing

Adding extra cores costs very little extra memory

This ensures the software is ready for data centers with more cores and less memory per core

Meeting the HL-LHC Challenge

Graphics processing units (GPU)



CPU

Small number of high power cores
Optimized for complex serial tasks

GPU

Large number of low power cores
Optimized for massively parallel tasks
(e.g. graphics), machine learning

Memory

Future Neutrino Experiments

DUNE DEEP UNDERGROUND NEUTRINO EXPERIMENT

Sanford Underground Research Facility

Fermilab

800 miles
(1300 kilometers)

EXISTING LABS

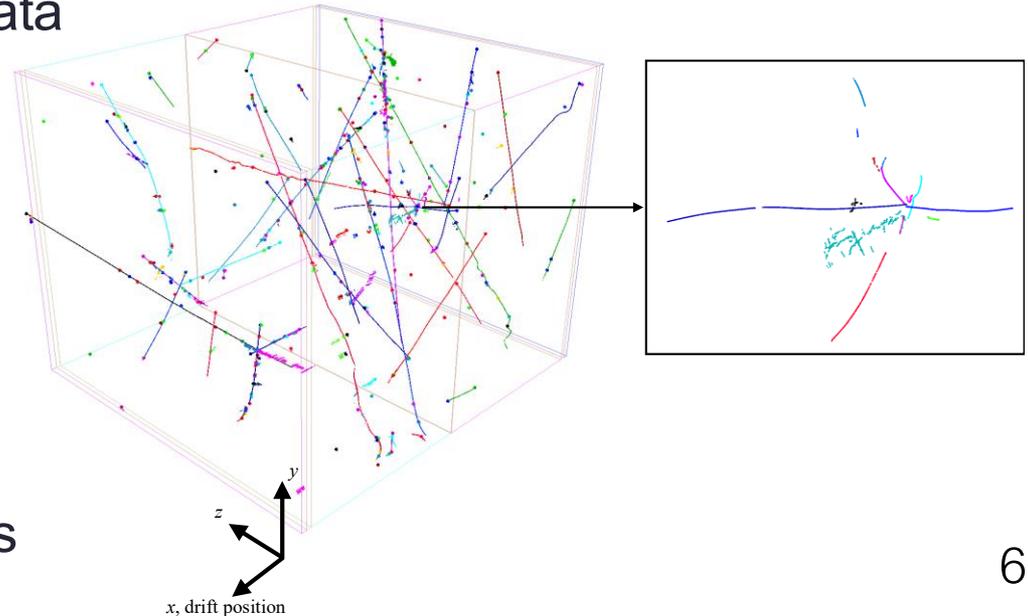
UNDERGROUND PARTICLE DETECTOR

PARTICLE DETECTOR

NEUTRINO PRODUCTION

PROTON ACCELERATOR

- DUNE computing needs include
 - Up to 30 PB/year of raw data
 - 10-15 years of running
 - 1,200 collaborators
 - Complex codes
 - Precision calibrations



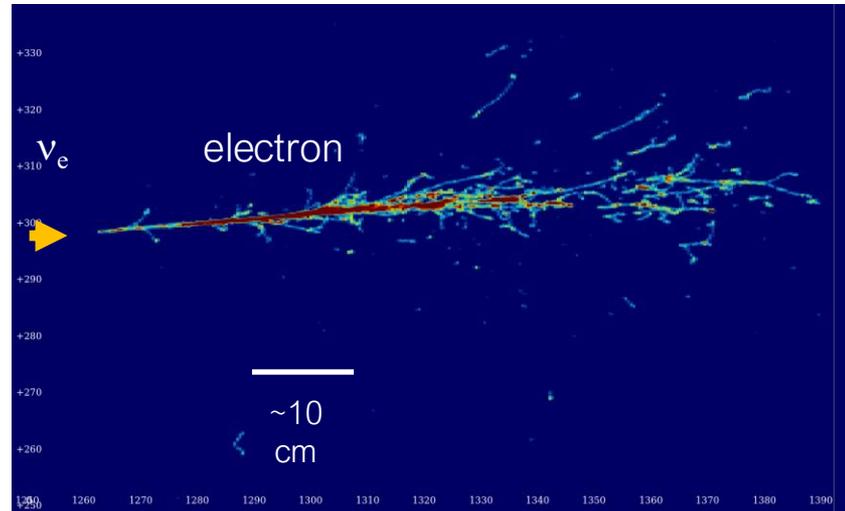
- Adopting many common solutions

Particle Detection in DUNE

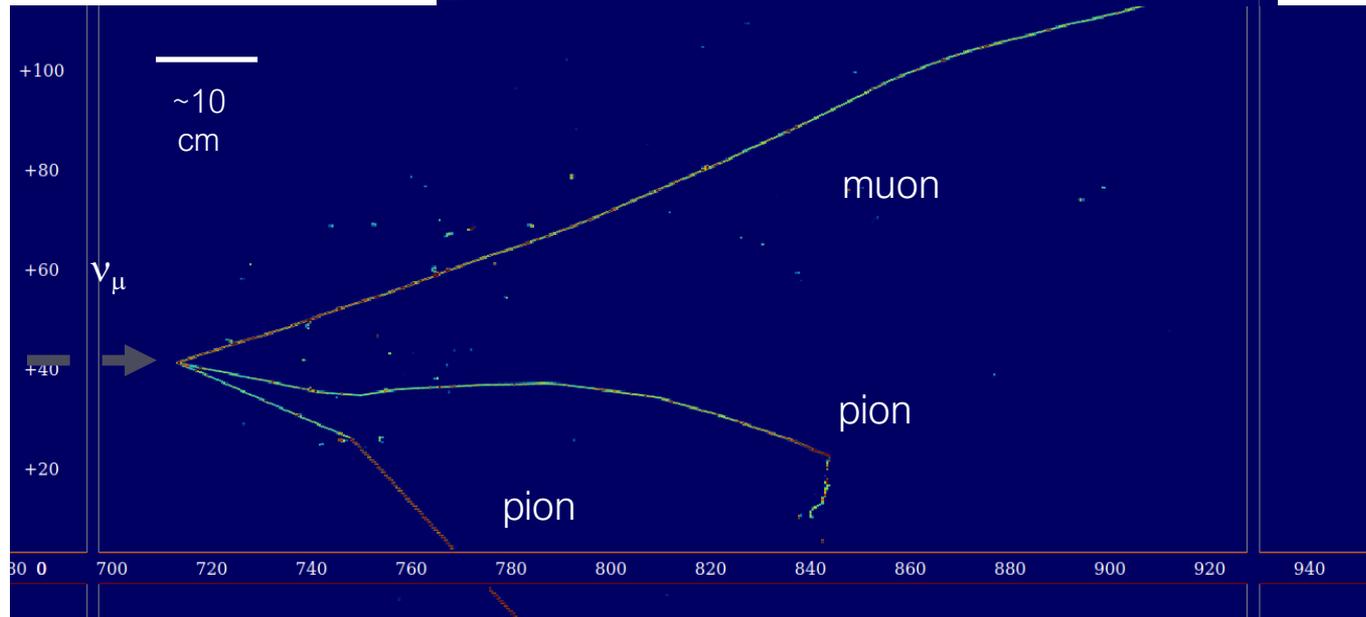


How do you tell a ν_μ from a ν_e ?

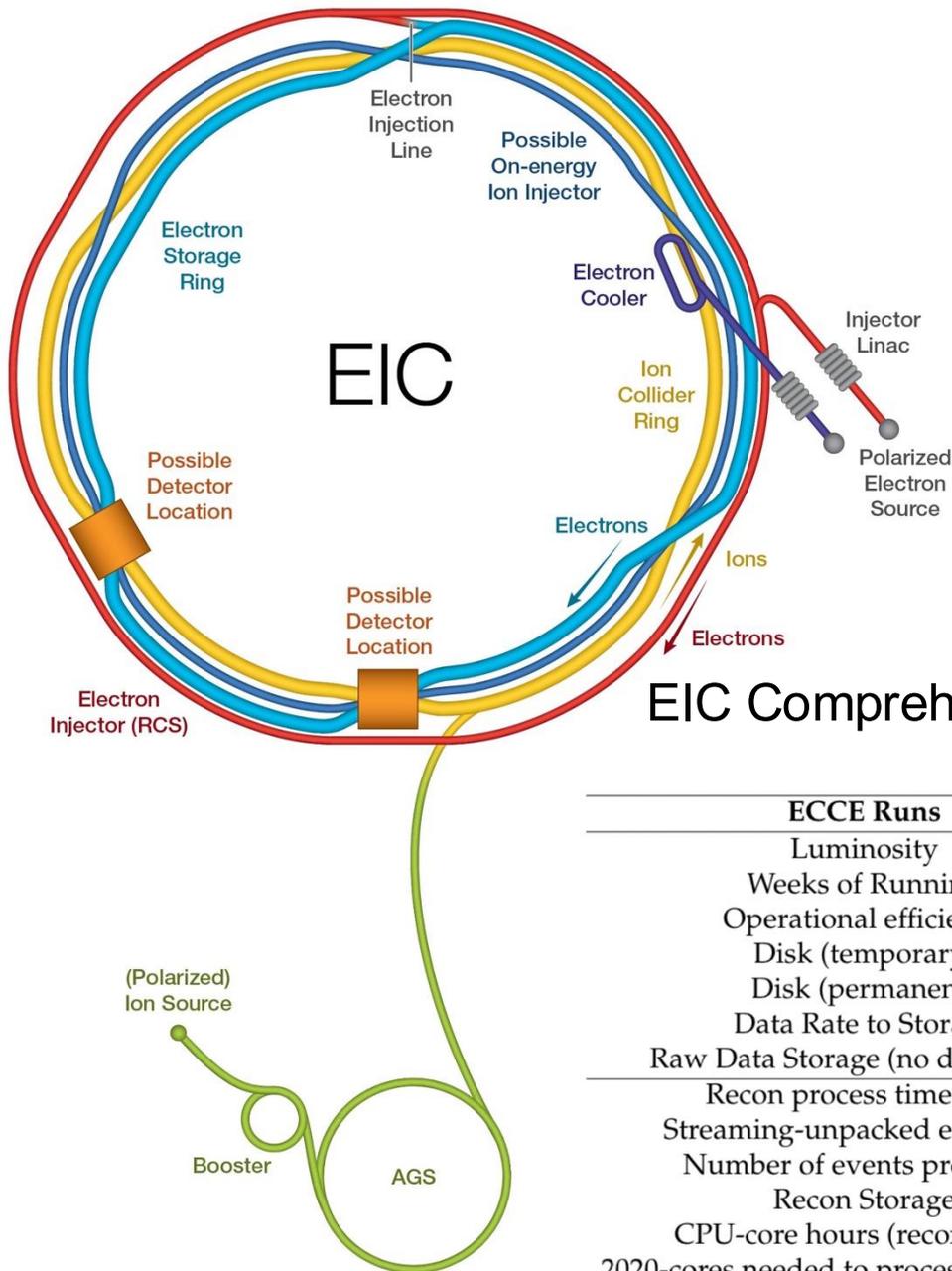
Liquid Argon detectors can distinguish muons from electrons



Challenge to instrument $\sim 50,000 \text{ m}^3$ with cm granularity and no dead material



Future Nuclear Physics Experiments



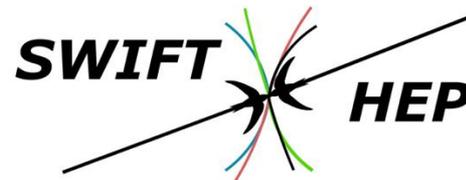
- Electron-Ion Collider plans several runs & experiments
 - Polarized electrons & protons
 - Polarized electrons & light ions
 - Electrons and heavy ions
- Significant computing needs, key goal is rapid turnaround of data for physics analysis

EIC Comprehensive Chromodynamics Experiment (ECCE)

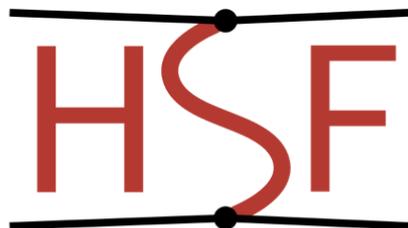
ECCE Runs	year-1	year-2	year-3
Luminosity	$10^{33} \text{cm}^{-2} \text{s}^{-1}$	$2 \times 10^{33} \text{cm}^{-2} \text{s}^{-1}$	$10^{34} \text{cm}^{-2} \text{s}^{-1}$
Weeks of Running	10	20	30
Operational efficiency	40%	50%	60%
Disk (temporary)	1.2PB	3.0PB	18.1PB
Disk (permanent)	0.4PB	2.4PB	20.6PB
Data Rate to Storage	6.7Gbps	16.7Gbps	100Gbps
Raw Data Storage (no duplicates)	4PB	20PB	181PB
Recon process time/core	5.4s/ev	5.4s/ev	5.4s/ev
Streaming-unpacked event size	33kB	33kB	33kB
Number of events produced	121 billion	605 billion	5,443 billion
Recon Storage	0.4PB	2PB	18PB
CPU-core hours (recon+calib)	191Mcore-hrs	953Mcore-hrs	8,573Mcore-hrs
2020-cores needed to process in 30 weeks	38k	189k	1,701k

Outlook on Challenges of the Future

- Future experiment needs require changes in how analysis performed in the future, including:
 - More efficient software
 - Use more machine learning / artificial intelligence methods
 - New computational technologies (e.g. GPUs)
- Opportunity to leverage developments from broader data science community & the broader physics community
 - Synergies between high energy physics, nuclear physics & astrophysics communities



Organizing the HEP community



The HEP Software Foundation facilitates cooperation and **common efforts** in High Energy Physics software and computing internationally.

- The HSF (<http://hepsoftwarefoundation.org>) was created in early 2015 as a means for organizing our community to address the software challenges of future projects such as the HL-LHC. The HSF has the following objectives:
 - Catalyze new common projects
 - Promote commonality and collaboration in new developments to make the most of limited resources
 - Provide a framework for attracting effort and support to Software & Computing projects
 - Provide a structure to set priorities and goals for work in common projects

HSF-India Project



HSF-India is a 5 year project funded by the US National Science Foundation that aims to build international research software collaborations between US, European, & India based researchers to reach the science goals of experimental particle, nuclear & astroparticle research

<https://research-software-collaborations.org/>

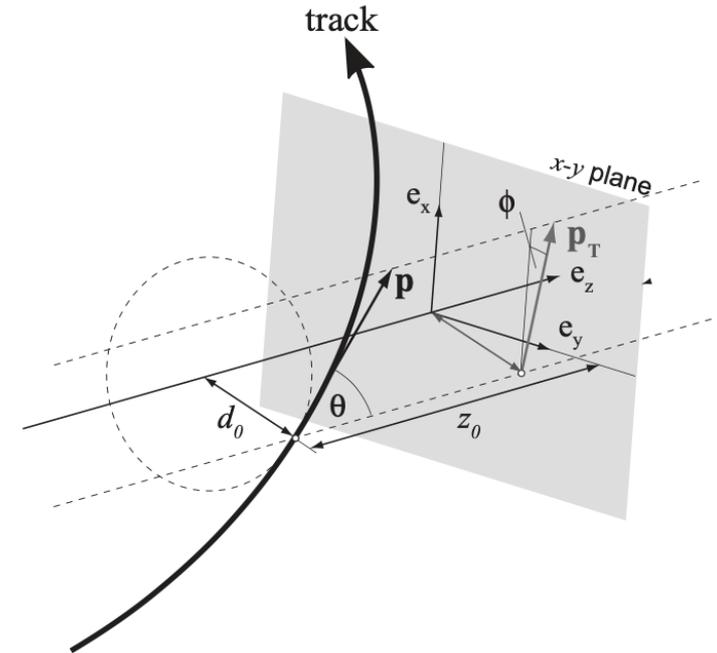
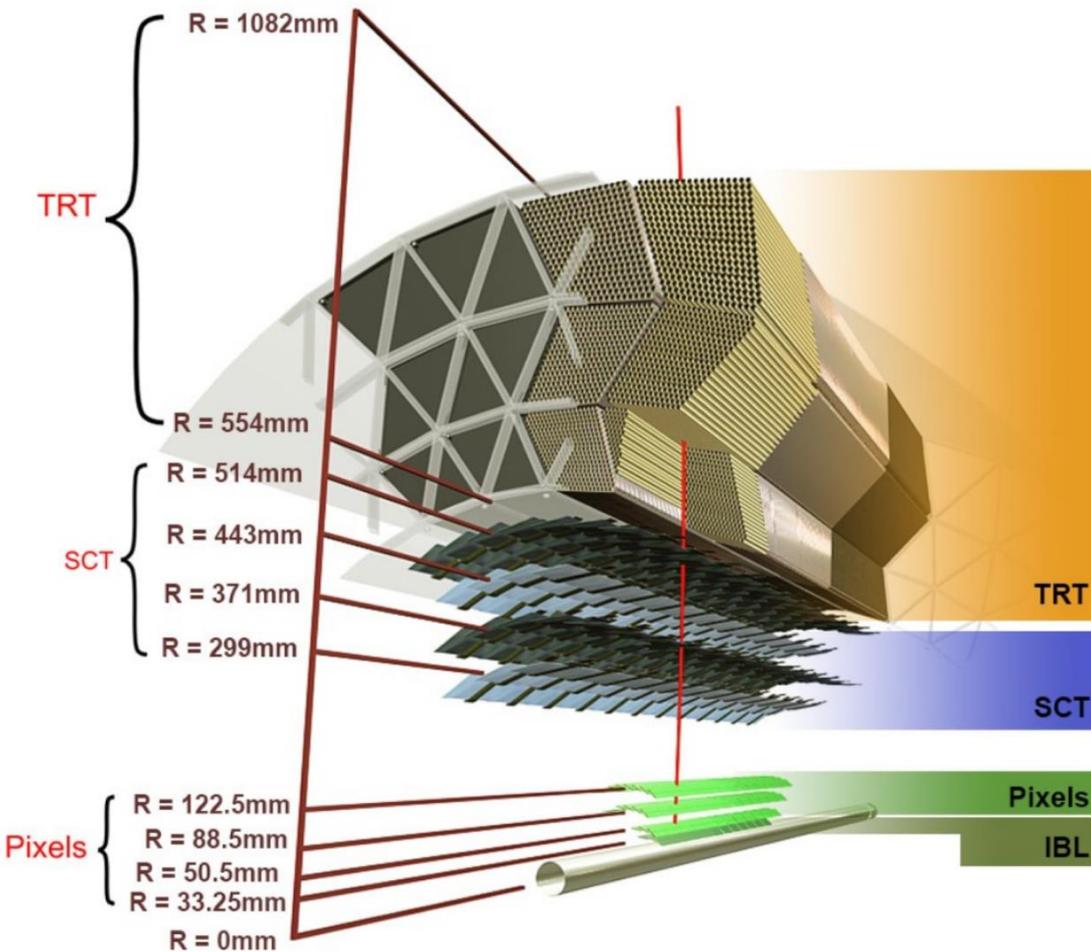
- Given the growing complexity of our scientific data and collaborations, software collaborations are increasingly important to raise the collective productivity of our research community
- Intended as a long-term investment in international team science
- Funding available for
 - Fellowships
 - Researcher exchanges
 - Training events
 - including this event!



Princeton University: Peter Elmer, David Lange (PI)
University of Massachusetts, Amherst: Rafael Coelho
Lopes de Sa, Verena Martinez Outschoorn

Additional Slides

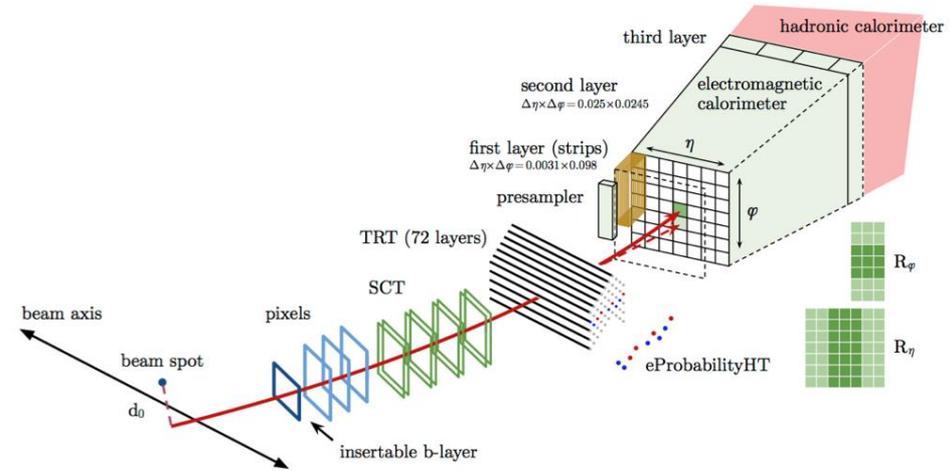
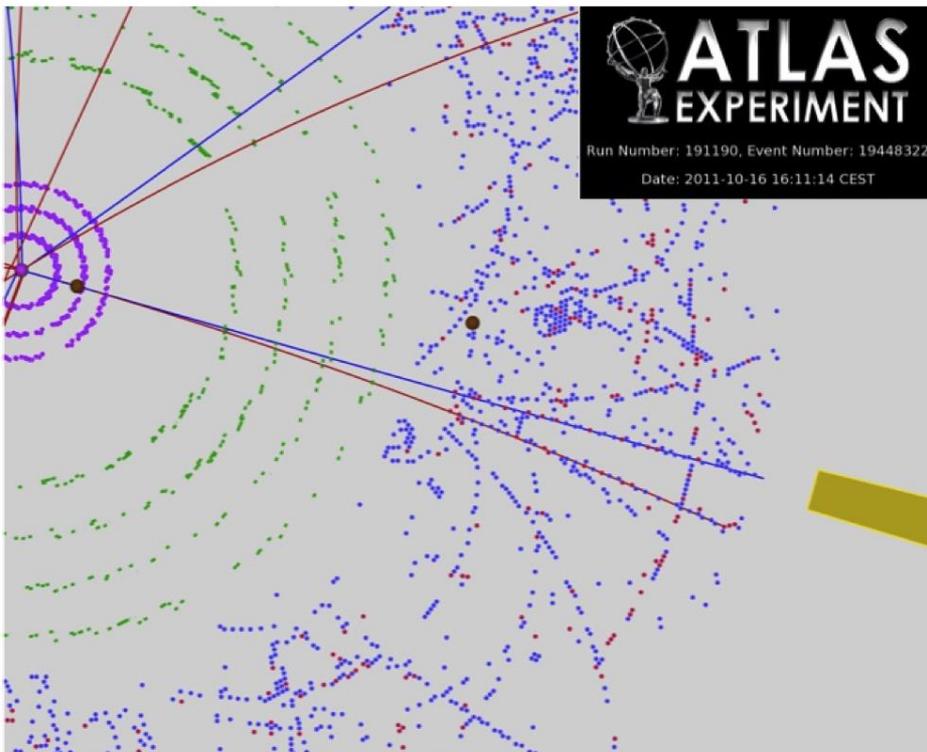
A bit more on inner tracker reconstruction



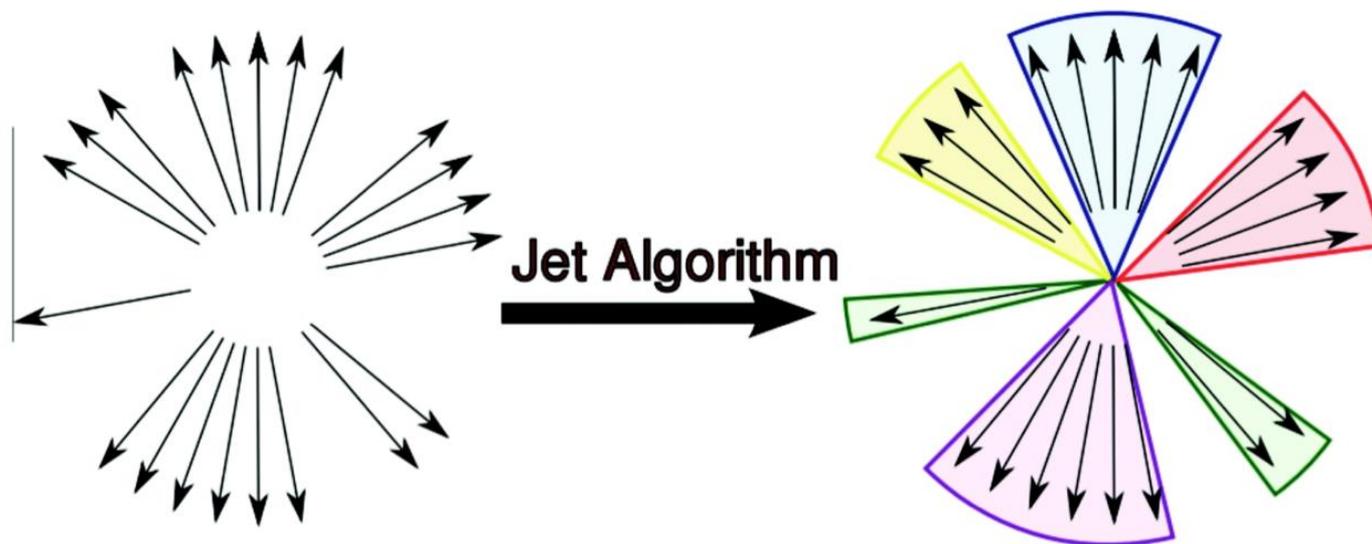
Global track parameters e.g. wrt. perigee

$$\left(d_0, z_0, \phi, \theta, \frac{q}{p} \right)$$

A bit more on electron reconstruction

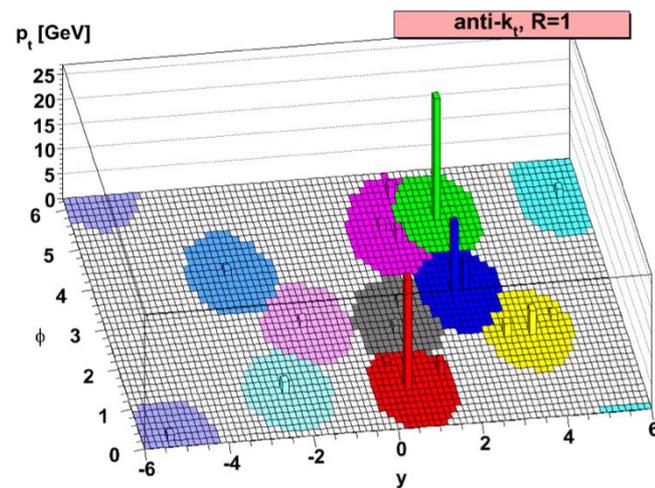
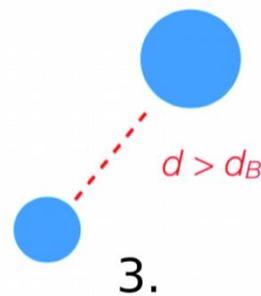
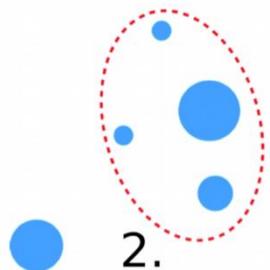
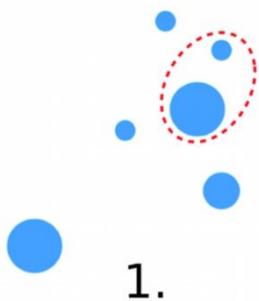


A bit more on jet reconstruction

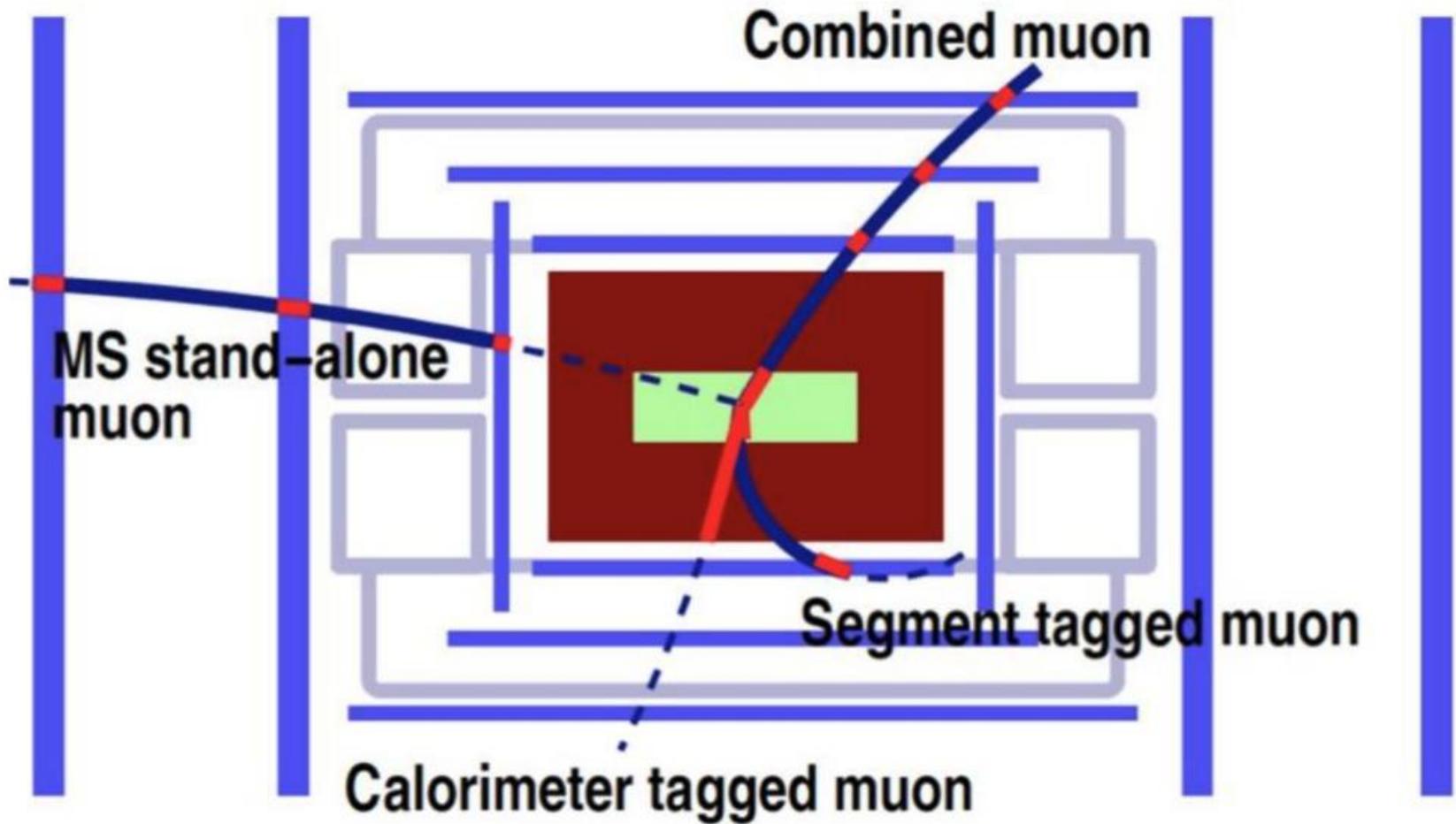


$$d = \min(p_{T,i}^{-1}, p_{T,j}^{-1}) \frac{\Delta_{i,j}^2}{R^2}$$

$$d_B = p_{T,i}^{-2}$$



A bit more on muon reconstruction



Analysis at the HL-LHC

- Analysis dataset size will increase substantially → challenge to process samples in a timely way
 - the time to process samples are a bottleneck, it is increasingly taking longer to carry out analysis
 - want to improve in the future reducing the processing time & using better tools
 - analyst time is critical

	<u>LHC (Run 1&2)</u>		<u>HL-LHC (Run 4+)</u>
Analysis Dataset size	10 TB		1,000 TB
Target Scan Turnaround time	Weeks		Hours
Analysis team size (physicists)	5-10		< 5
Primary analysis resource	Laptop		Analysis Facility