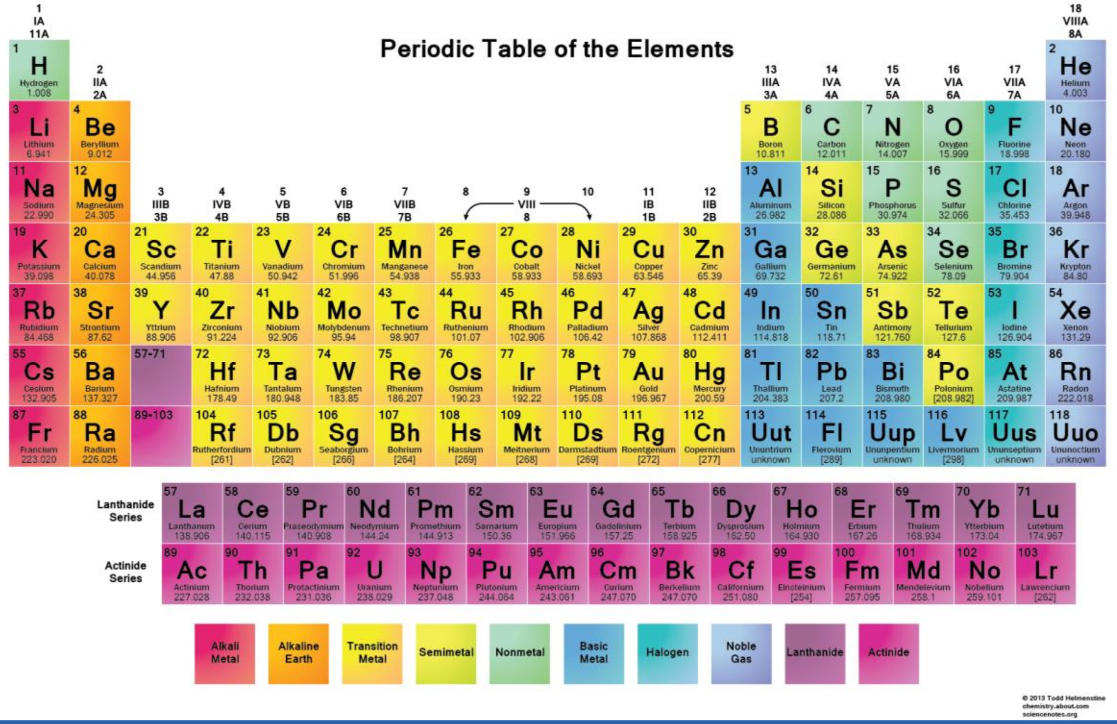
Bits et particules:

comment les FPGA ont capturé le boson de Higgs et ciblent désormais la matière noire

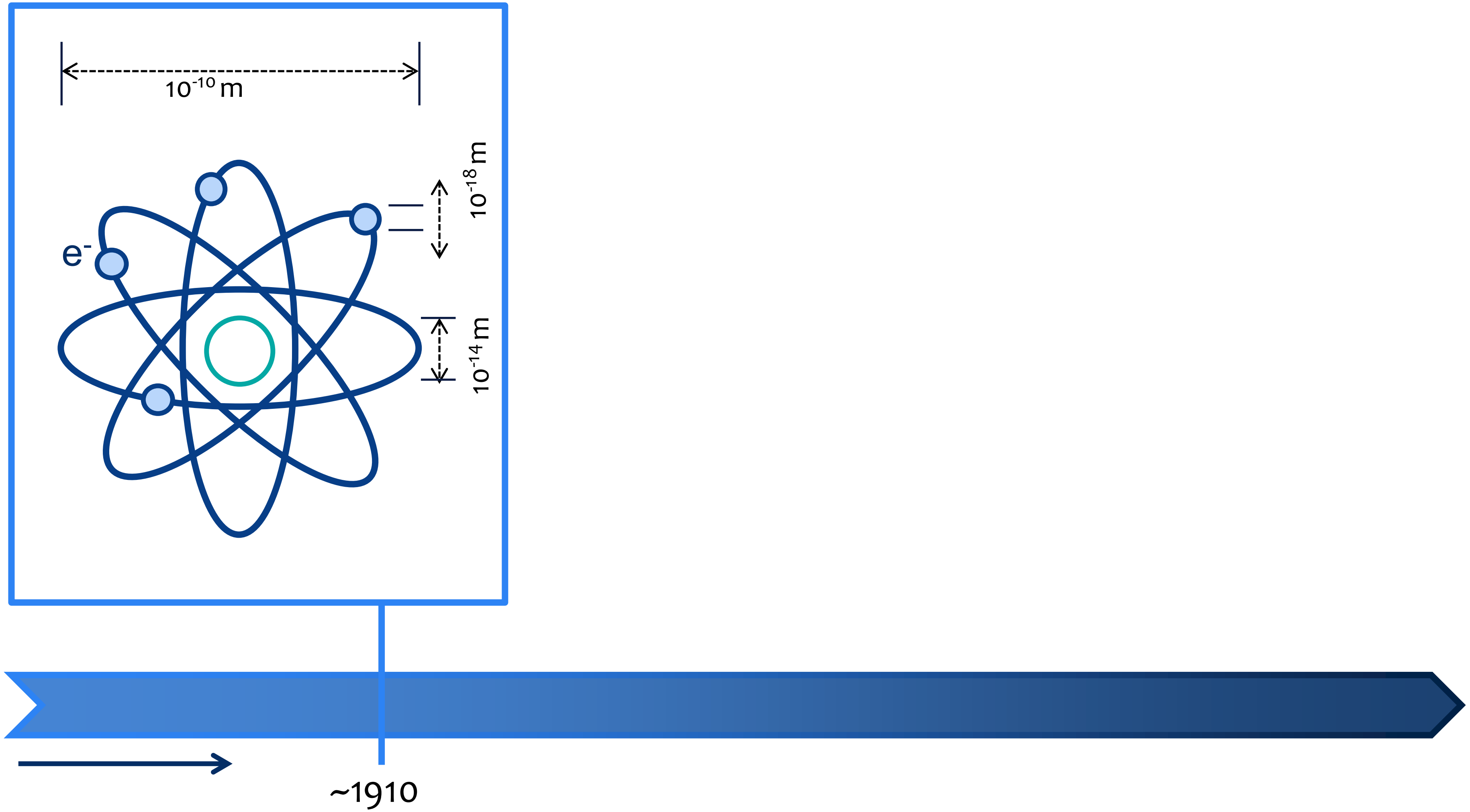
Anna Sfyrla (Univ. of Geneva)

Matière

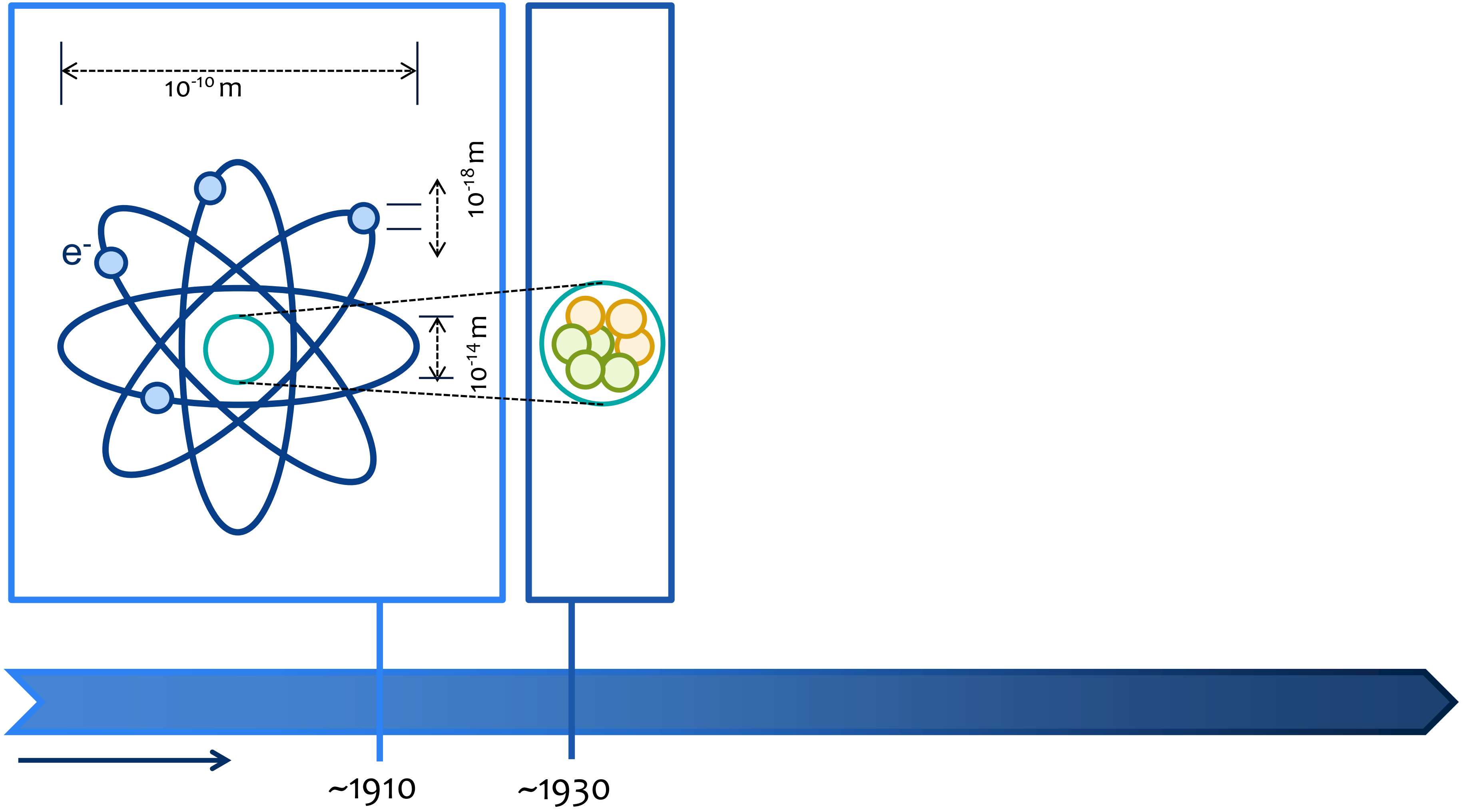


~400 BC 0 AC ~1870

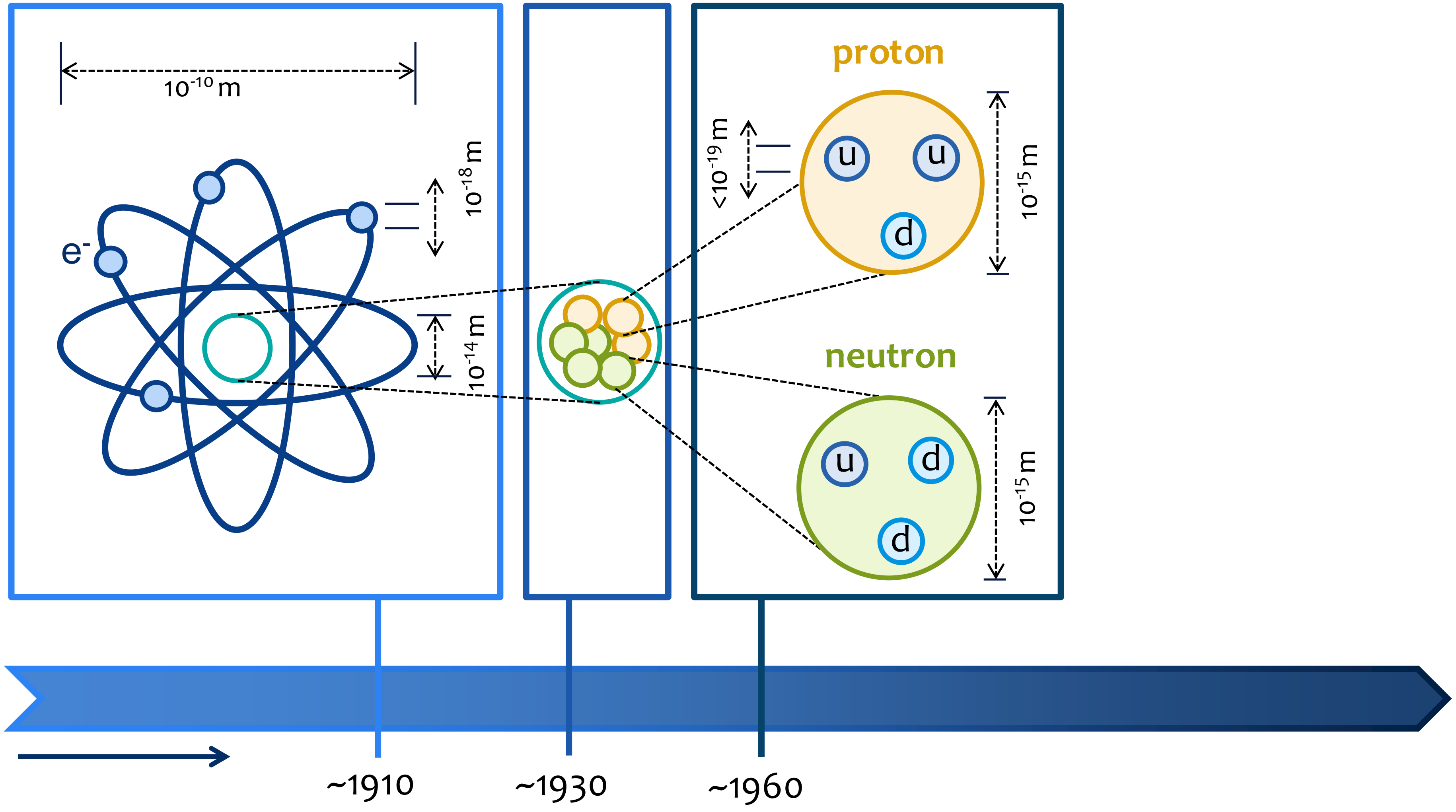
Matière



Matière

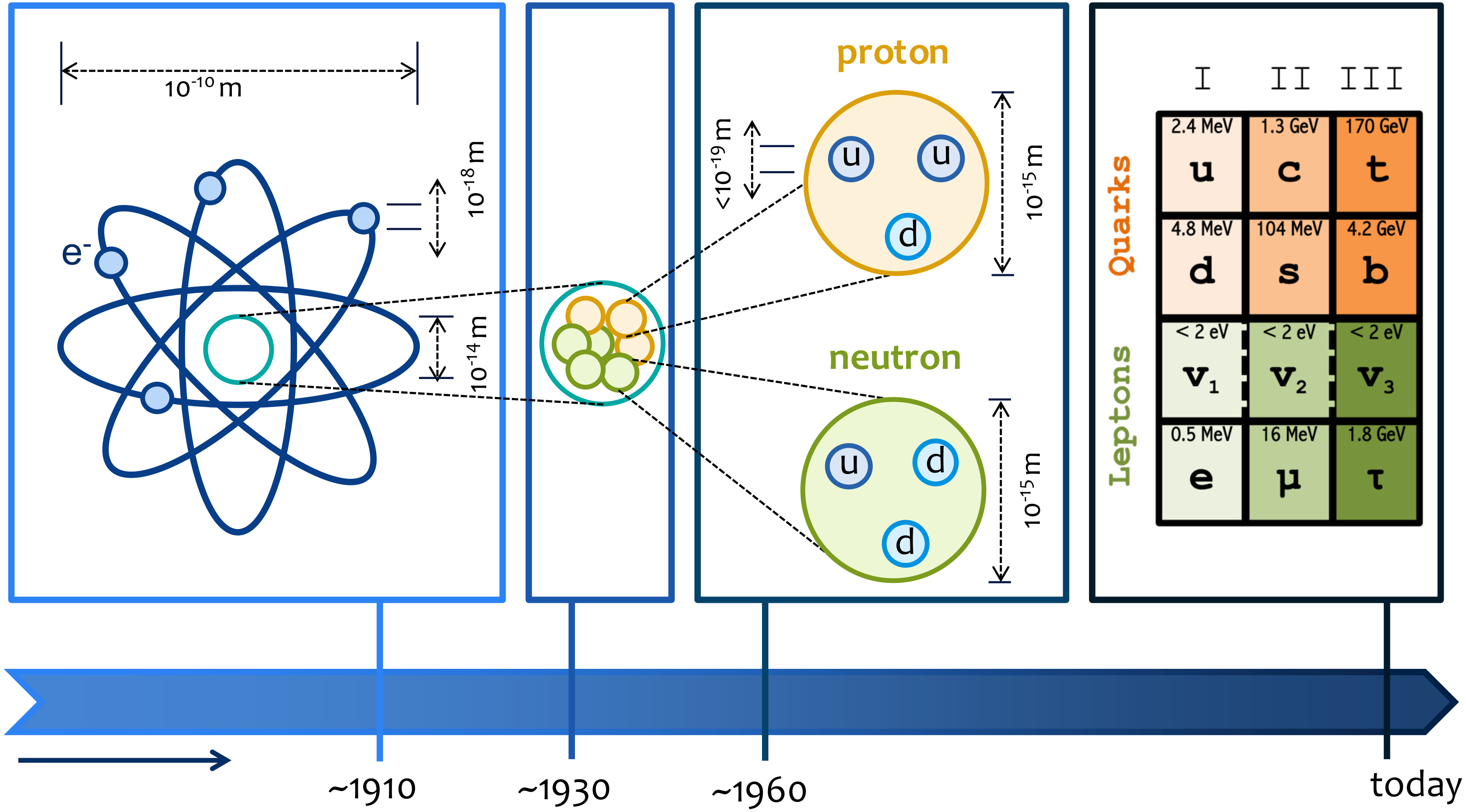


Matière



10^{-10} m
e⁻
10^{-18} m
10^{-14} m
proton
u
u
d
$<10^{-19}$ m
10^{-15} m
neutron
u
d
d
10^{-15} m
~1910
~1930
~1960

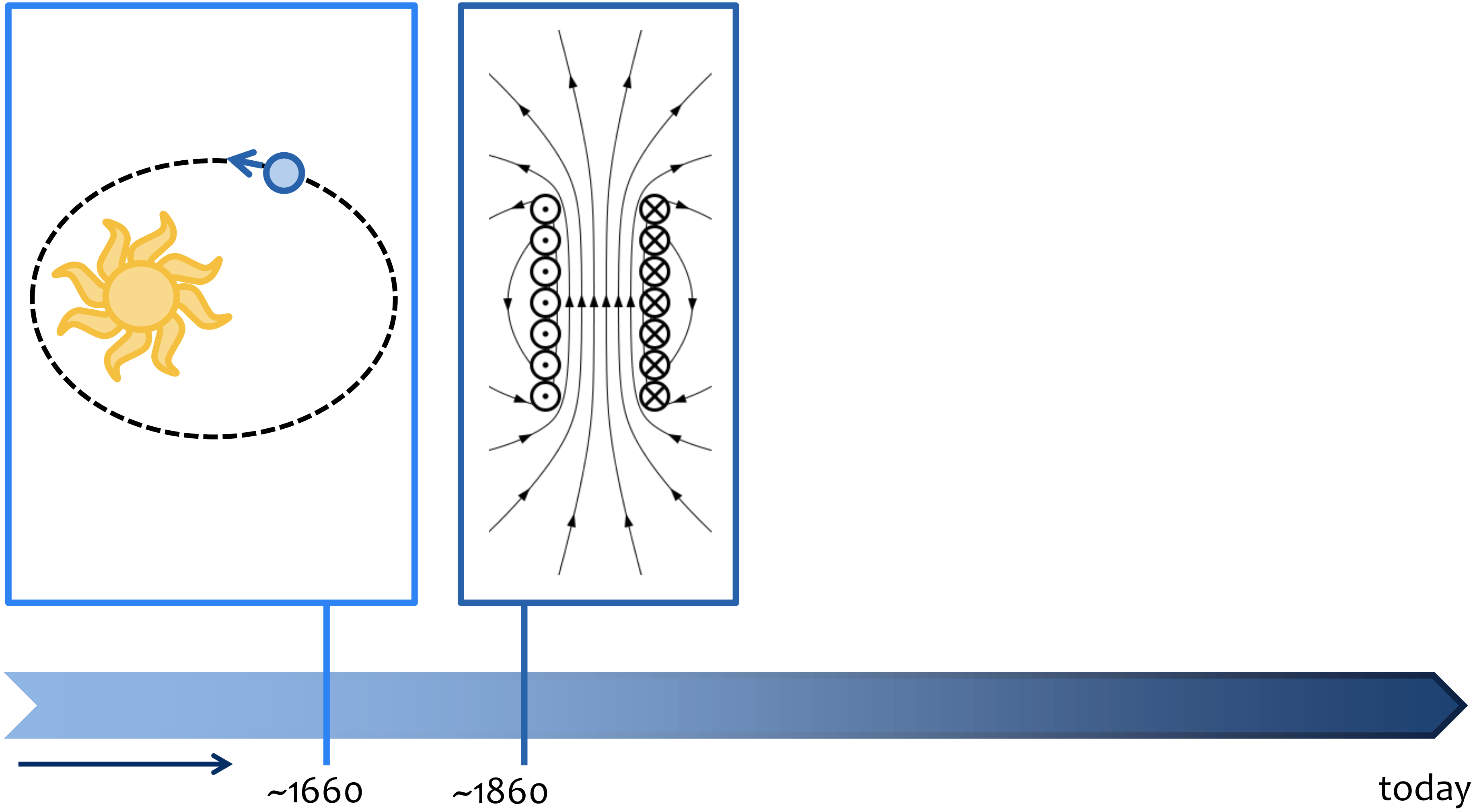
Matière



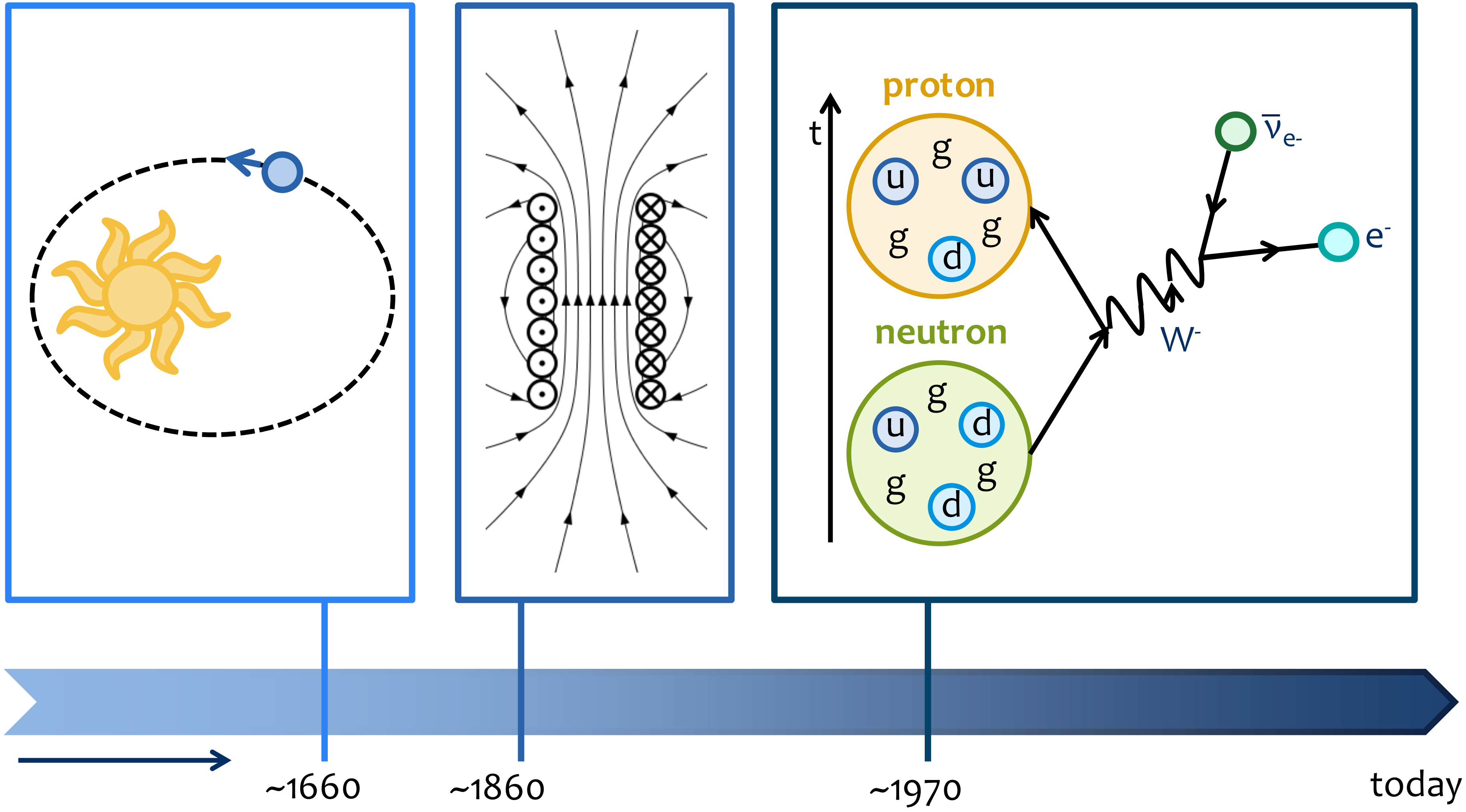
10⁻¹⁰ m
e⁻
10⁻¹⁸ m
10⁻¹⁴ m
proton
<10⁻¹⁹ m
10⁻¹⁵ m
u
u
d
neutron
u
d
d
10⁻¹⁵ m
I
II
III
Quarks
Leptons
2.4 MeV
u
1.3 GeV
c
170 GeV
t
4.8 MeV
d
104 MeV
s
4.2 GeV
b
< 2 eV
ν1
< 2 eV
ν2
< 2 eV
ν3
0.5 MeV
e
16 MeV
μ
1.8 GeV
τ
~1910
~1930
~1960
today

Forces

Forces

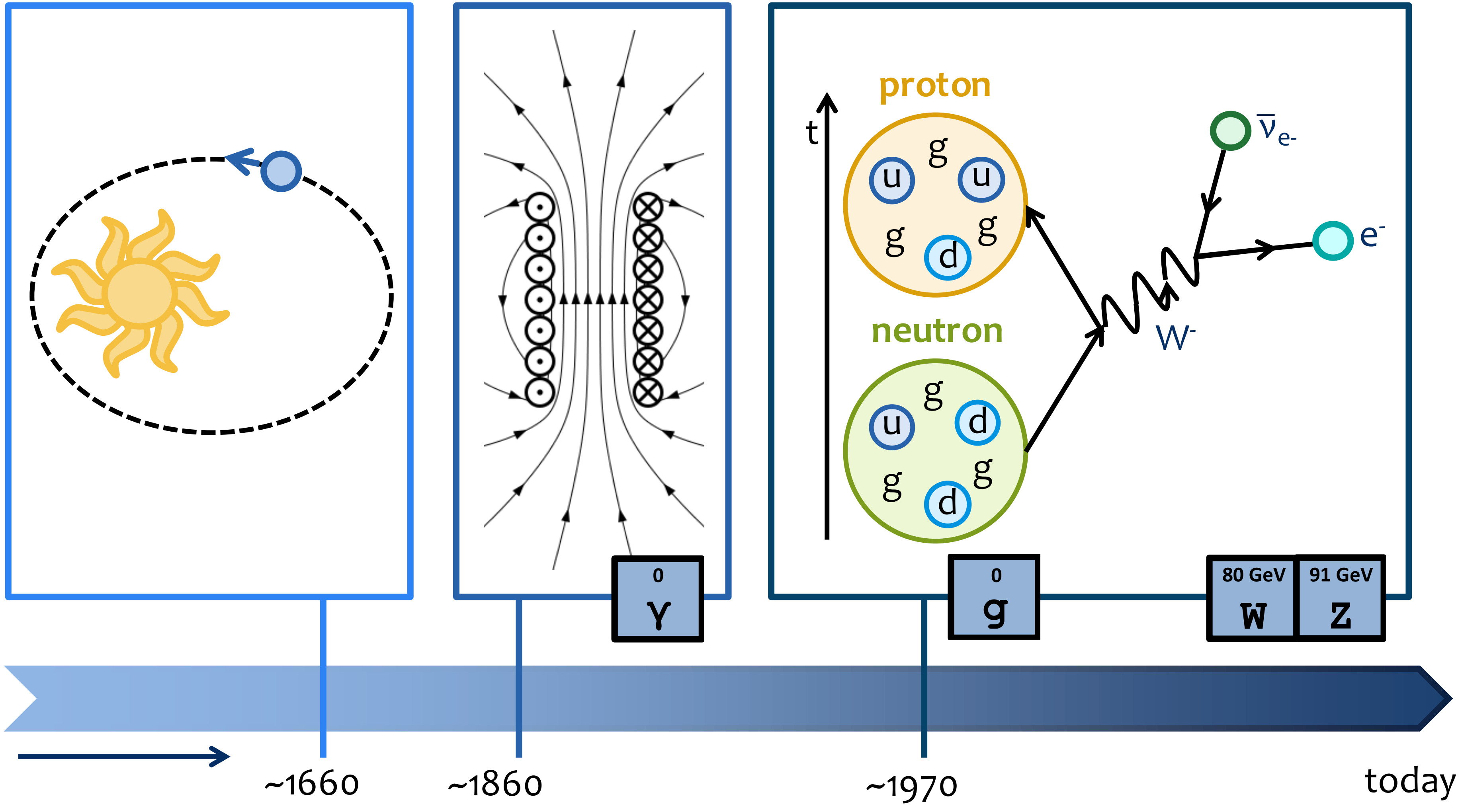


Forces

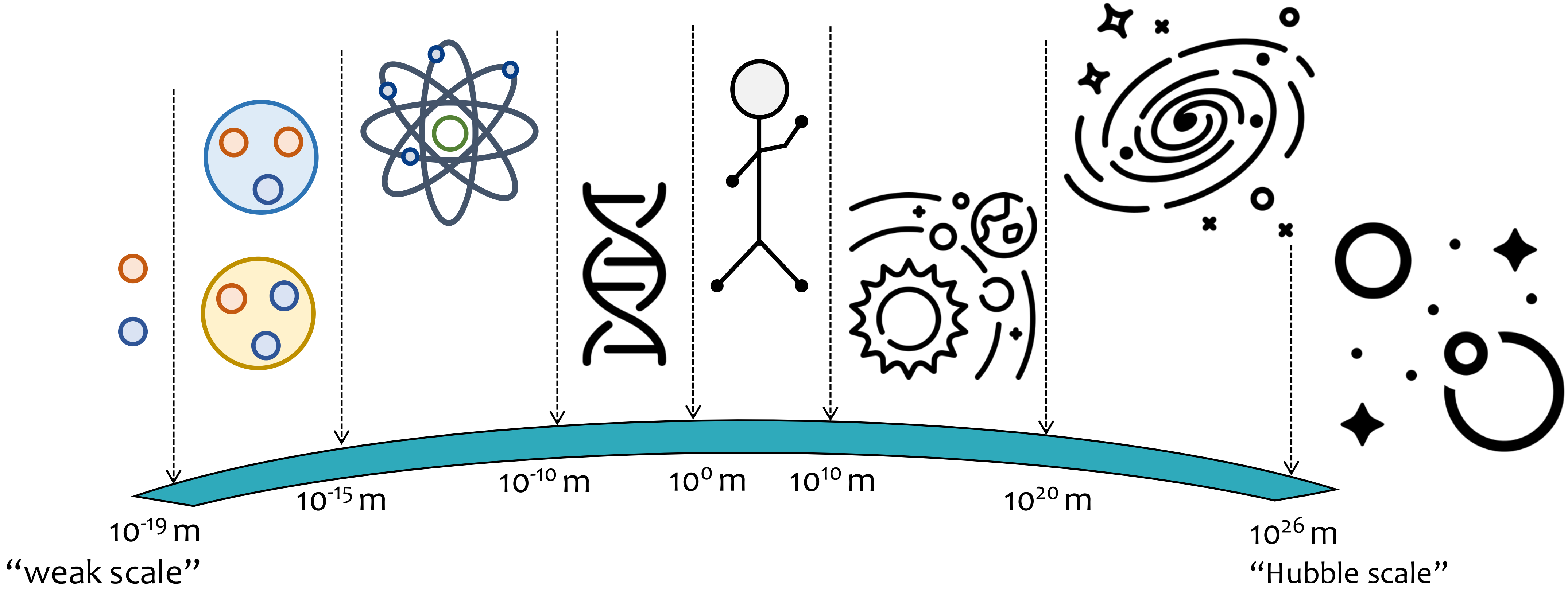


Forces

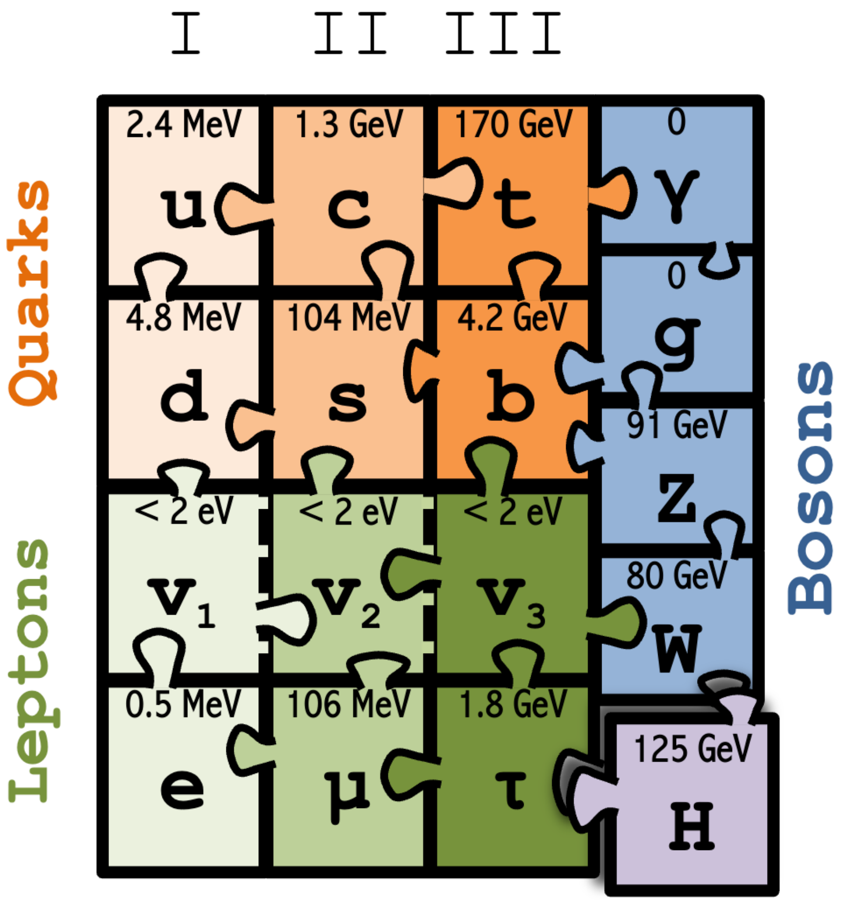
Force carriers: Bosons

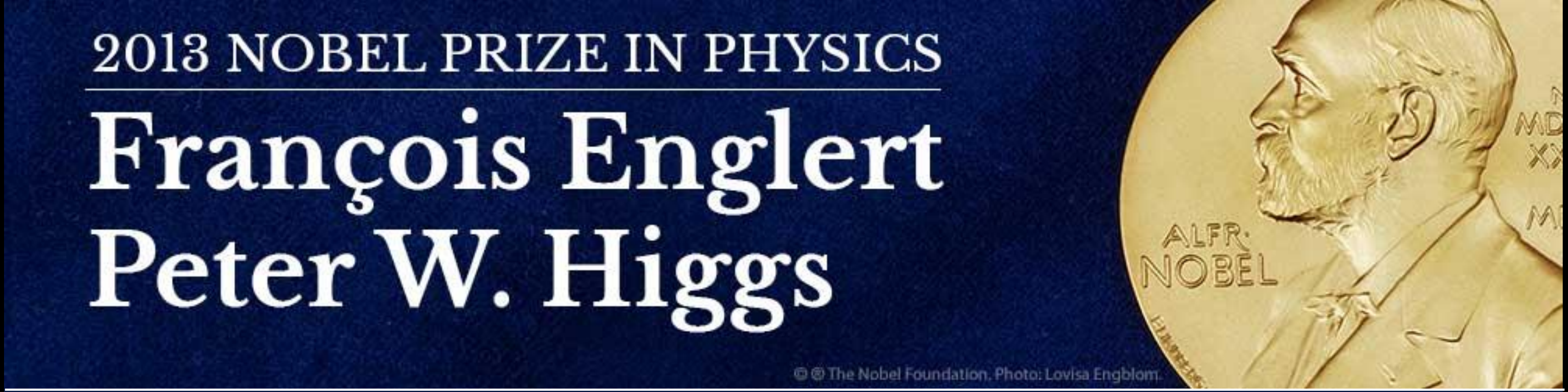


Échelle des Forces



Le modèle standard





Le prix Nobel de physique 2013 a été décerné conjointement à François Englert et Peter W. Higgs

« pour la découverte théorique d'un mécanisme qui contribue à notre compréhension de l'origine de la masse des particules subatomiques, et qui a récemment été confirmé par la découverte de la particule fondamentale prédite, grâce aux expériences ATLAS et CMS menées au Grand collisionneur de hadrons du CERN ».

Le boson de Higgs, c’est quoi ?

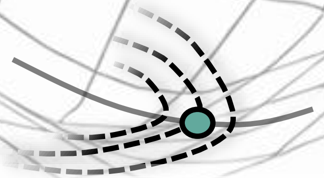
Le boson de Higgs = la signature observable du champ de Higgs

Le champ de Higgs = un champ invisible present partout dans l’éspace

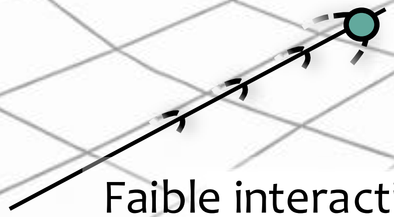
Les particules qui interagissent avec ce champ acquièrent une masse

Toutes les particules n’interagissent pas de la même façon

Plus l’interaction est forte, plus la masse est grande



Forte interaction: grande masse



Faible interaction: petite masse

Le boson de Higgs, c'est quoi ?

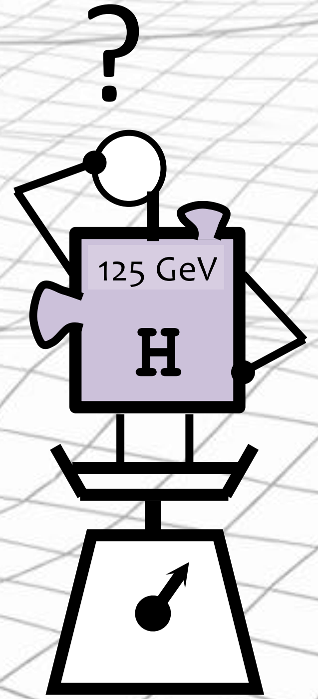
Le boson de Higgs = la signature observable du champ de Higgs

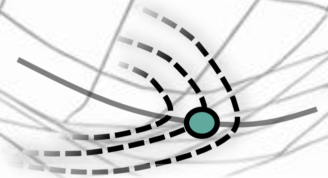
Le champ de Higgs = un champ invisible present partout dans l'éspace

Les particules qui interagissent avec ce champ acquièrent une masse

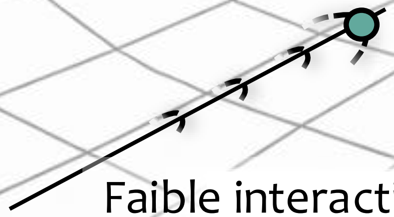
Toutes les particules n'interagissent pas de la même façon

Plus l'interaction est forte, plus la masse est grande



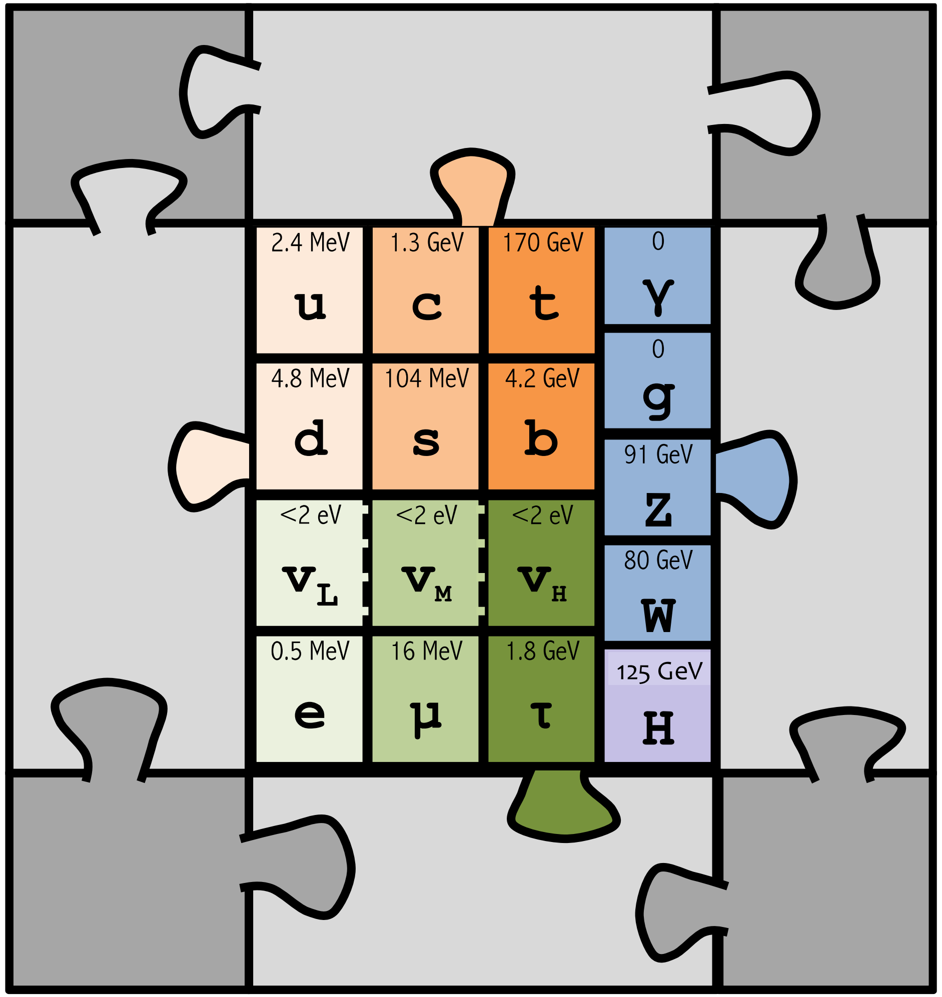


Forte interaction: grande masse



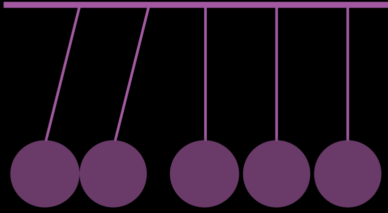
Faible interaction: petite masse

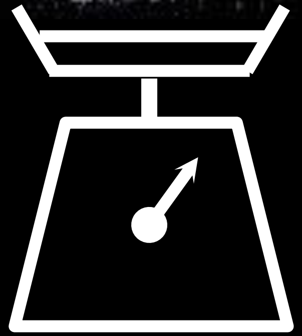
…juste une pièce du puzzle cosmique

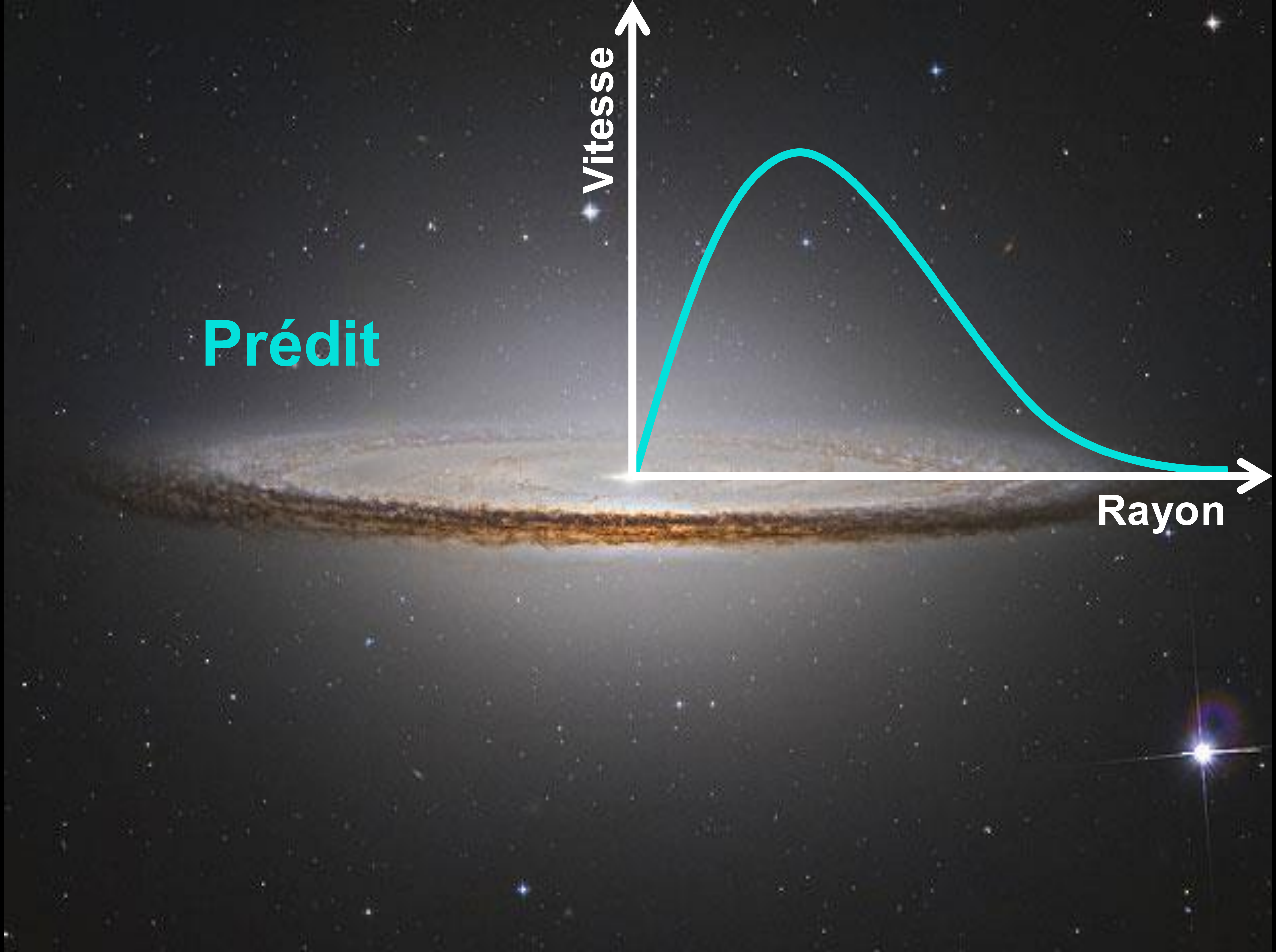


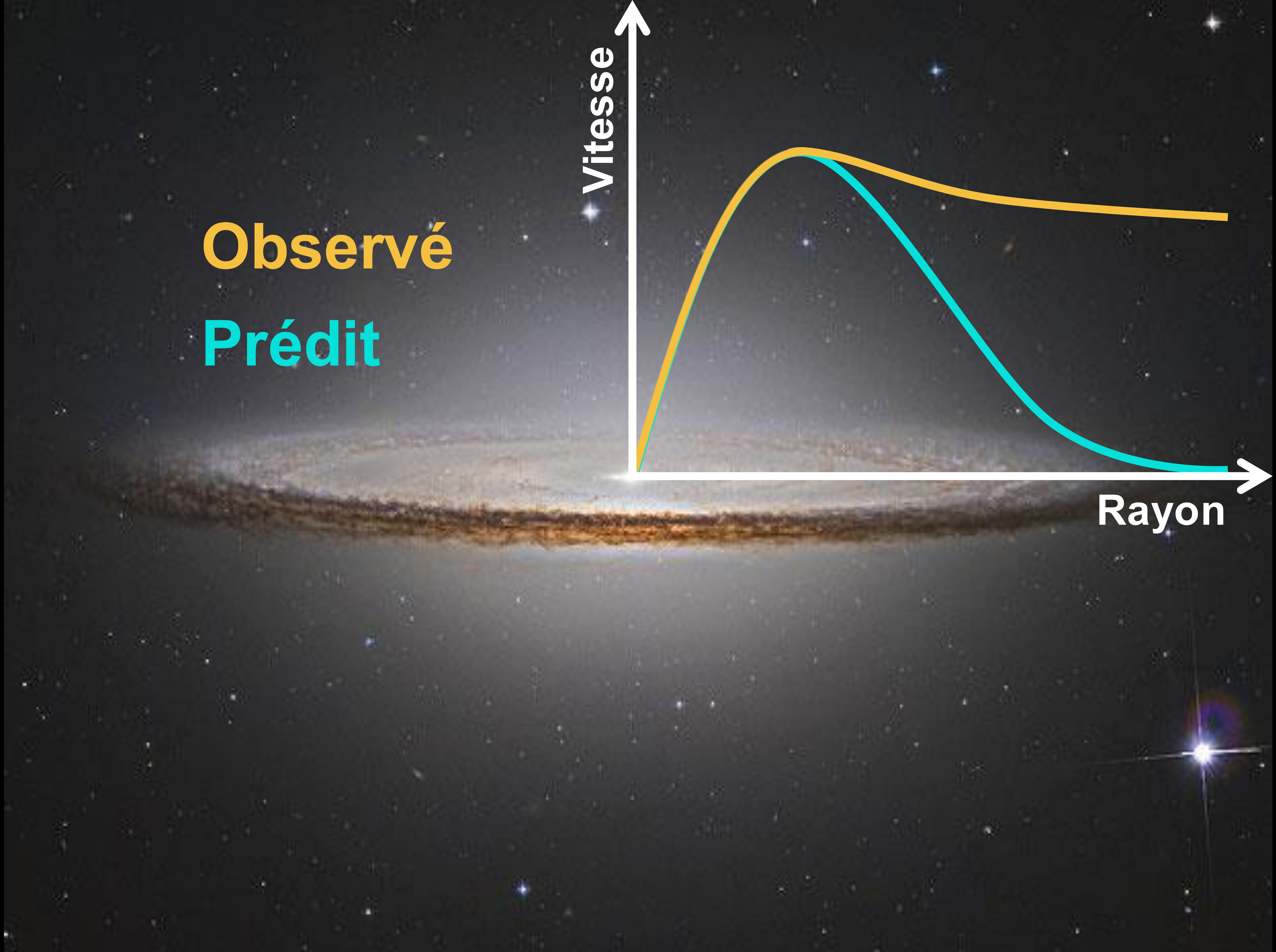
Théorie de la gravitation 

Lois du mouvement



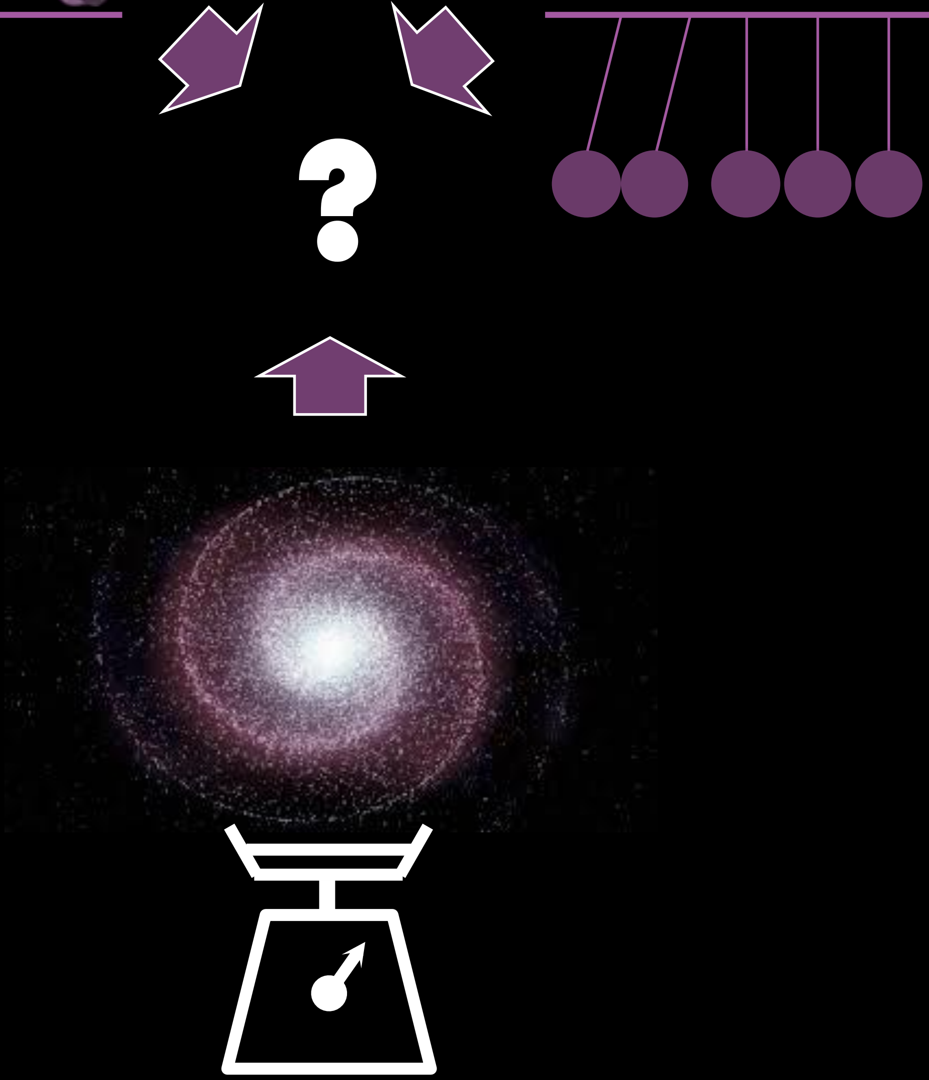



Vitesse
Prédit
Rayon

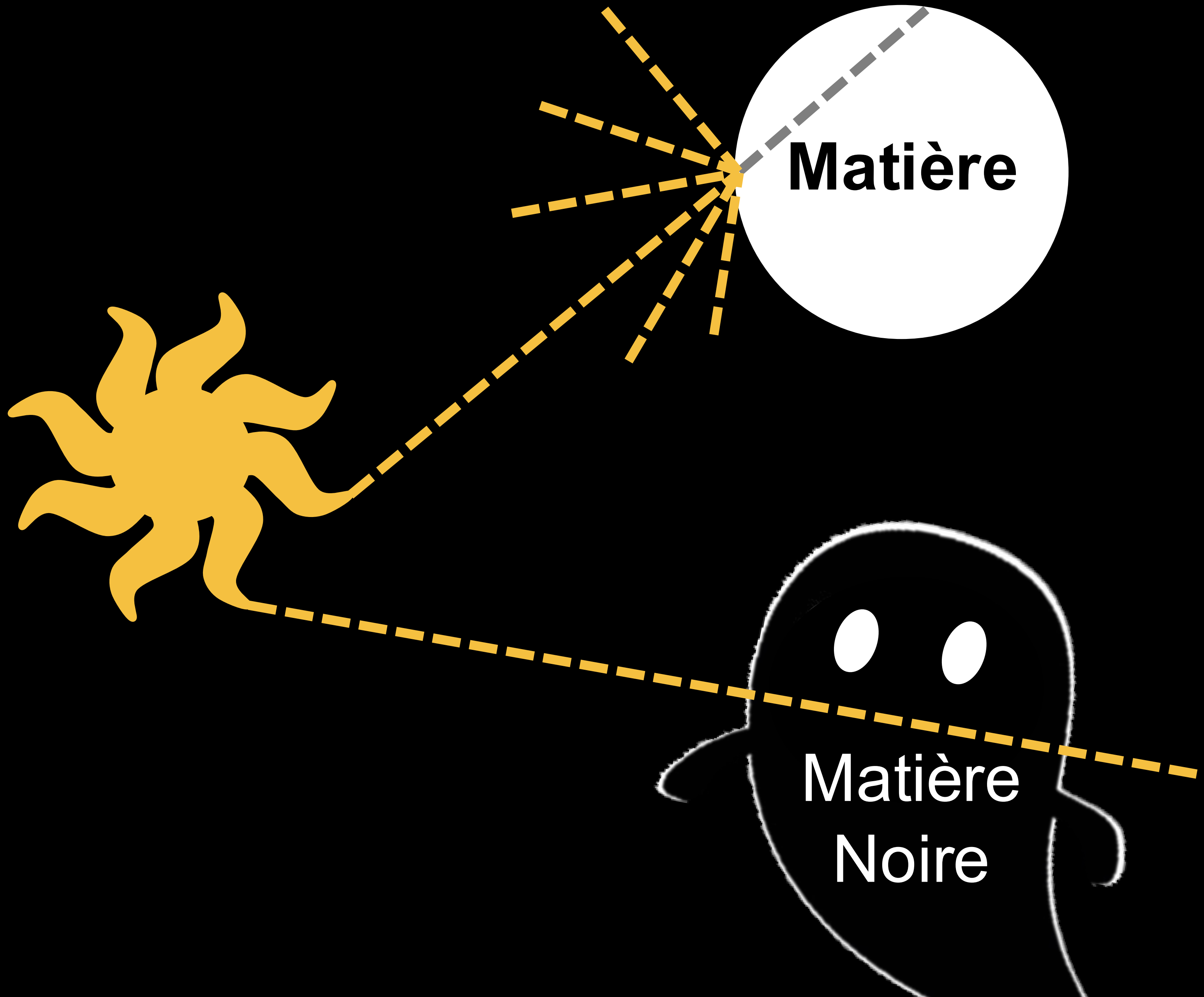

Observé
Prédit
Vitesse
Rayon

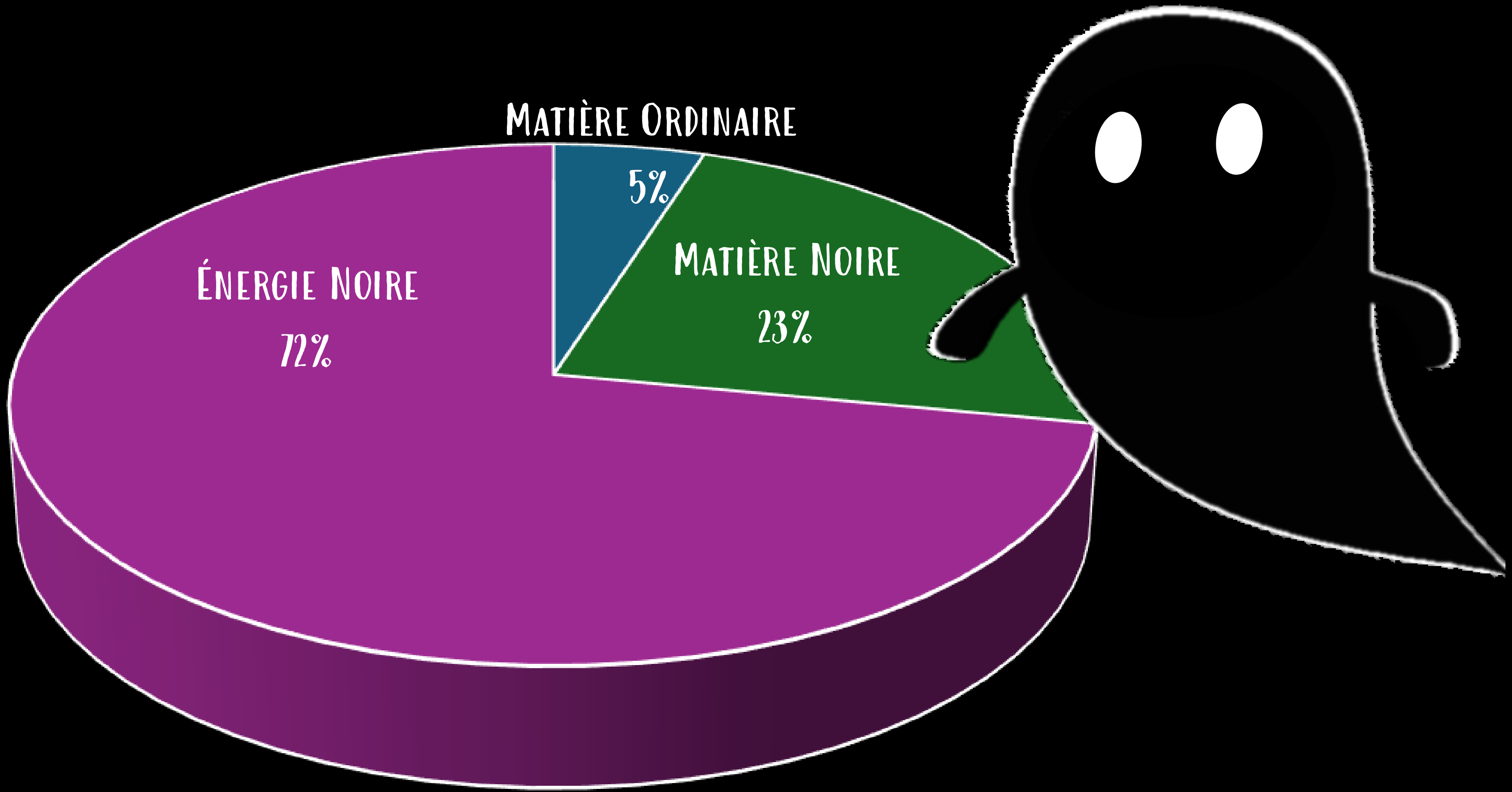
Théorie de la gravitation

Lois du mouvement

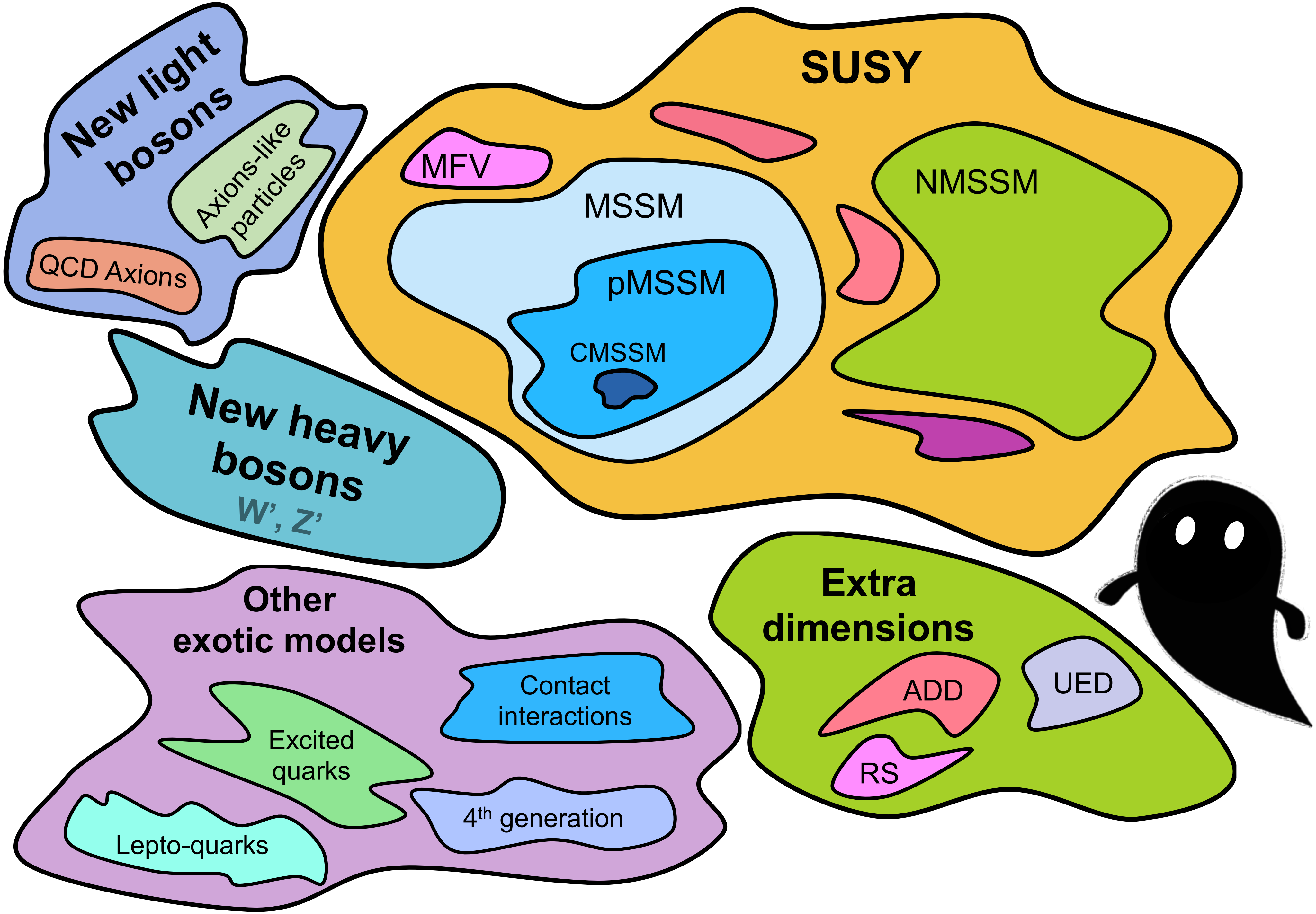



Matière
Noire

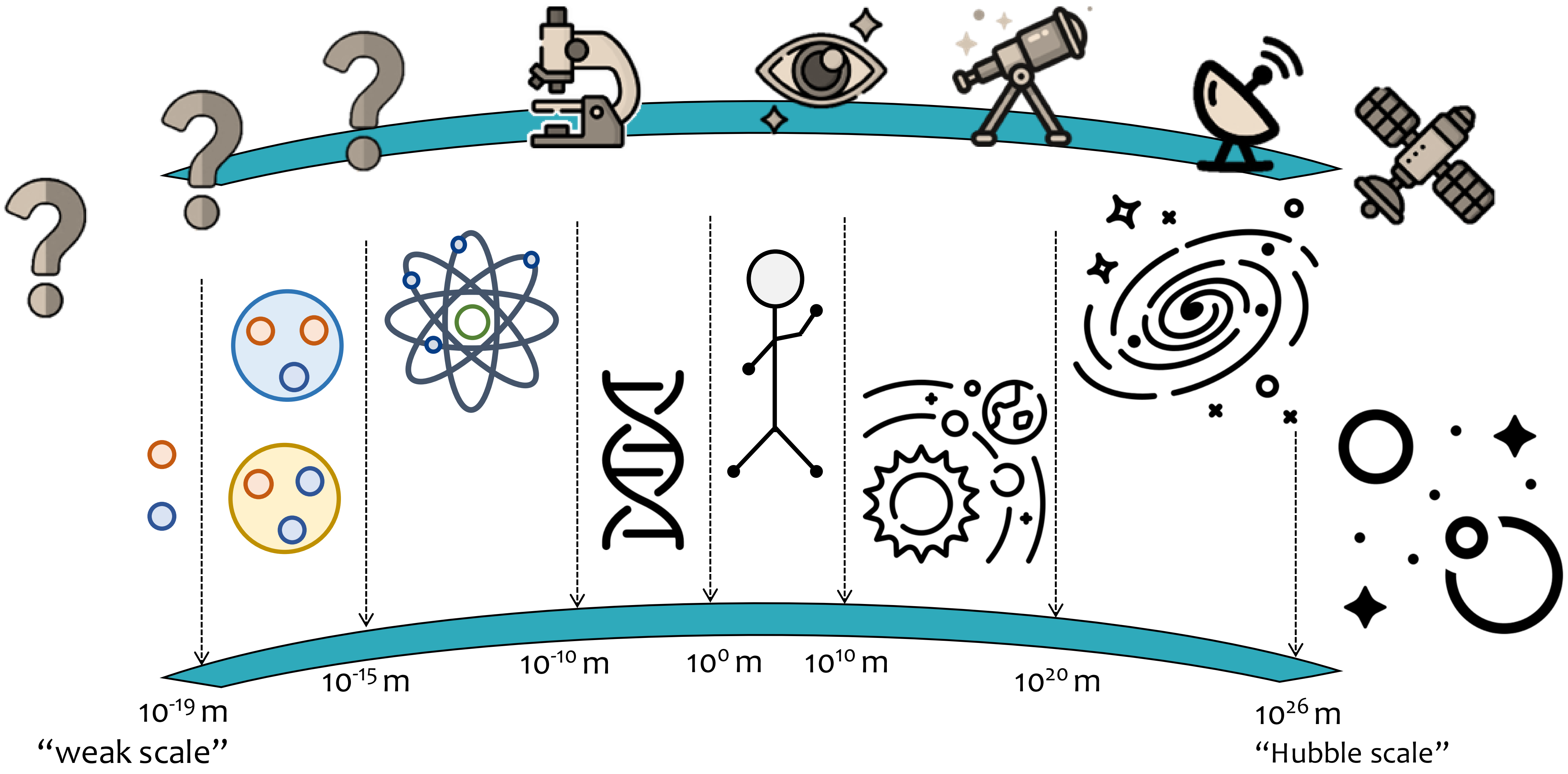

Matière
Matière
Noire

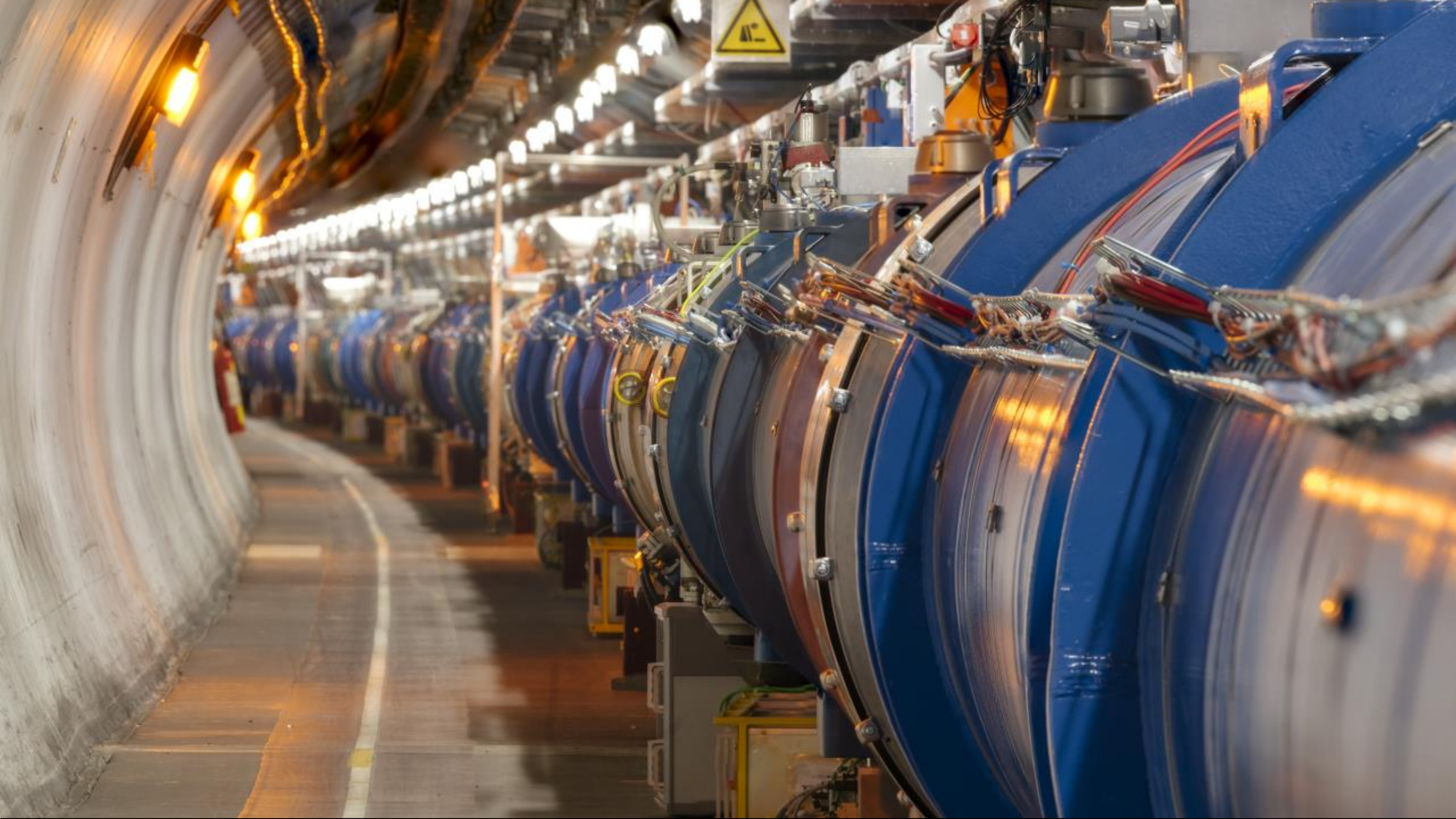

Matière Ordinaire
5%
Matière Noire
23%
Énergie Noire
72%

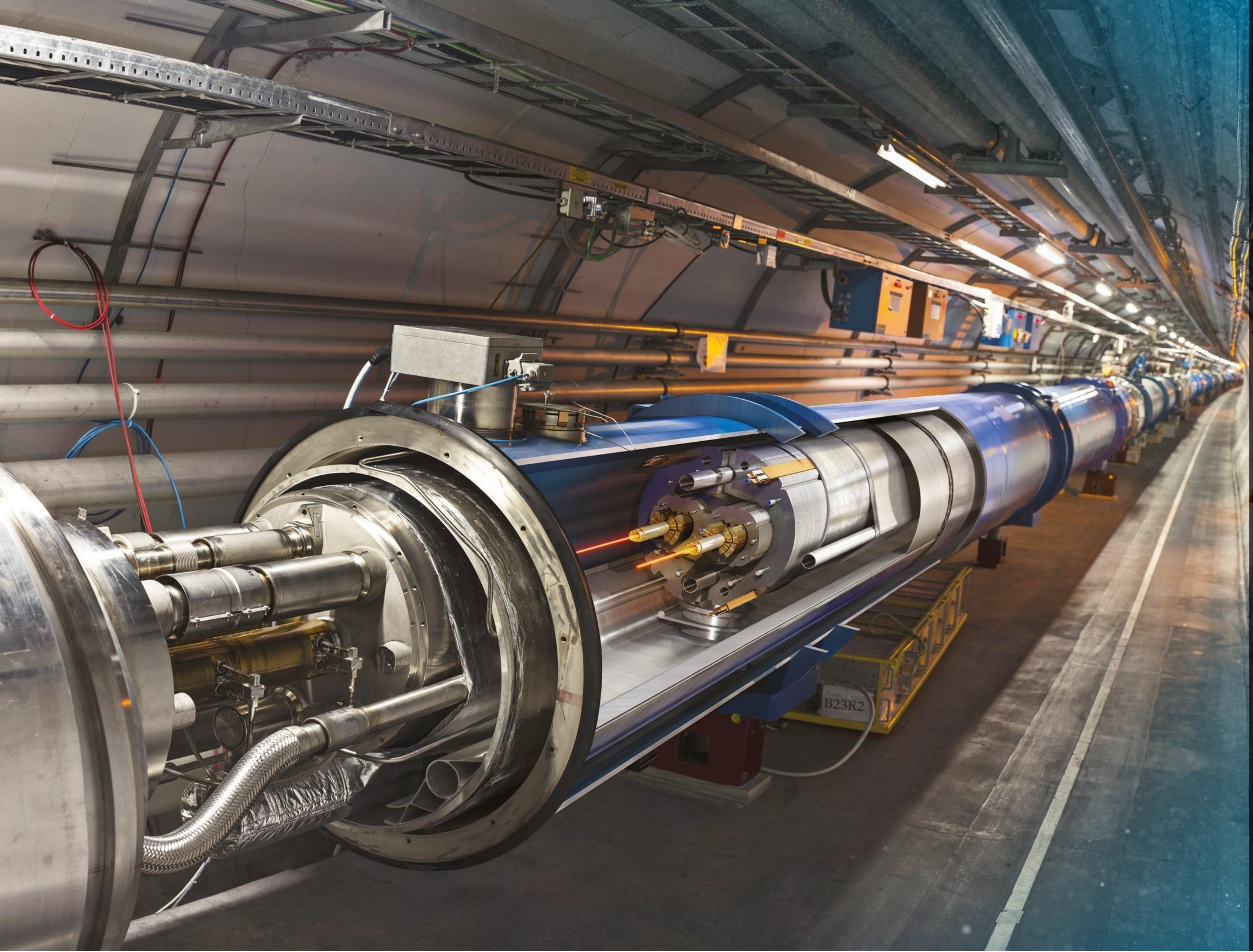
?


New light bosons
Axions-like particles
QCD Axions
SUSY
MFV
MSSM
pMSSM
CMSSM
NMSSM
New heavy bosons
W', Z'
Other exotic models
Contact interactions
Excited quarks
4th generation
Lepto-quarks
Extra dimensions
ADD
UED
RS

Échelle d'observations





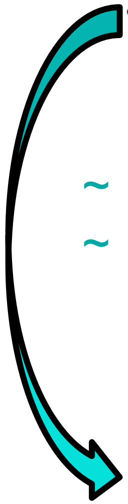


Large Hadron Collider (LHC)

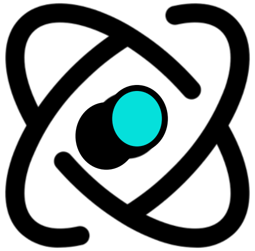
- Anneau de 30 km de circonférence
- ~ 100 m sous terre
- ~ 9300 aimants supraconducteurs à 1,9 K (–271 °C) pour guider et focaliser les faisceaux
- Protons et ions accélérés à plusieurs TeV puis mis en collision.

Quelle énergie d'un Proton dans le faisceau du LHC ?

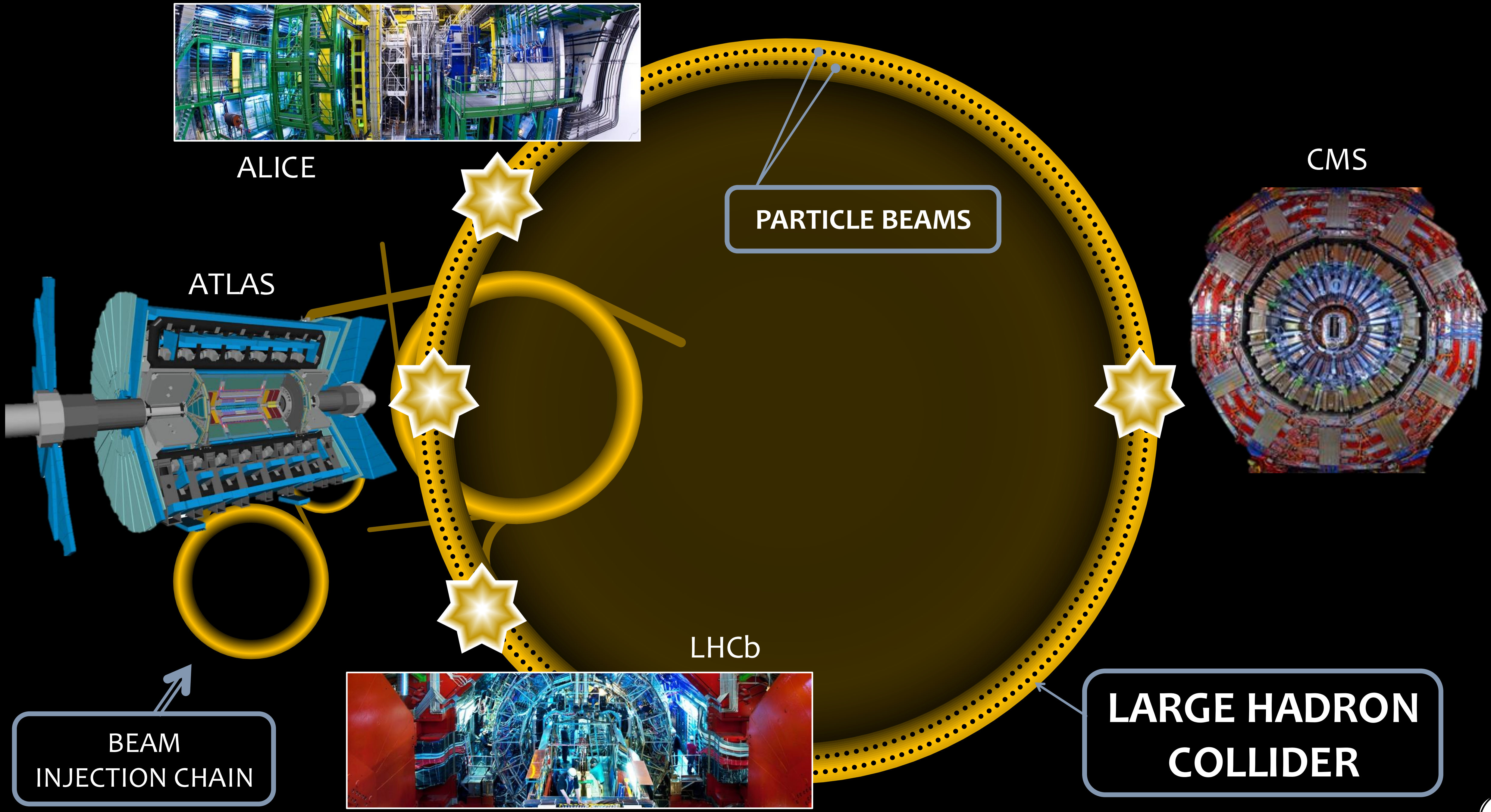
Energie d'un **moustique** qui vole
masse = 2 mg = 2 x 10^{-6} kg
vitesse = 1 m/s
Energie cinetique du moustique = ½ m v^2 = 10^{-6} J = 6.25 TeV

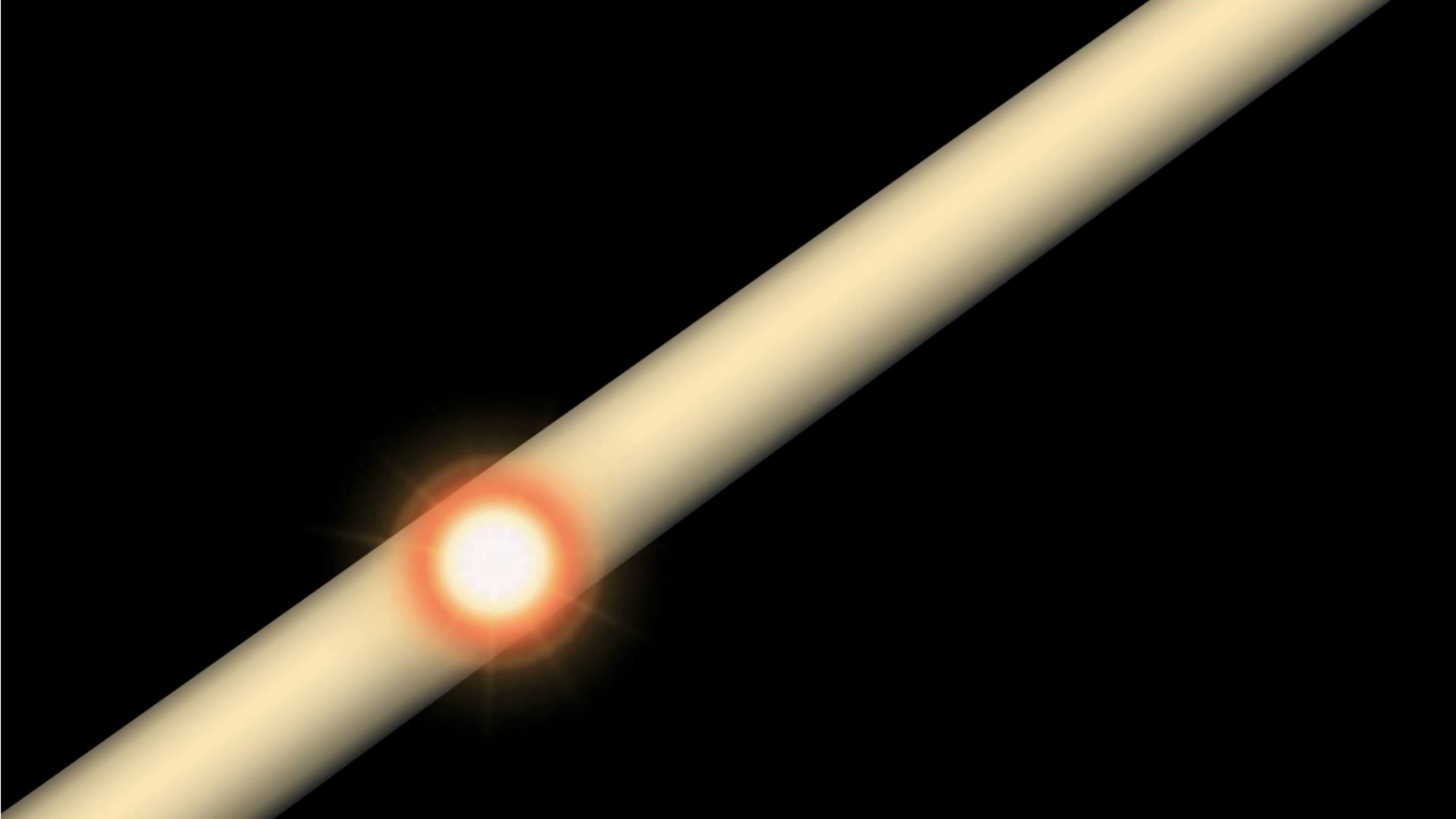


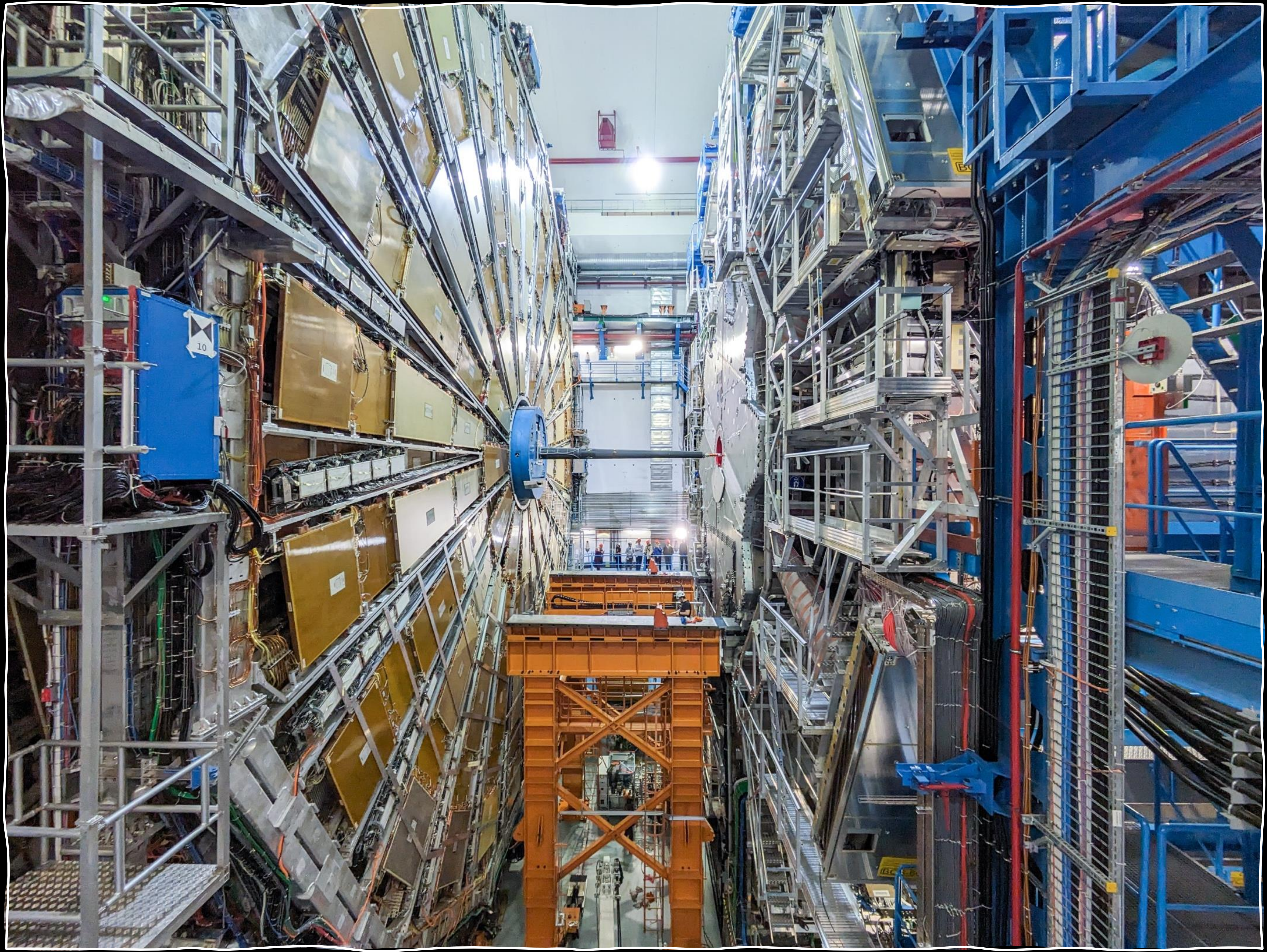
~ 10^{21} fois plus léger !
~ 10^{8} fois plus rapide !



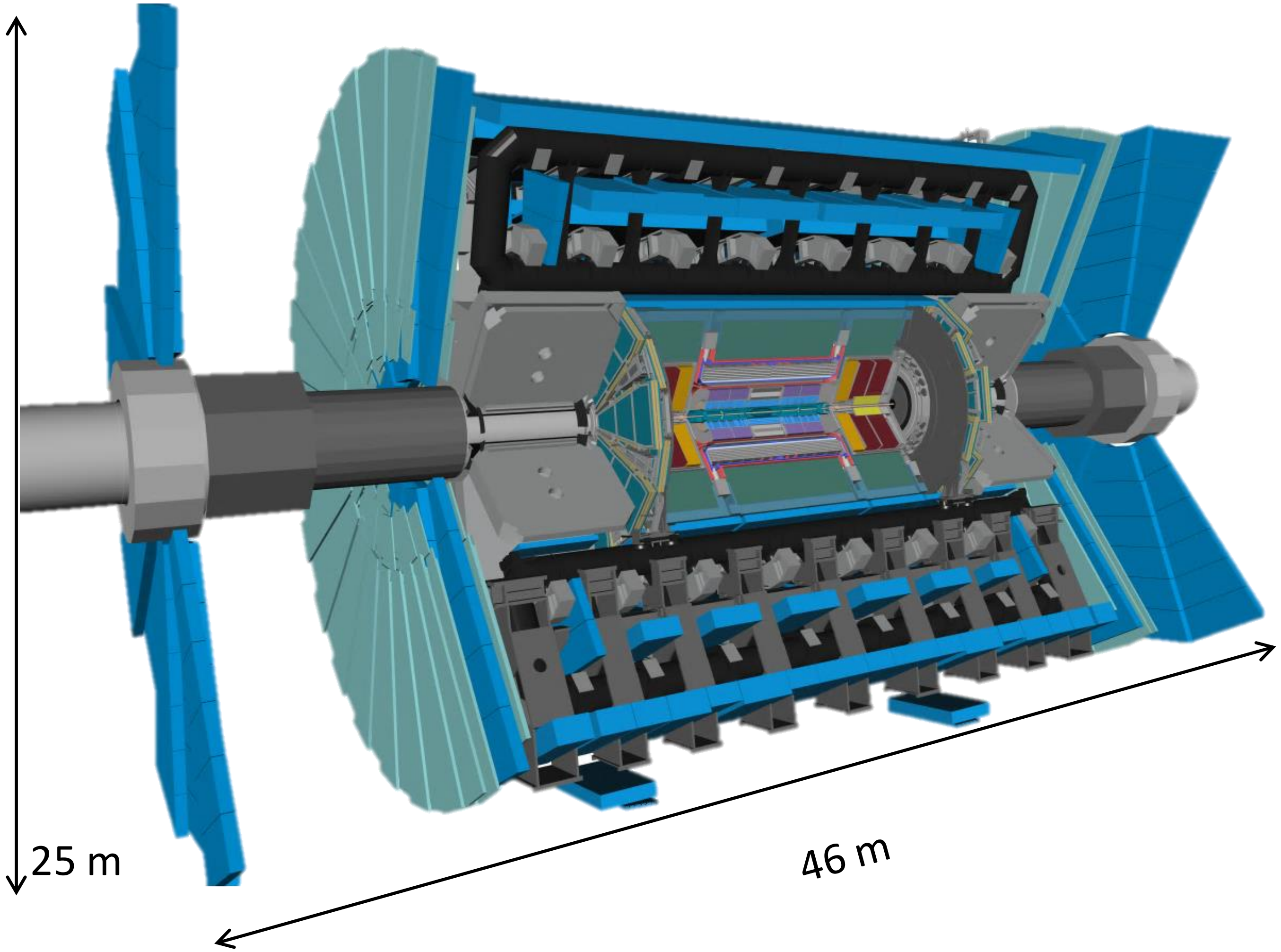
Energie d'un **proton** au LHC: 6.8 TeV
masse = ~ 10^{-27} kg
vitesse = ~ 300 000 000 m/s = ~ c (vitesse de la lumière)


ALICE
CMS
PARTICLE BEAMS
ATLAS
LHCb
BEAM
INJECTION CHAIN
LARGE HADRON
COLLIDER




10

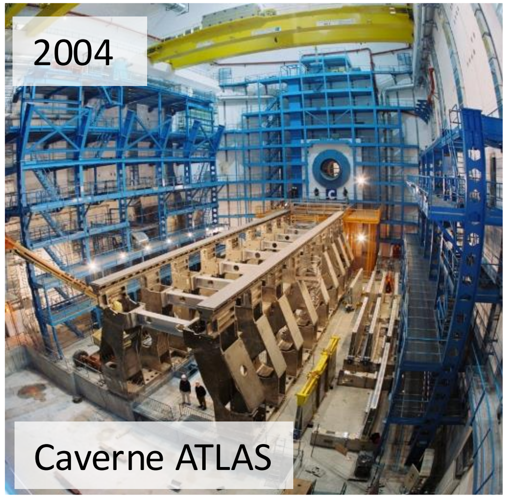
Le détecteur ATLAS en chiffres

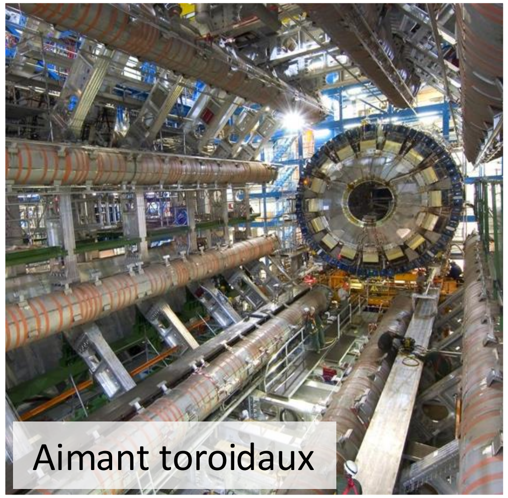


- Poids: 7 kilotonnes 
- Aimants supraconducteurs de 2 à 4 T
- Position des particules enregistrée avec une precision de l'ordre de 10 μm
- 100 millions de canaux

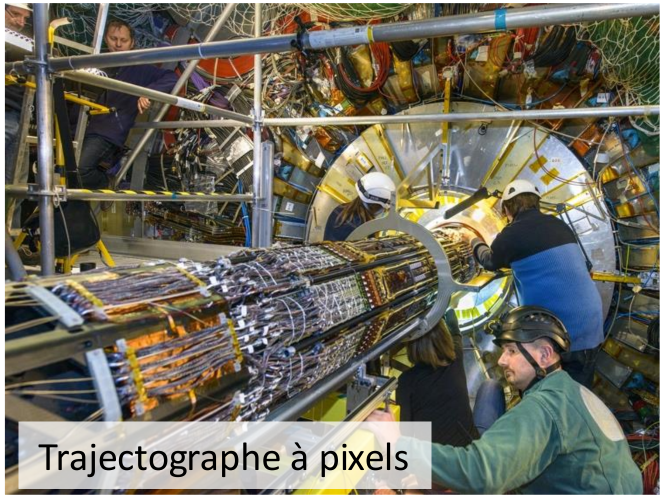
- 1 milliard de collisions par seconde
- 1 000 événements par seconde enregistrés
- 500 PB de données sur disque et sur bande
- 0.5 millions de coeurs CPU utilisées 7/24

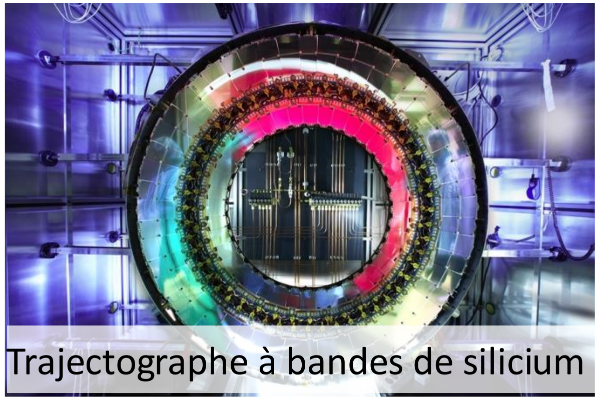
- 100 milliards d'événements collectés

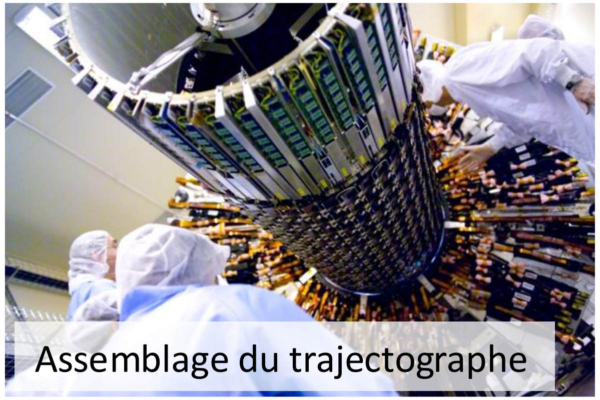

2004
Caverne ATLAS

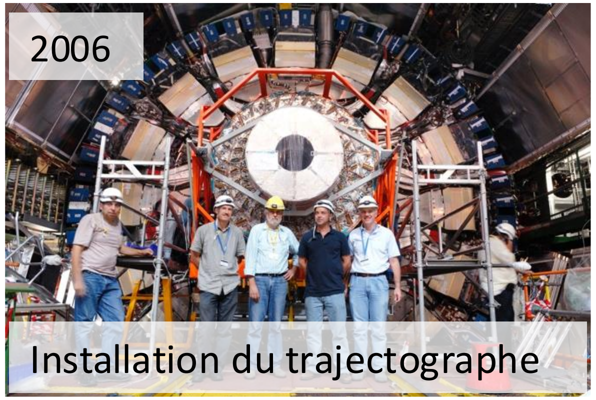

Aimant toroidaux


Calorimètre


Trajectographe à pixels

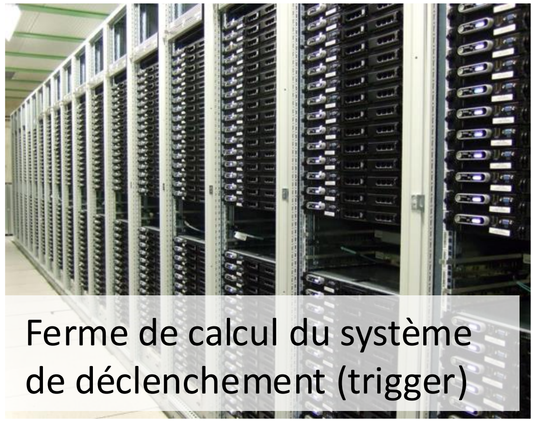

Trajectographe à bandes de silicium


Assemblage du trajectographe


2006
Installation du trajectographe

ATLAS
EXPERIMENT


2014
Installation du trajectographe à pixels le plus interne


Ferme de calcul du système de déclenchement (trigger)

(Aspects communs à tous les détecteurs du LHC)

- Capteurs rapides et résistants aux radiations
- Stabilité et précision des structures construites
- Systèmes de lecture extrêmement rapides pour un traitement à faible latence
- Infrastrucuture informatique capable de traiter des volumes de données gigantesques

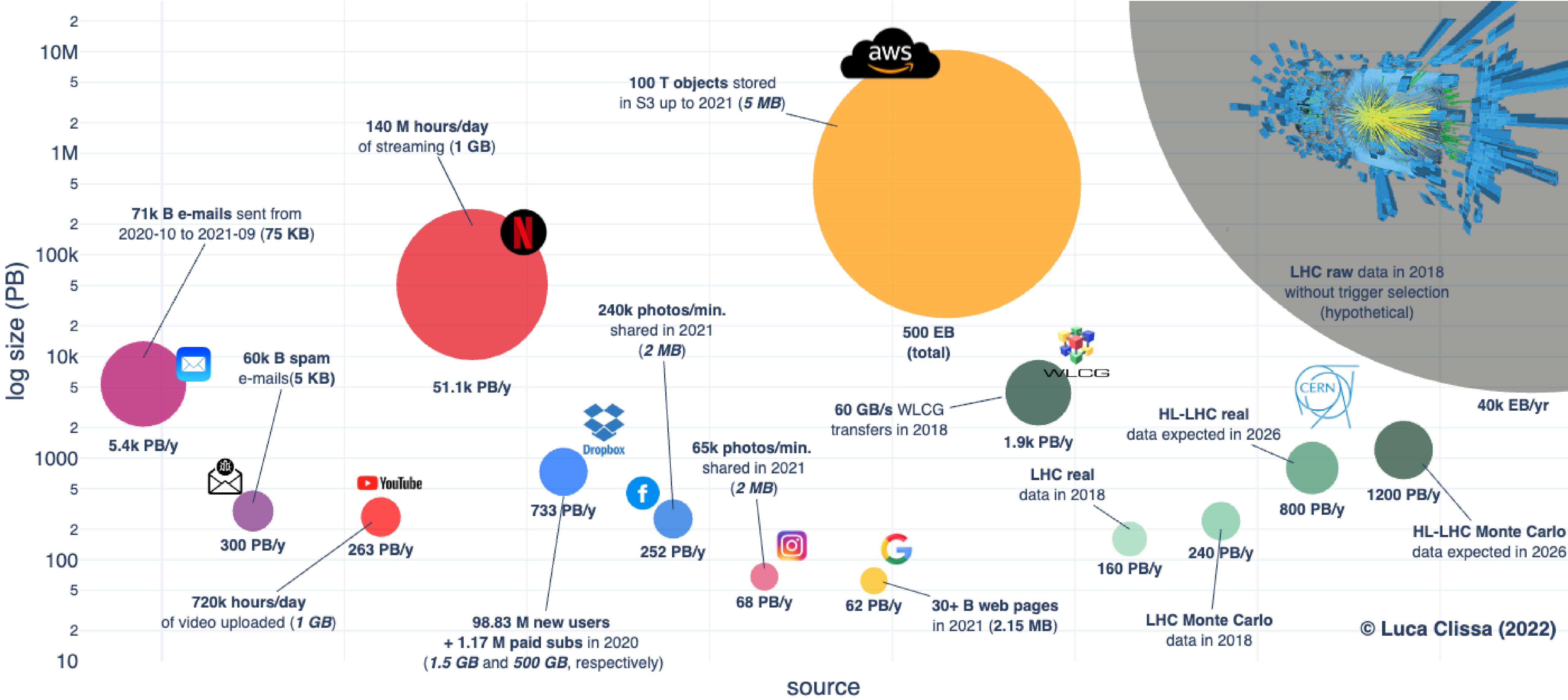
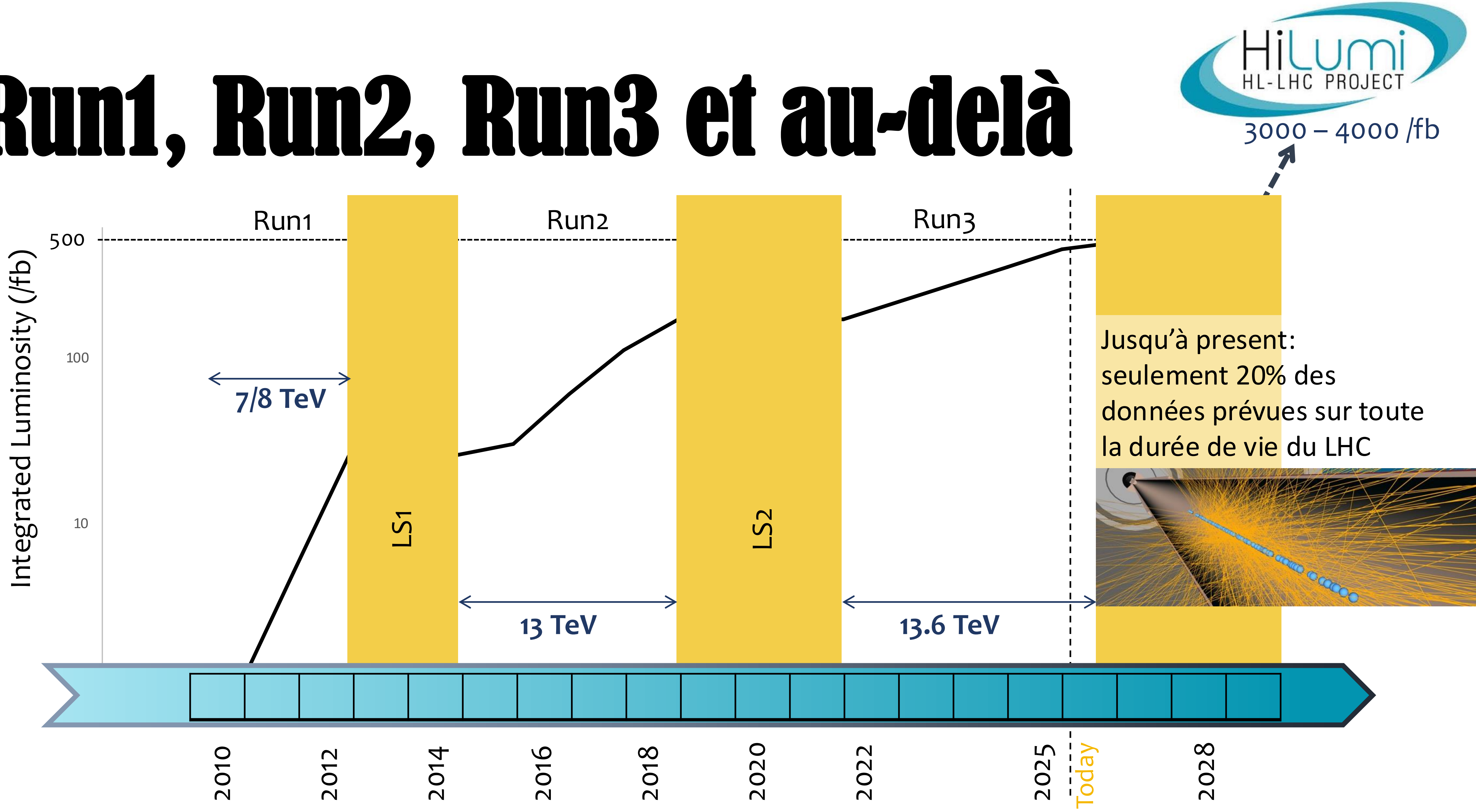


Figure 2.3: **Big Data sizes.** Bubble plot of the orders of magnitude of data produced by important big data players. The balloon areas illustrate the amount of data and the text annotations highlight the key factors considered in the estimates. Average per-unit sizes are reported in parentheses, where italic indicates measures reconstructed based on likely assumptions because no references were found.

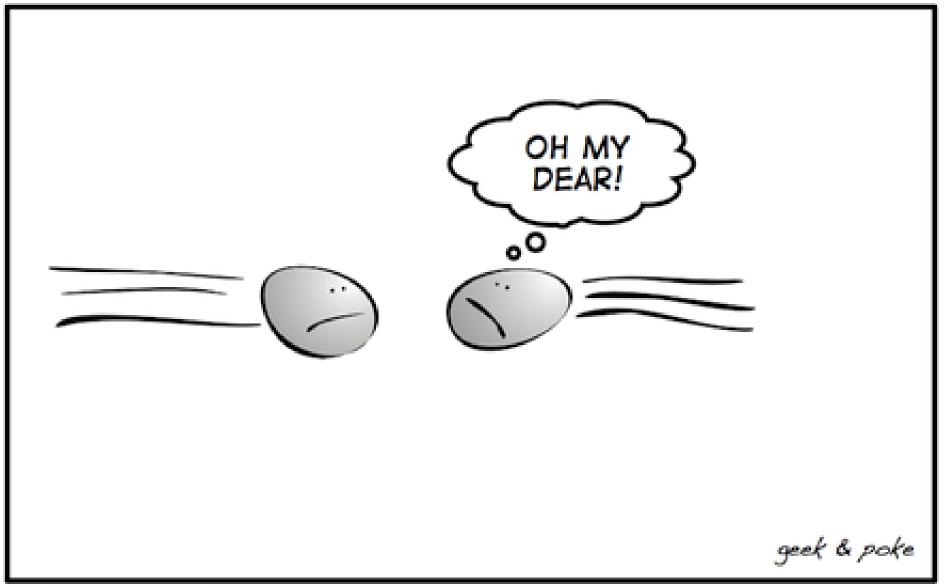
https://clissa.github.io/BigData2021/BigData2021.html

Run1, Run2, Run3 et au-delà



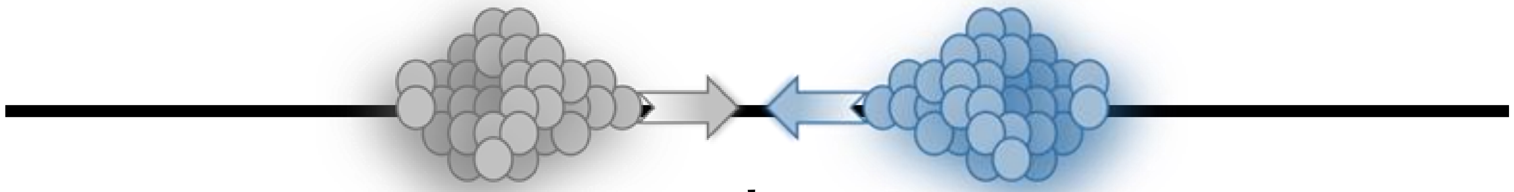
Jusqu'à present:
seulement 20% des
données prévues sur toute
la durée de vie du LHC

La collision proton-proton



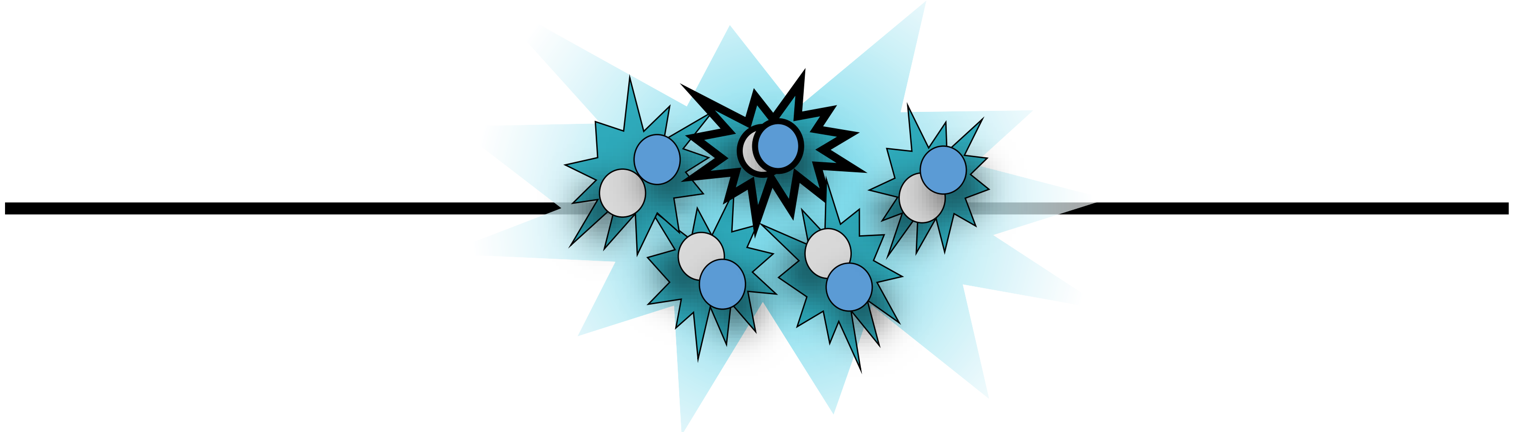
LATELY INSIDE THE LHC:
2 PROTONS 0.000000000000000001 SEC BEFORE THE COLLISION

La collision proton-proton



Paquets de protons
~ 1.8 x 10^{11} protons par paquet
Espacement entre paquets: 25 ns

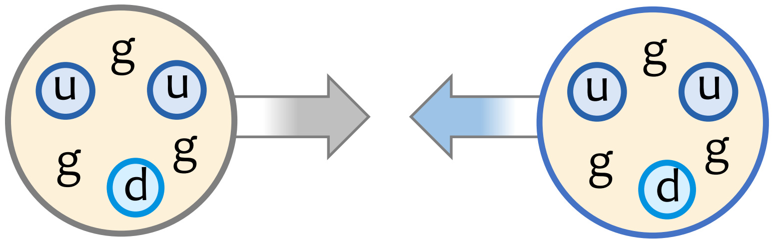
La collision proton-proton



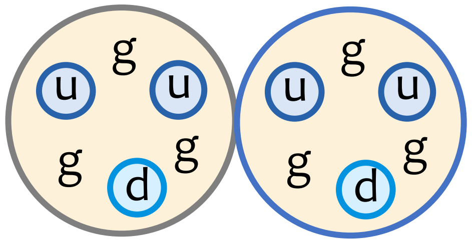
~ 1.8 x 10^{11} protons par paquet
Espacement entre paquets: 25 ns

Jusqu'à 60–70 collisions proton-proton par croisement de paquets

La collision proton-proton

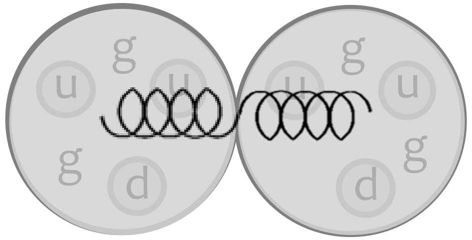


La collision proton-proton

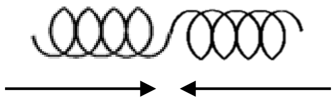


La collision proton-proton

Les “partons” (constituents of the protons) de chaque proton peuvent interagir
Par exemple ici: deux gluons

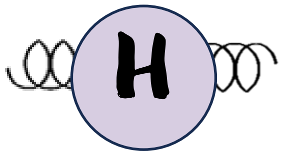


La collision proton-proton



La collision proton-proton

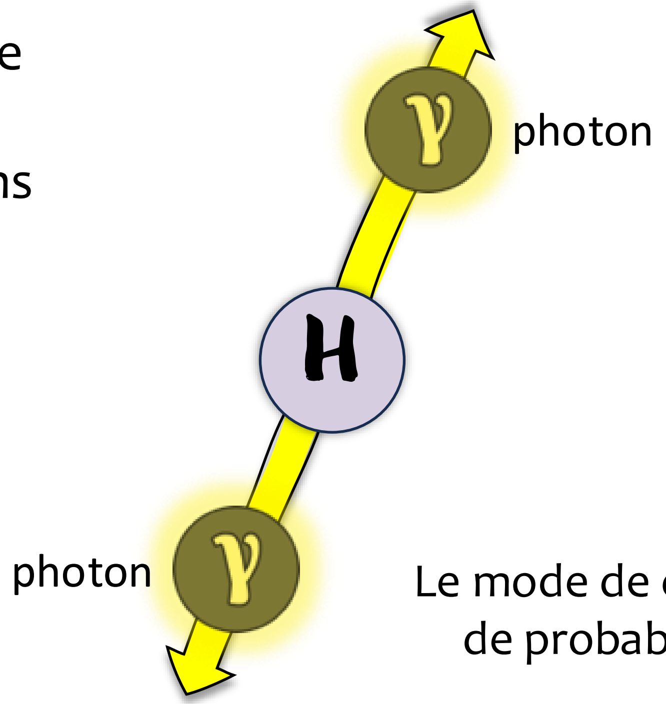
L'interaction parton-parton conduit à la production de particles
Par exemple ici: le boson de Higgs



La fréquence à laquelle certaines particules sont produites
dépend de probabilités calculables dans le cadre théorique

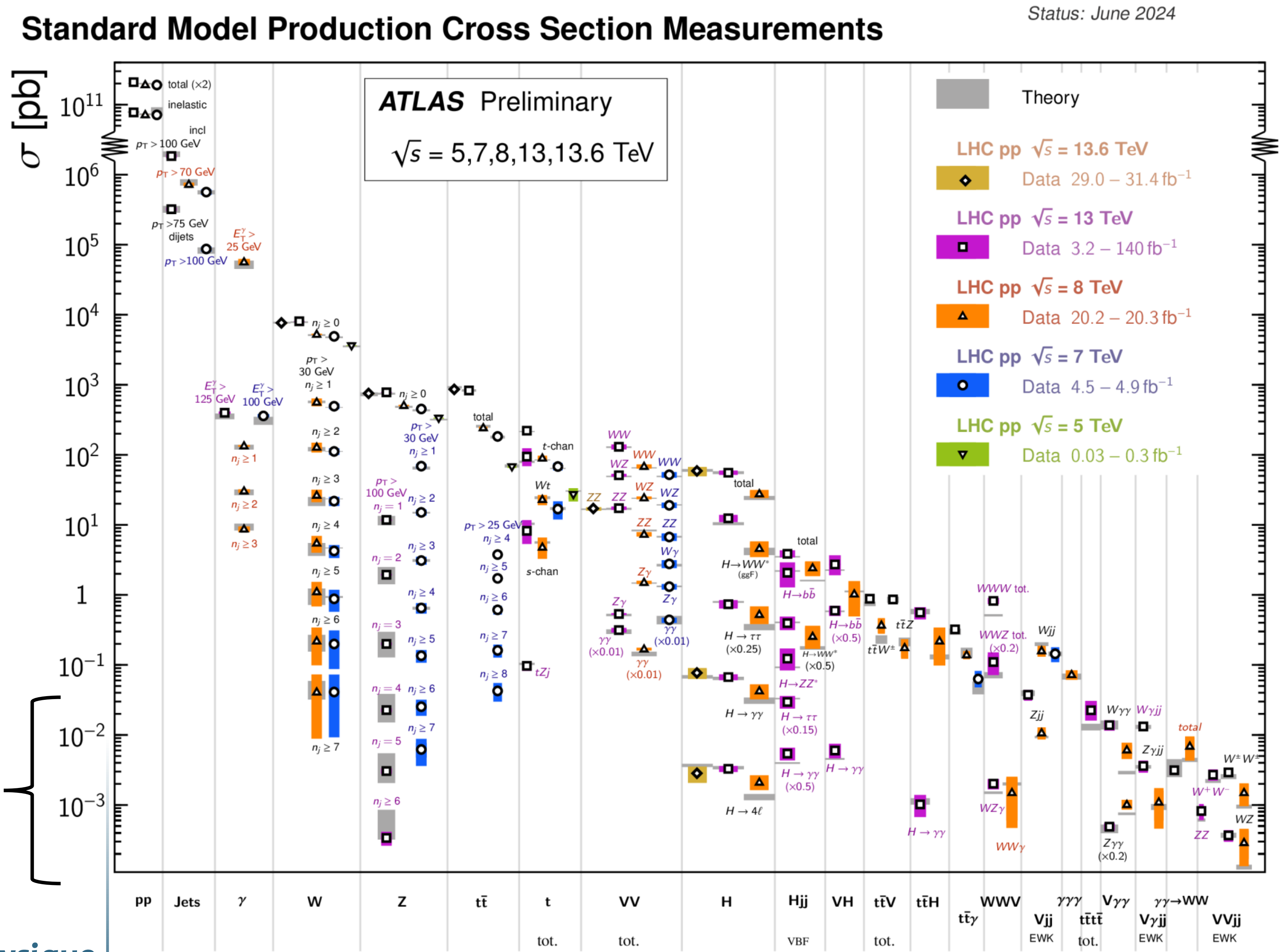
La collision proton-proton

Le boson de Higgs se désintègre
presque instantanément
Par exemple ici: en deux photons



Le mode de désintégration de chaque particule dépend
de probabilités calculables dans le cadre théorique

État du Modèle Standard



Certains aspects fondamentaux restent à explorer

Nouvelle physique

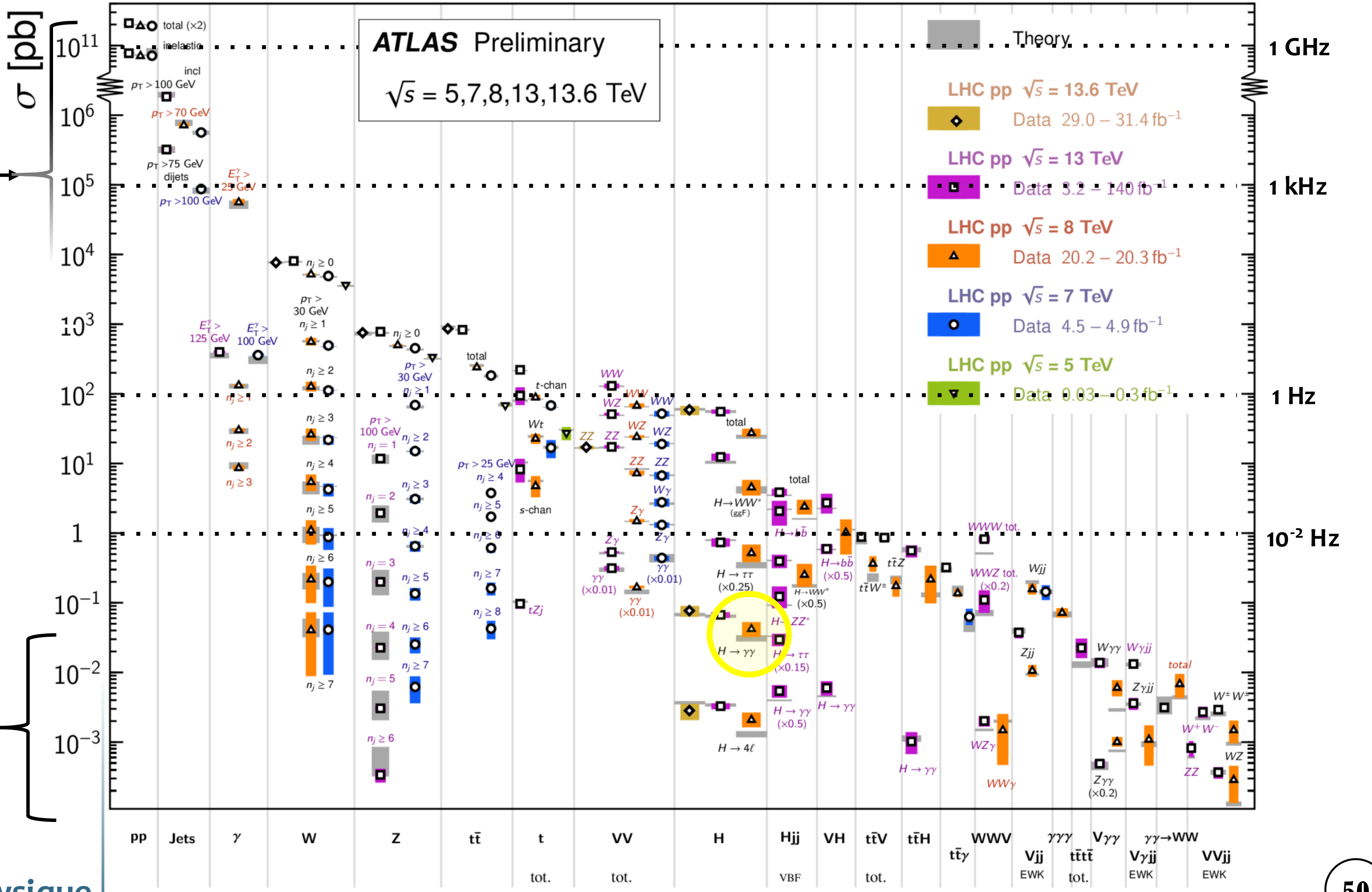
Reminder: $\sigma = \frac{\#\text{ events}}{L}$

Event Rate

$L_{inst} = 10^{34}\ cm^{-2}s^{-1}$

État du Modèle Standard

Standard Model Production Cross Section Measurements



Le taux d'événements est dominé par les "processus de fond"

Certains aspects fondamentaux restent à explorer

Nouvelle physique

Triggering Challenge

Défi du déclenchement

Les événements non sélectionnés sont perdus à jamais

Débits de données extrêmes

- Collisions proton-proton toutes les 25 ns (40 MHz)
- Environ 100 millions de canaux de détection
- O(MB) par collision (après suppression des zéros)

→ Débit de données brut > 100 TB/s

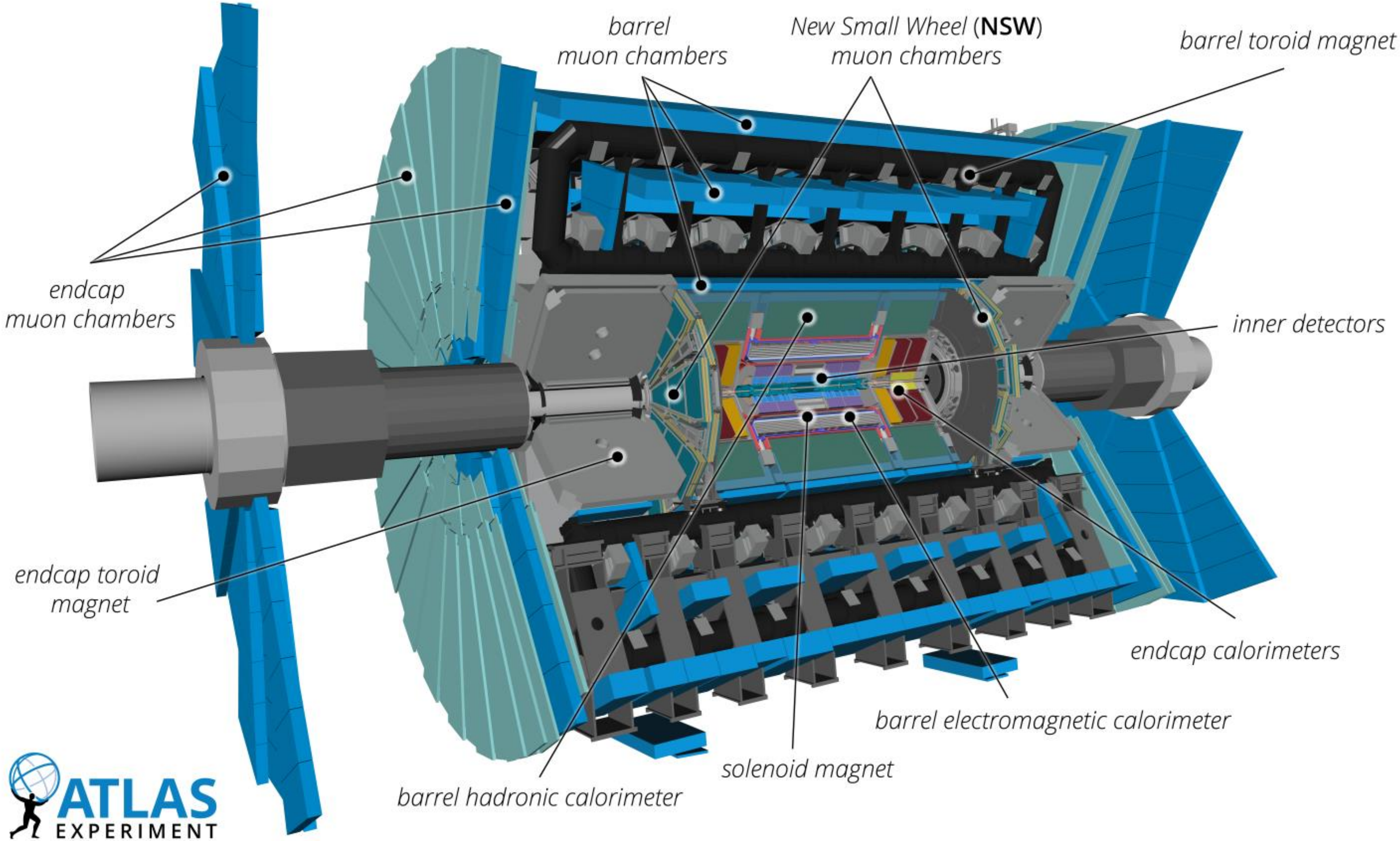
→ **Impossible** de tout lire ou de tout stocker!

Contraintes du système

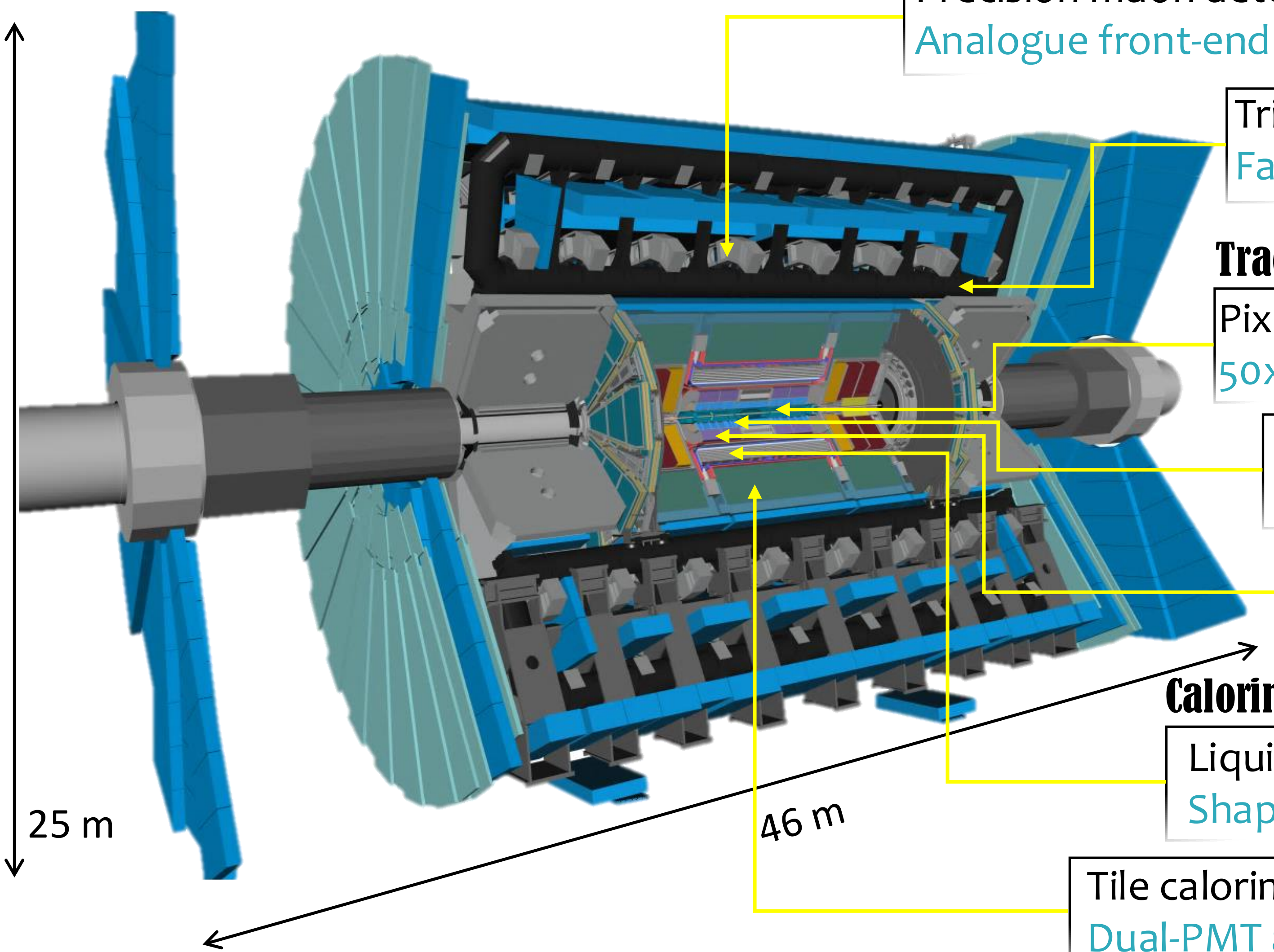
- Bande passante de lecture du détecteur (electronique)
- Latence disponible pour traitement
- Ressources de calcul disponibles

Objectifs du système de déclenchement

- Identifier et sélectionner les événements dignes d'être conservés à un taux soutenable: O(kHz) sur 40 MHz
- La décision d'acceptation ou de rejet doit être rapide et fiable
- Maintenir une large acceptation en physique
  - y compris pour des phénomènes nouveaux inconnus
- Être robuste face au pile-up et au bruit du détecteur


barrel
muon chambers
New Small Wheel (NSW)
muon chambers
barrel toroid magnet
endcap
muon chambers
inner detectors
endcap toroid
magnet
endcap calorimeters
barrel electromagnetic calorimeter
solenoid magnet
barrel hadronic calorimeter
ATLAS
EXPERIMENT

Number of channels



Muon detectors

Precision muon detectors: **400 k channels**
Analogue front-end with time / charge digitization

Trigger muon detectors: **700 k channels**
Fast analogue front-end with binary trigger read-out

Tracking devices

Pixel detector: **80 M channels**
50x400 µm pixel ASICs

Silicon strip detector: **6 M channels**
Binary readout architecture

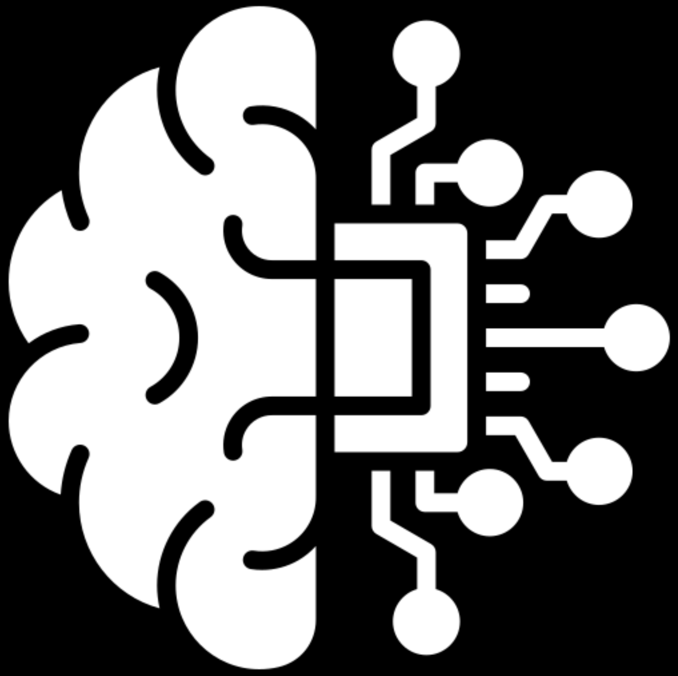
Transition radiation tracker: **350 k channels**
Analogue front-end with threshold discrimination

Calorimeters

Liquid-argon calorimeter: **180 k channels**
Shaped analog readout with multi-gain sampling

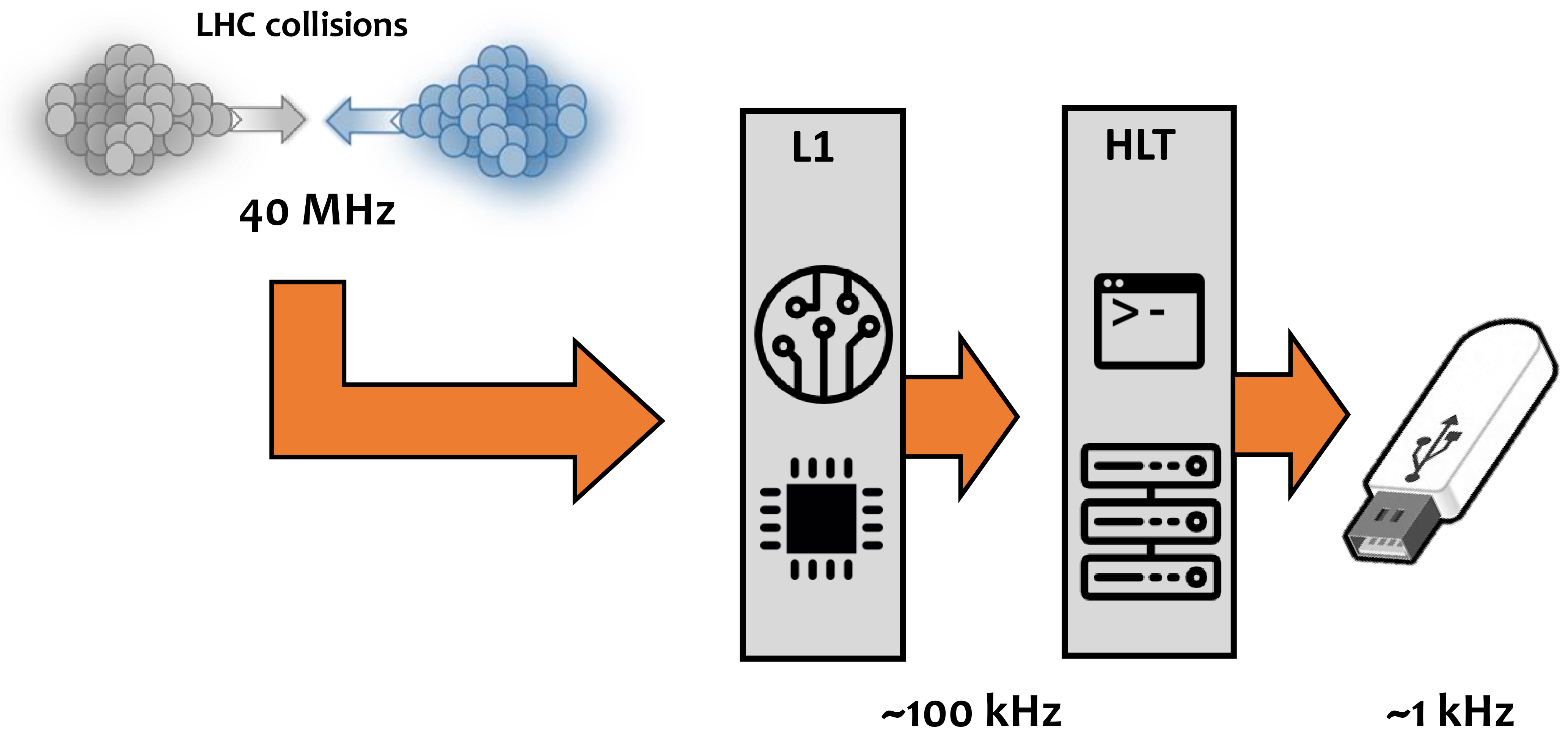
Tile calorimeter: **10 k channels**
Dual-PMT analogue readout

Trigger system:

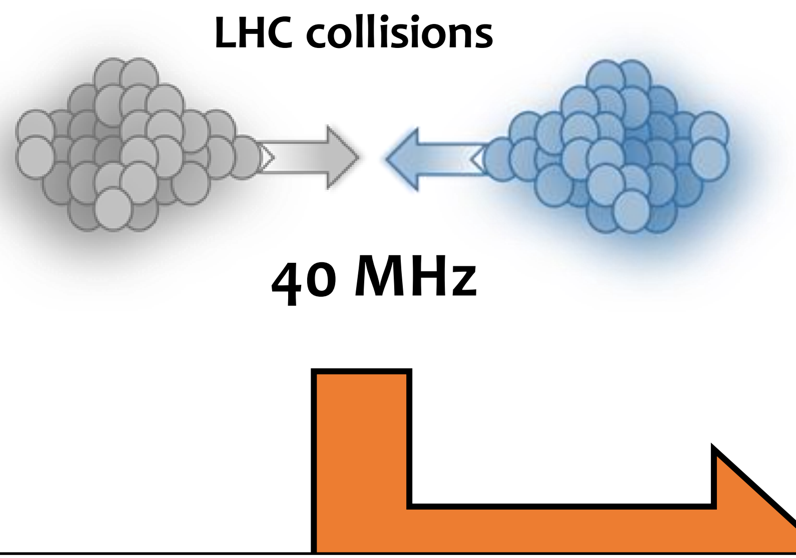


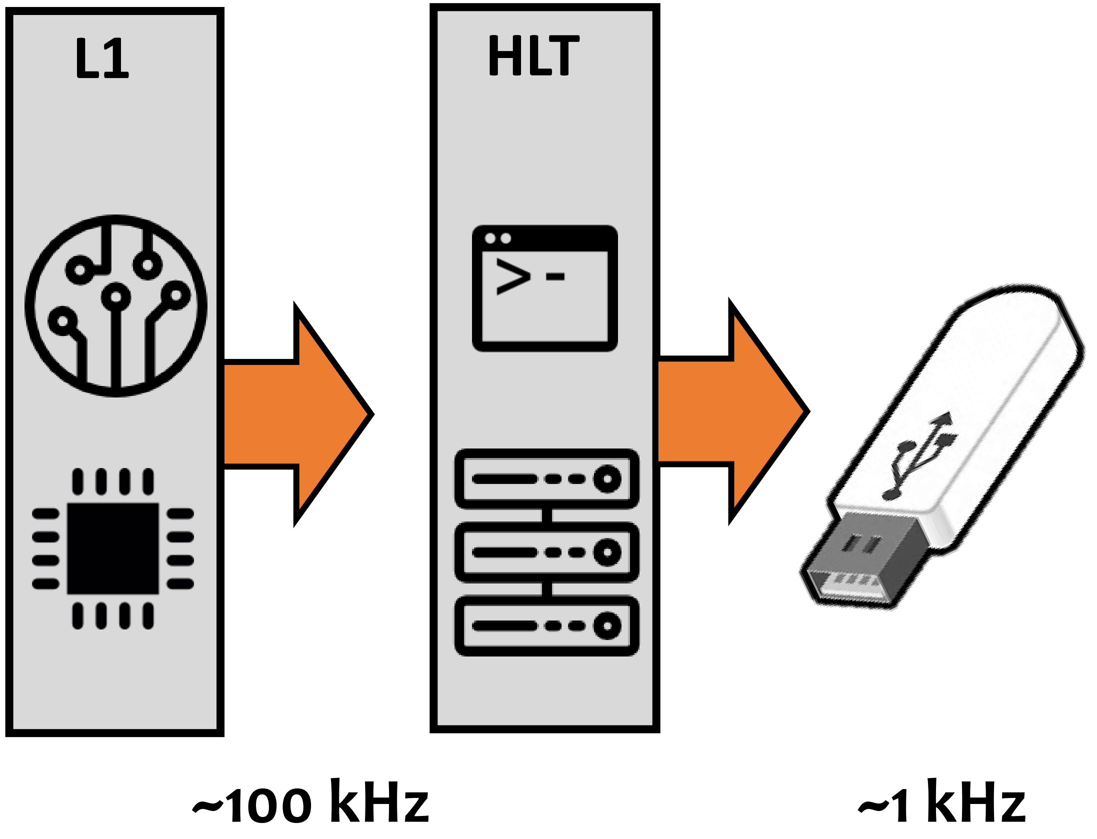
the experiment's brain

Triggering in Physics



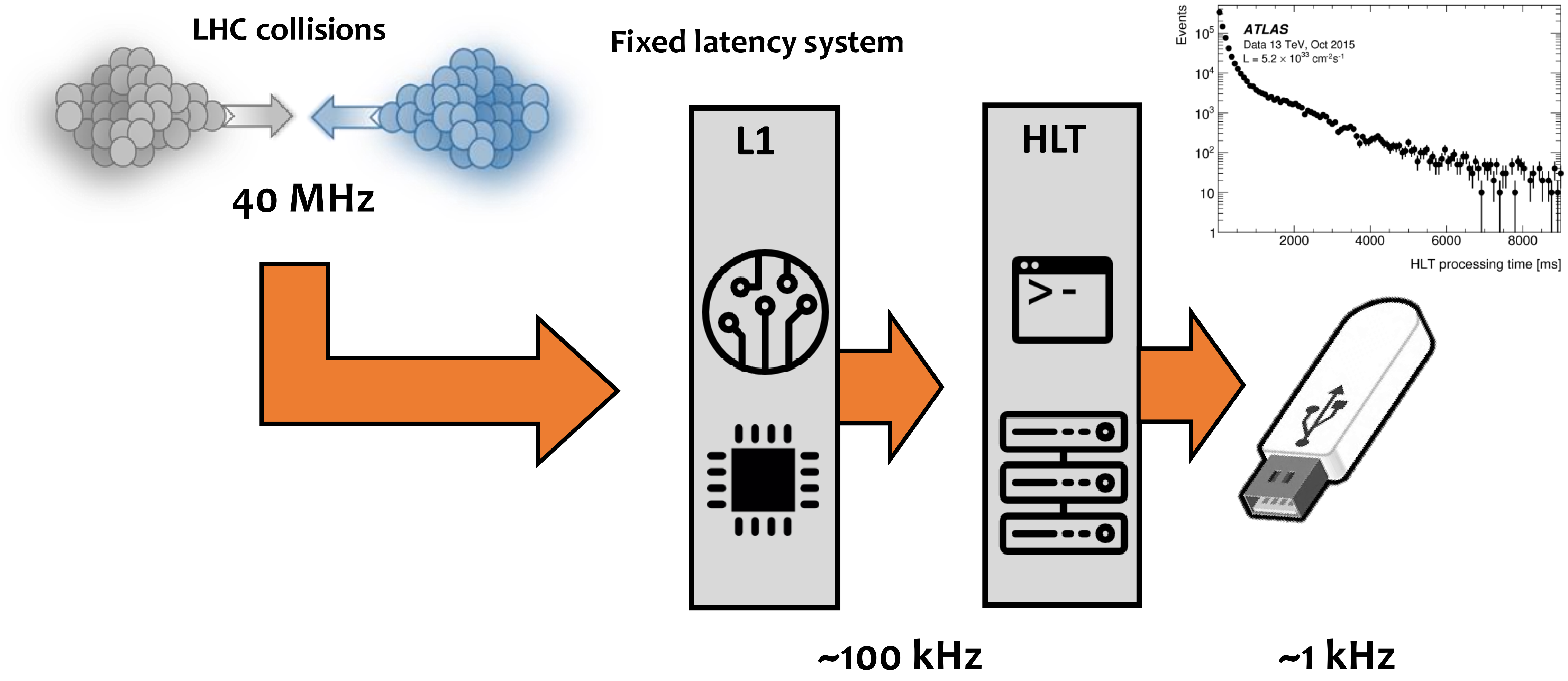
Triggering in Physics



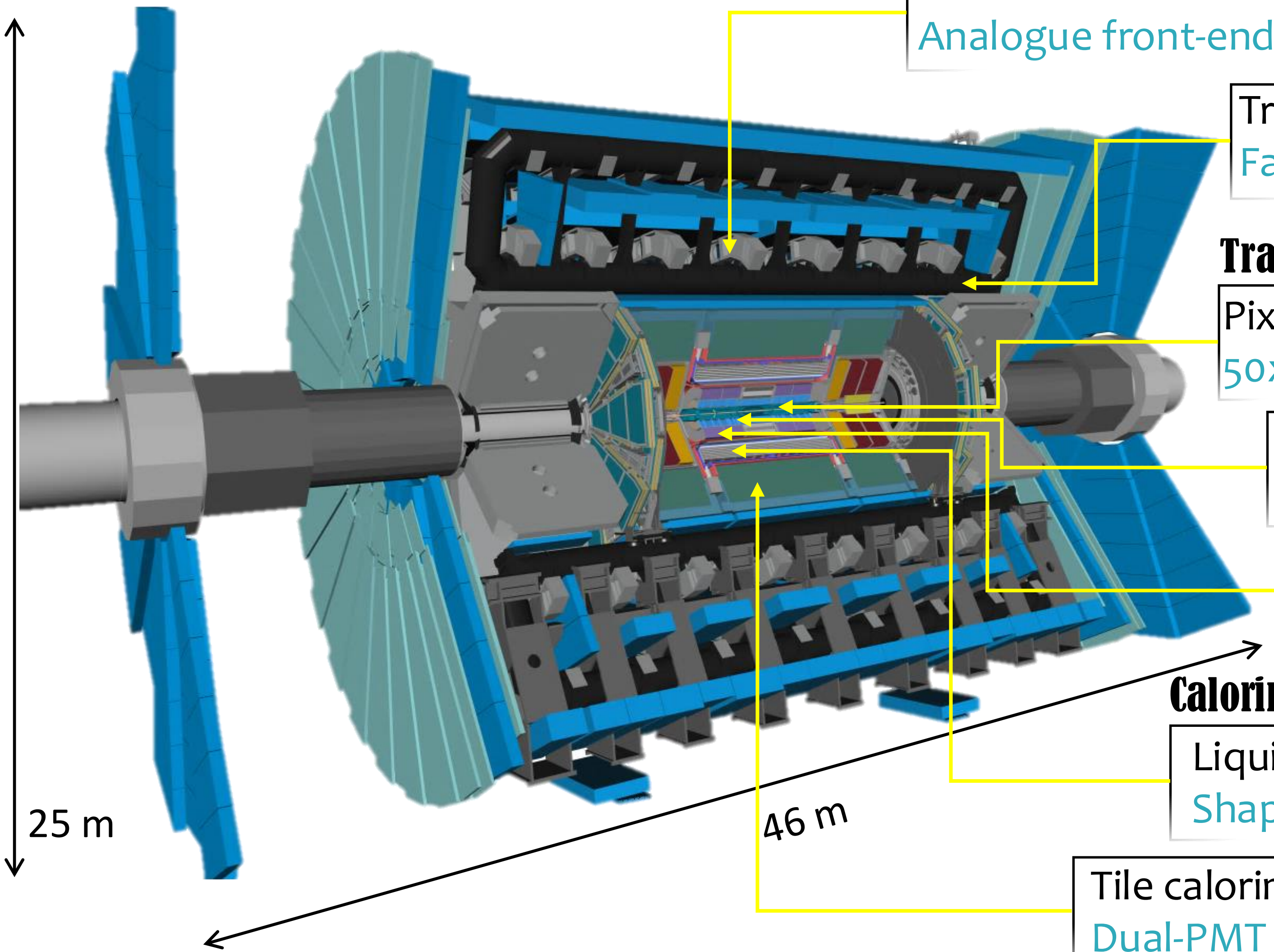


- Triggering details differ between LHC experiments
- General triggering concept in the large LHC experiments is similar, driven by huge data volumes
  - *LHCb has gone "triggerless"*

Triggering in Physics



Number of L1 trigger channels



Muon detectors

Precision muon detectors: **400 k channels**
Analogue front-end with time / charge digitization ✗

New primitives in the future

Trigger muon detectors: **700 k channels**
Fast analogue front-end with binary trigger read-out

10-20k logical channels

Tracking devices

Pixel detector: **80 M channels**
50x400 µm pixel ASICs ✗

Silicon strip detector: **6 M channels**
Binary readout architecture ✗

Transition radiation tracker: **350 k channels**
Analogue front-end with threshold discrimination ✗

Calorimeters

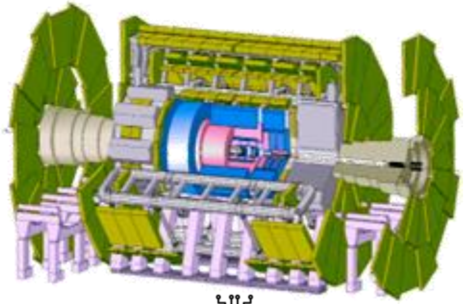
Liquid-argon calorimeter: **180 k channels**
Shaped analog readout with multi-gain sampling

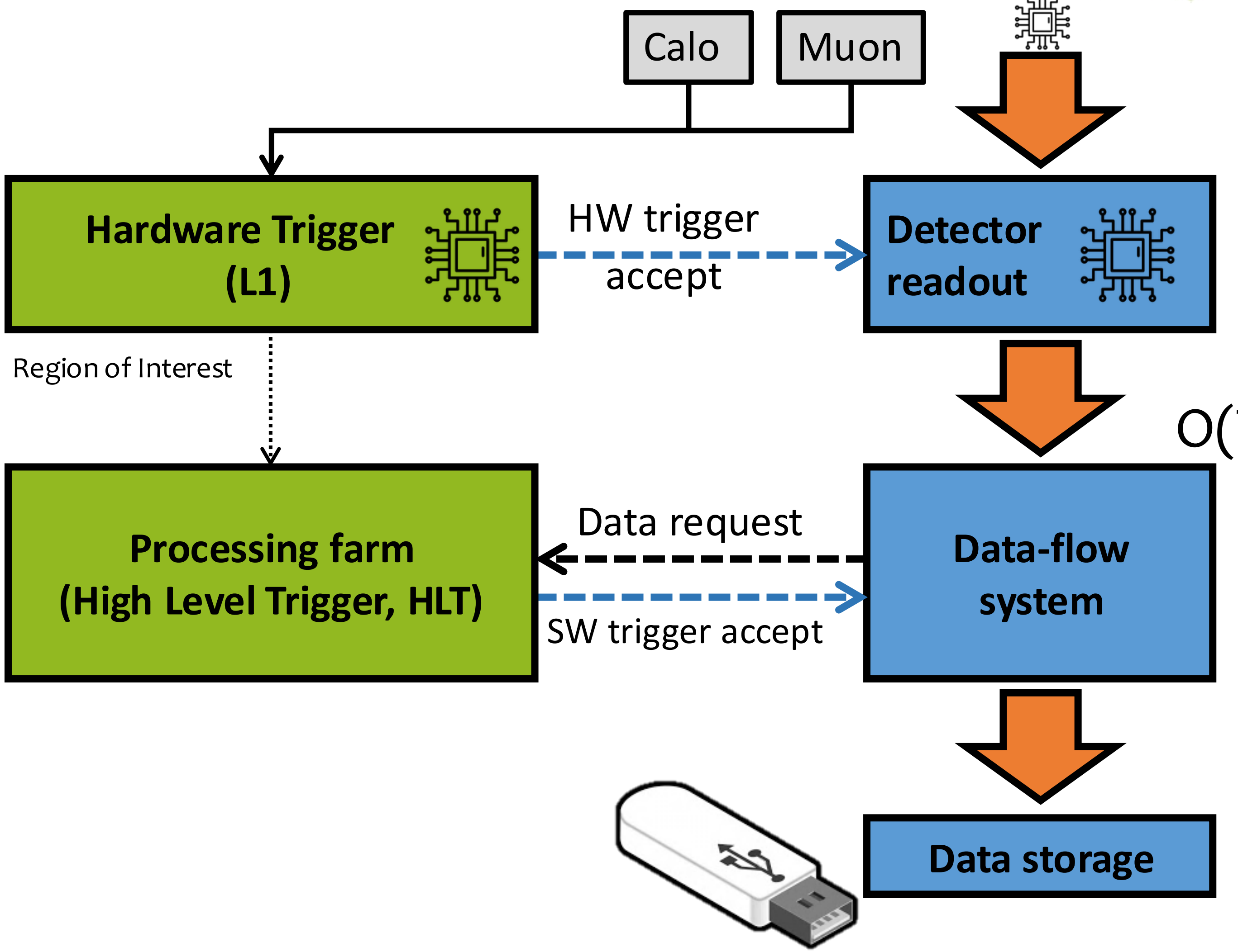
Tile calorimeter: **10 k channels**
Dual-PMT analogue readout

7'000 trigger 'towers'
100'000 'supercells'

Triggering in Physics

Architecture: Very simplified view





Now

40 MHz
(pile-up ~60)

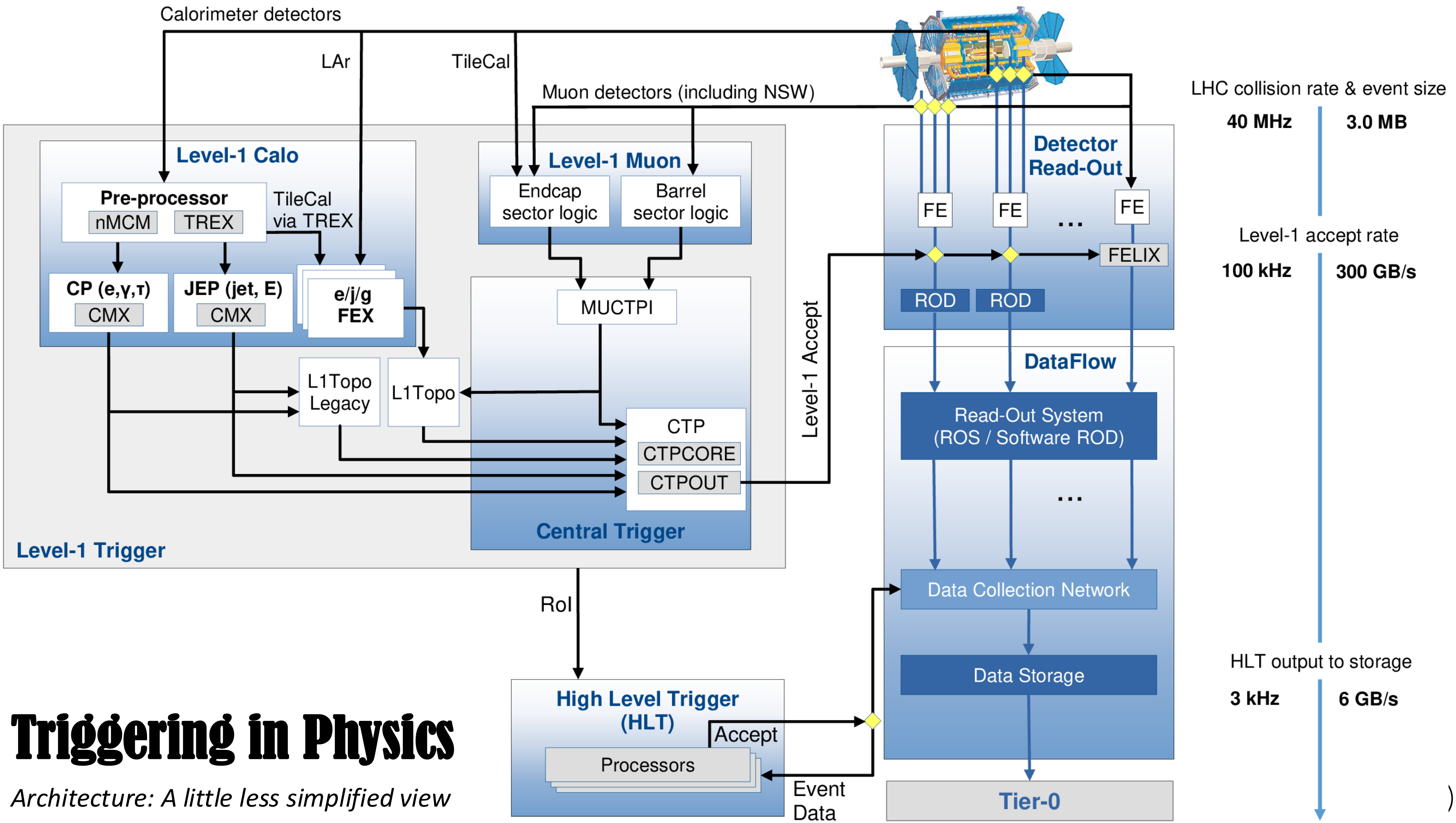
Max 2.5 µs
HW trigger accept
O(Tb/s), max 100 kHz

O(0.1-1 s)

Storage
O(Gb/s), O(kHz)

Triggering in Physics

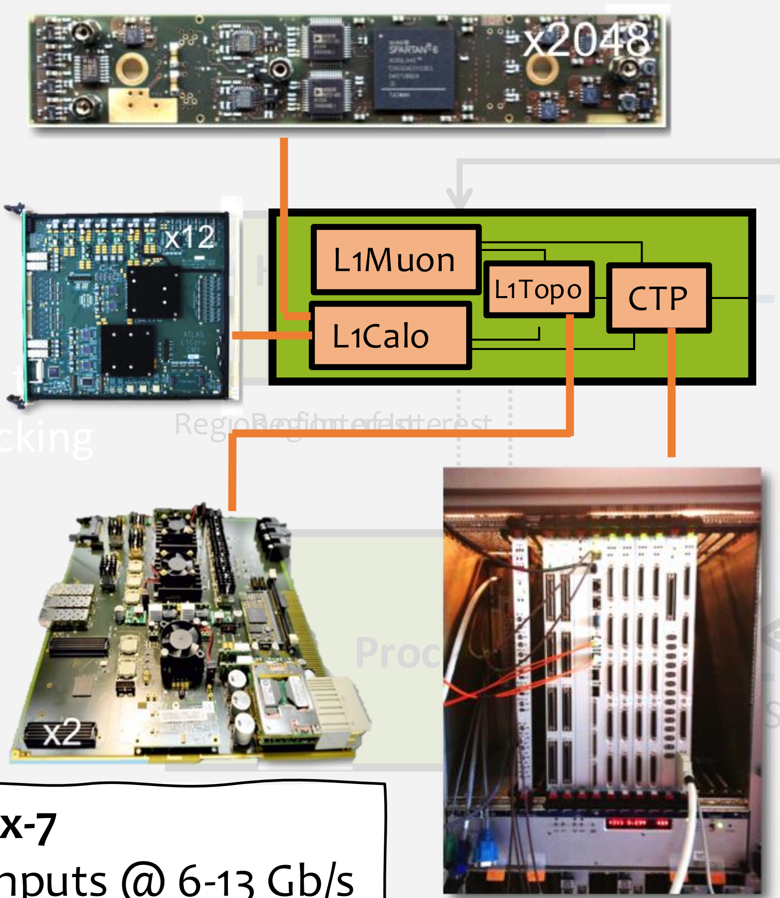
Architecture: A little less simplified view



Triggering at L1

Architecture: Very simplified view

- Low latency
- Deterministic
- Low jitter clocking



All L1 HW boards are custom-made boards
- Primarily using Virtex-6, Virtex-7, and UltraScale/UltraScale+ families, depending on subsystem and upgrade phase
- O(10–100) high-speed optical links per board
- 40 MHz synchronous operation
- Some **examples** given here

L1Calo – VU9P and VU13P
- massive I/O
  - 48-96 links per FPGA @ 25-28 Gb/s
  - O(Tb/s) input bandwidth
- $O(10^6)$ LUTs and $O(10^4)$ DSP slices
- Compute heavy

CTP – Virtex-7
- $O(10^2)$ inputs
- Accept rate 100 kHz
- Fixed latency O(10 ns)
- Low jitter clocking < 100 ps

L1Topo – Virtex-7
- 80 optical inputs @ 6-13 Gb/s
- Fixed latency O(100 ns)
- $O(10^6)$ LUTs

Now

30 MHz

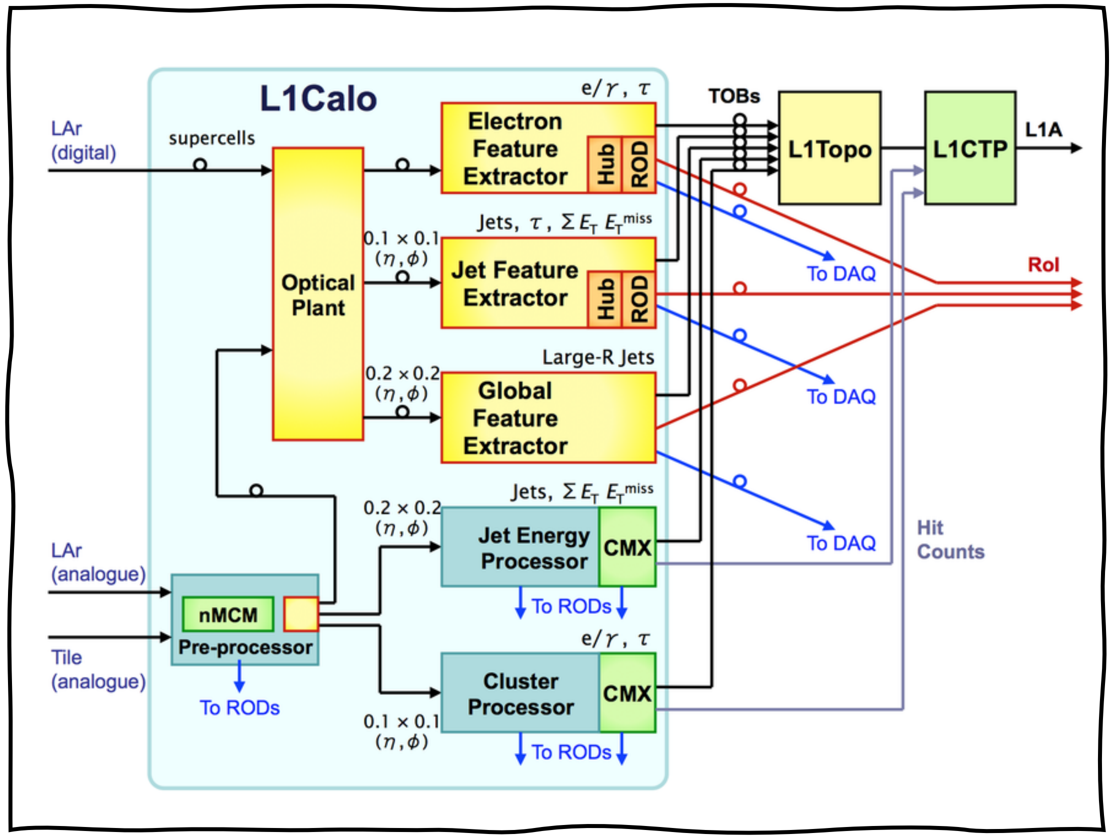
(pile-up ~60)

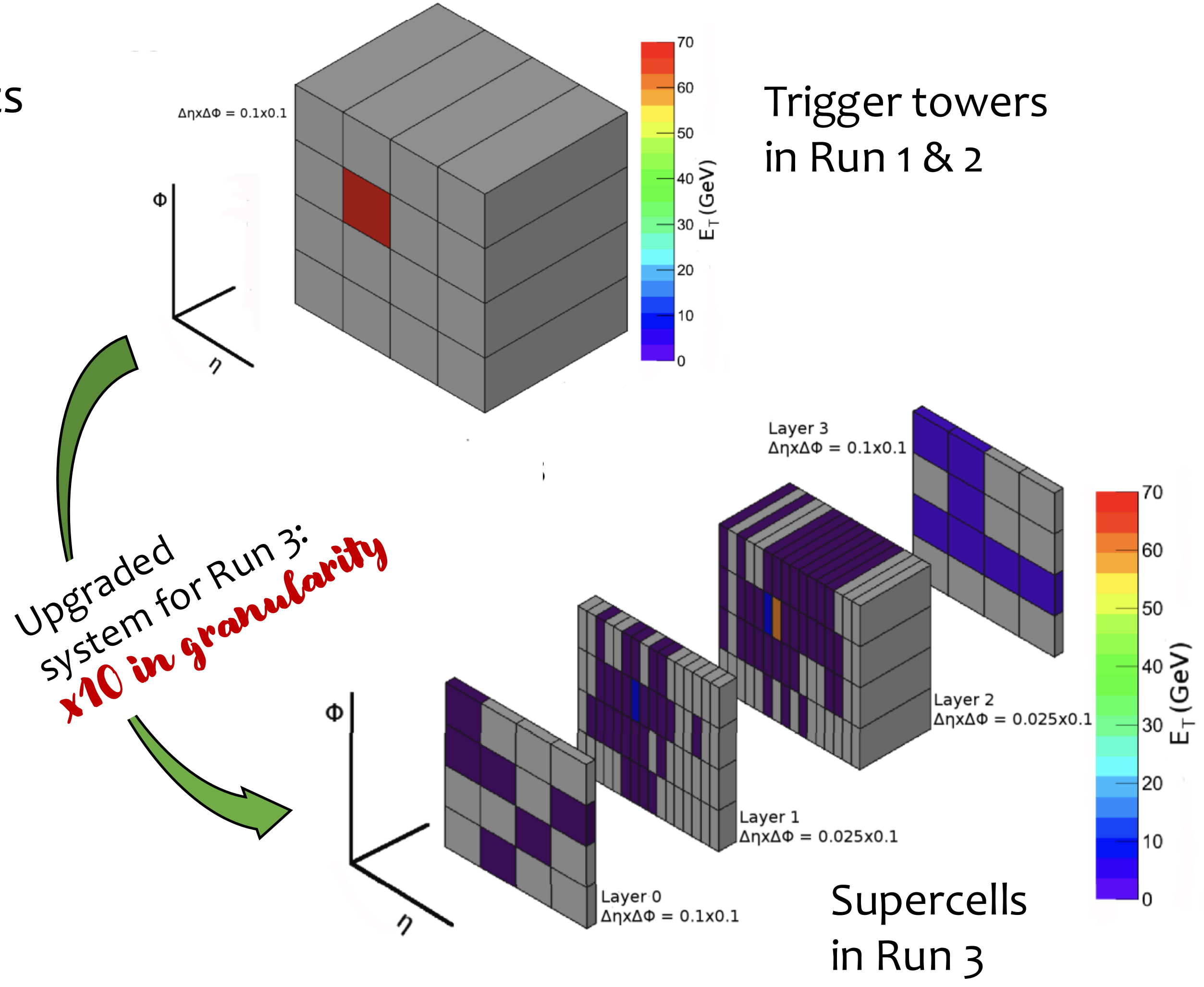
Storage

O(Gb/s), O(kHz)

Algorithms at L1? E.g. L1Calo

Many different boards with dedicated functionality, each running different algorithms using similar inputs

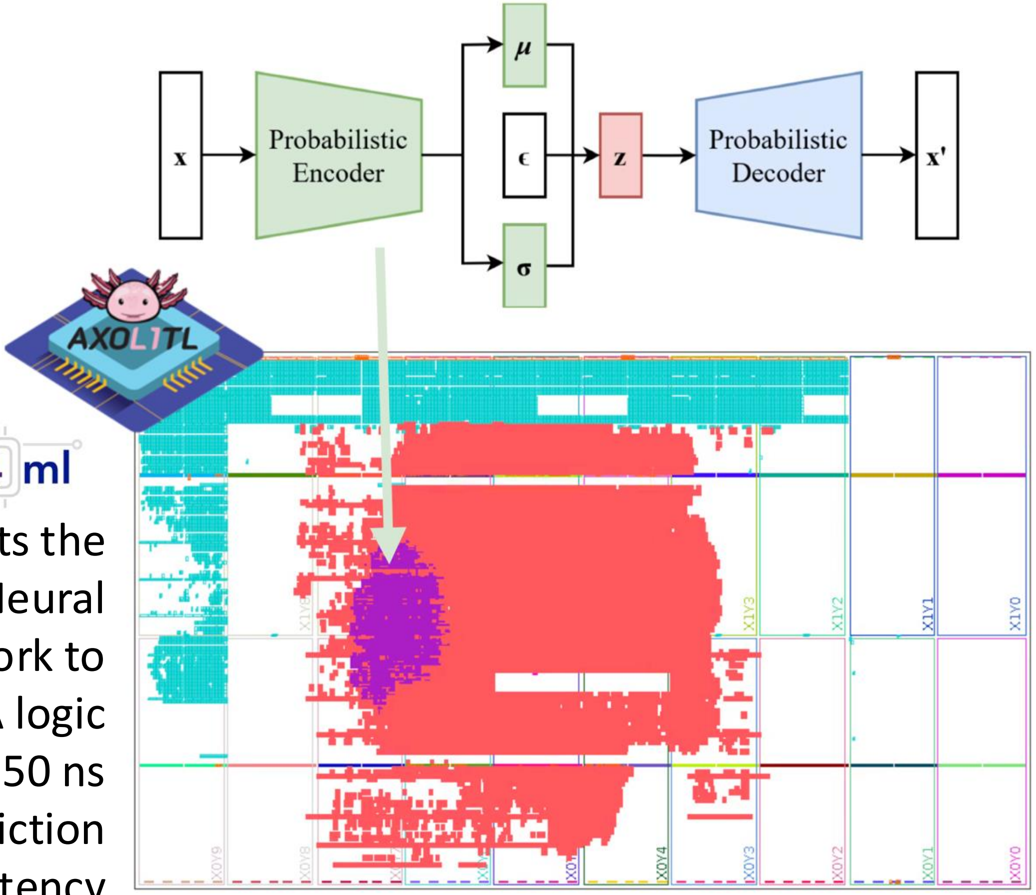




https://cds.cern.ch/record/1602230/; https://cds.cern.ch/record/1602235

Algorithms at L1? New: Anomaly Detection

First done at **CMS** as a Variational-Autoencoder model. Now also on ATLAS.

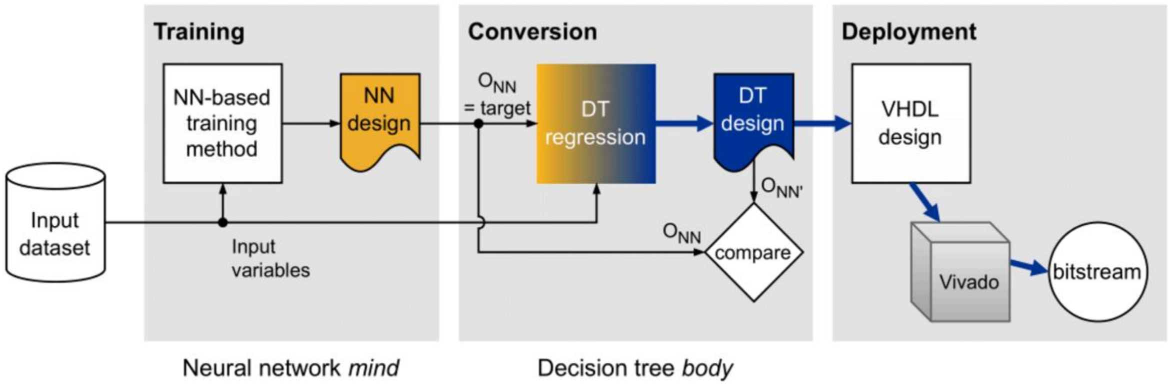


hls4ml converts the Neural Network to FPGA logic with 50 ns prediction latency

E.g. see https://indico.nikhef.nl/event/4875/contributions/20303/

Besides hls4ml-based implementation, ATLAS is also pursuing VAE+BDT Anomaly detection algorithm implemented with FWxMachina, optimised for BDT

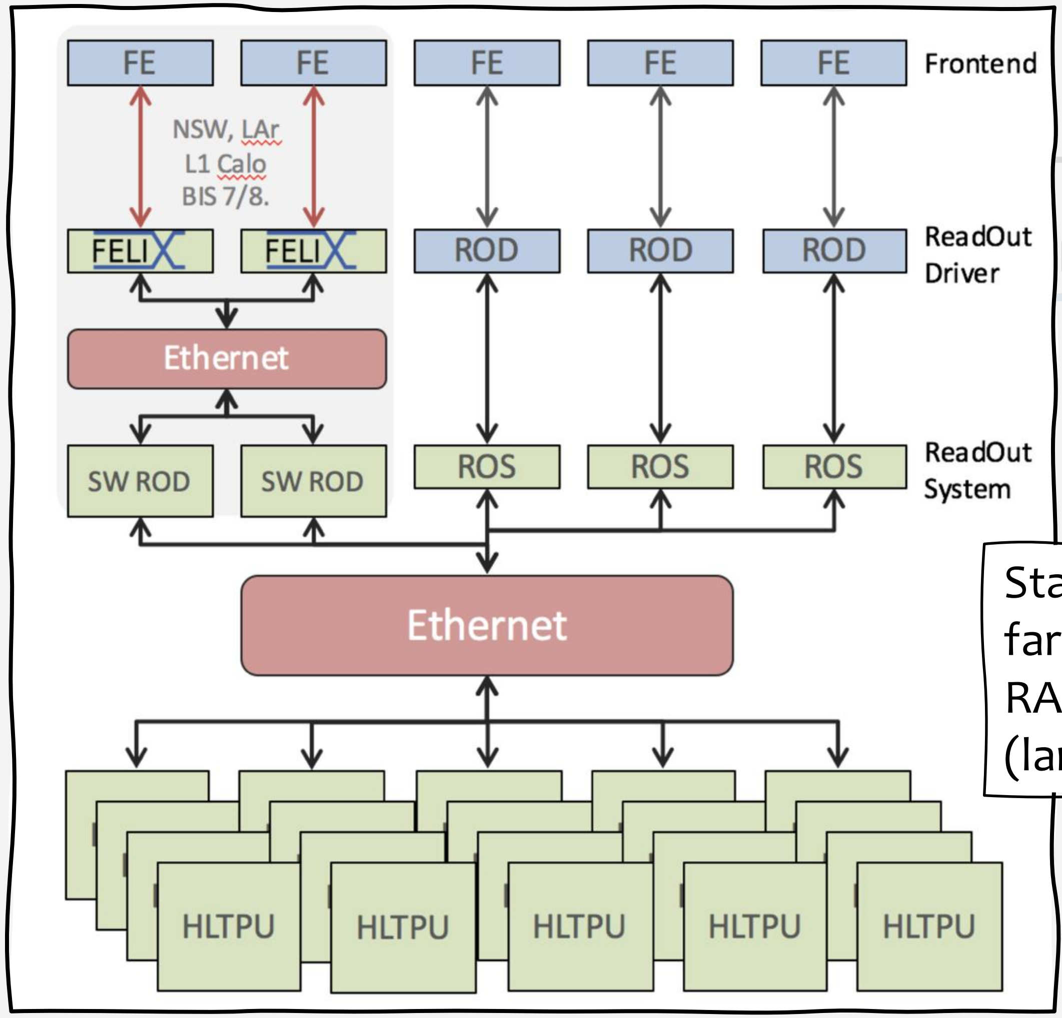
FWXMachina

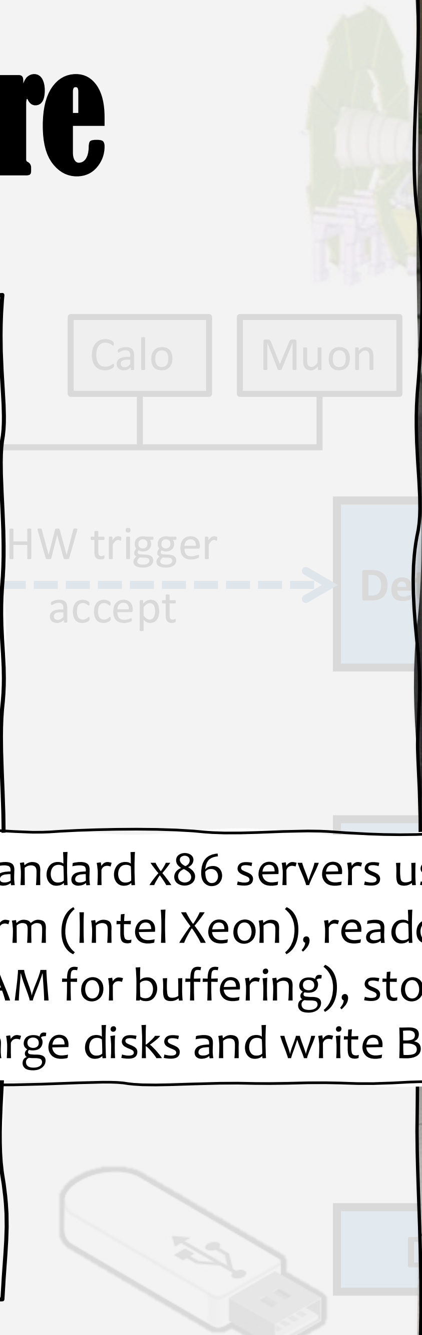


E.g. https://arxiv.org/pdf/2104.03408, https://www.fwx.pitt.edu/

Commodity hardware

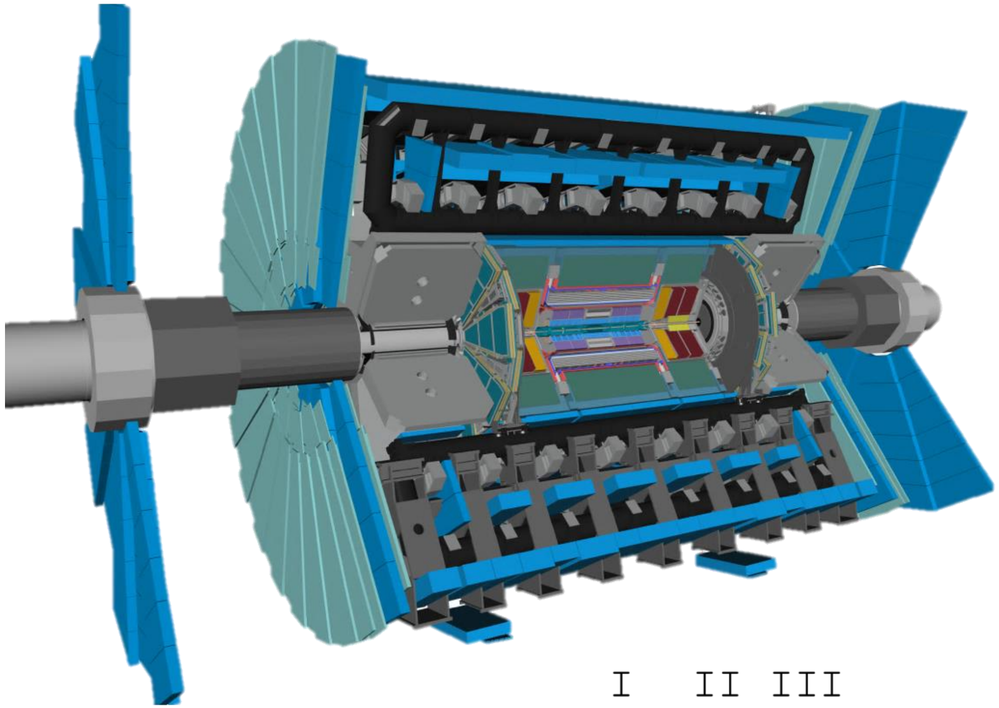
Architecture: Very simplified view

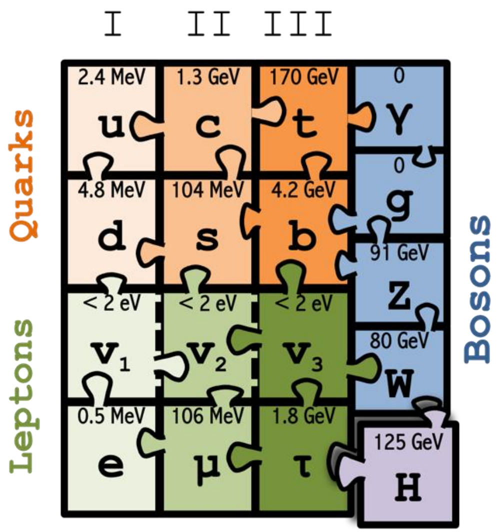




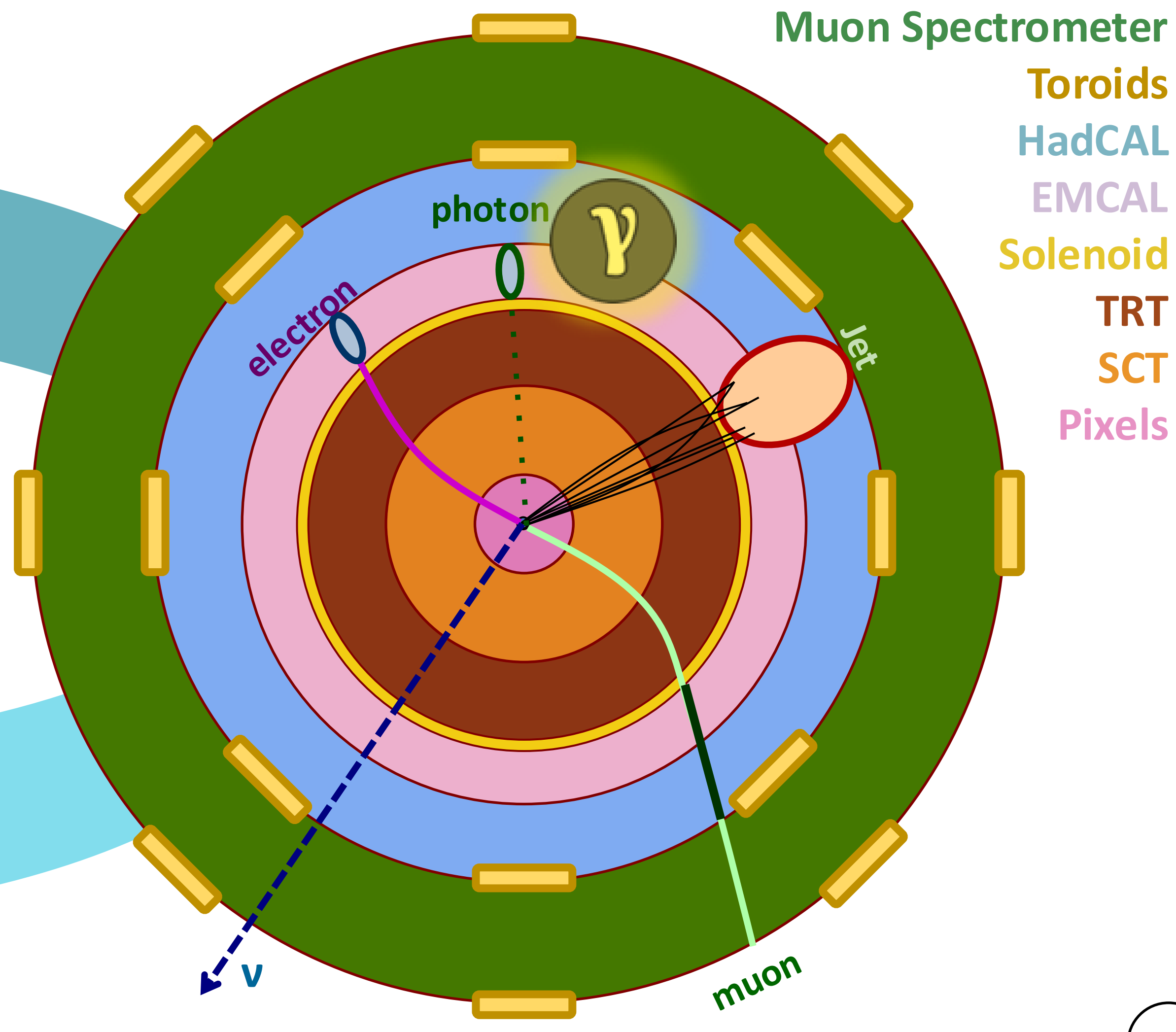
Standard x86 servers used for processing farm (Intel Xeon), readout system (large RAM for buffering), storage management (large disks and write BW)

What do we actually do at the trigger?

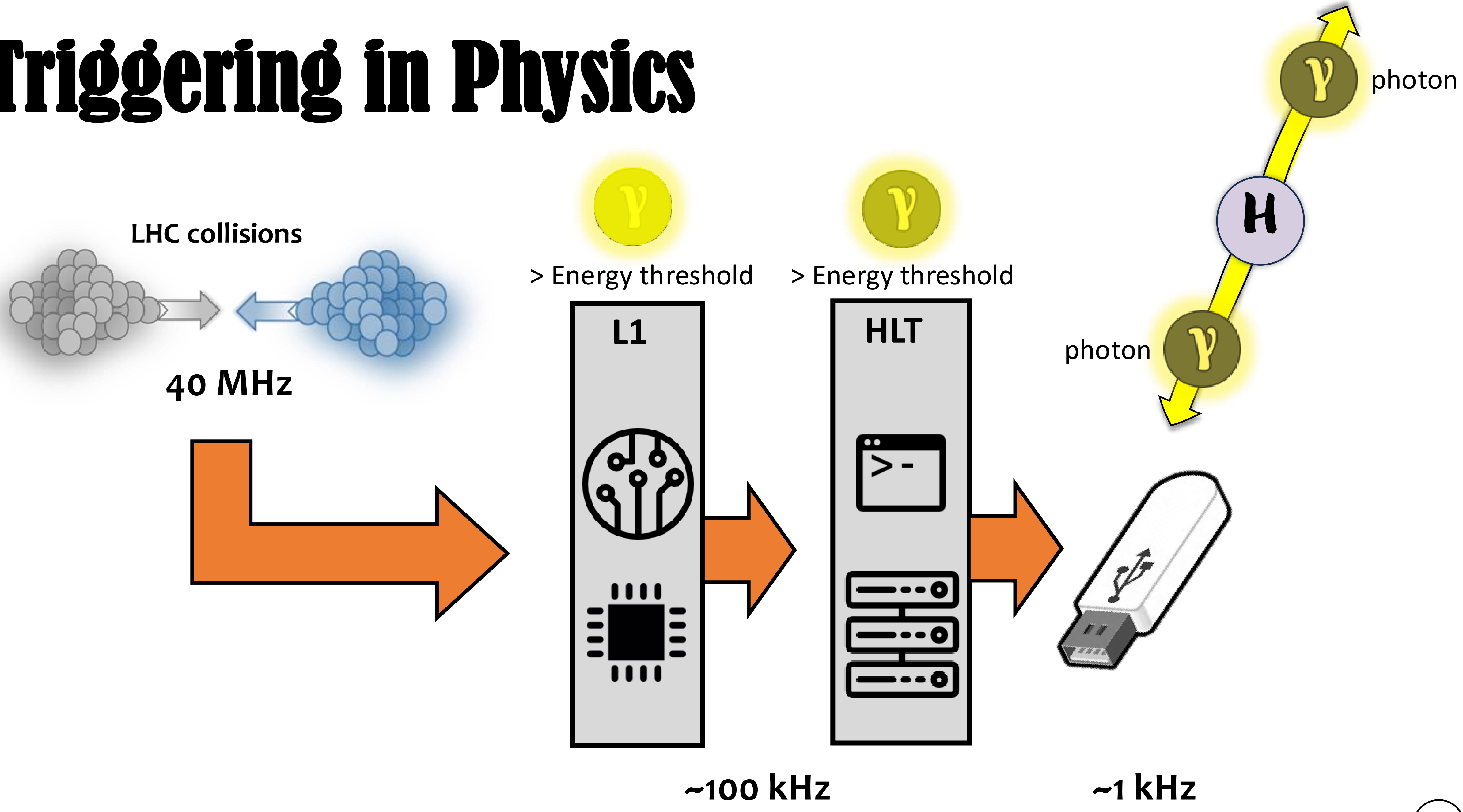




Simplified Detector Transverse View

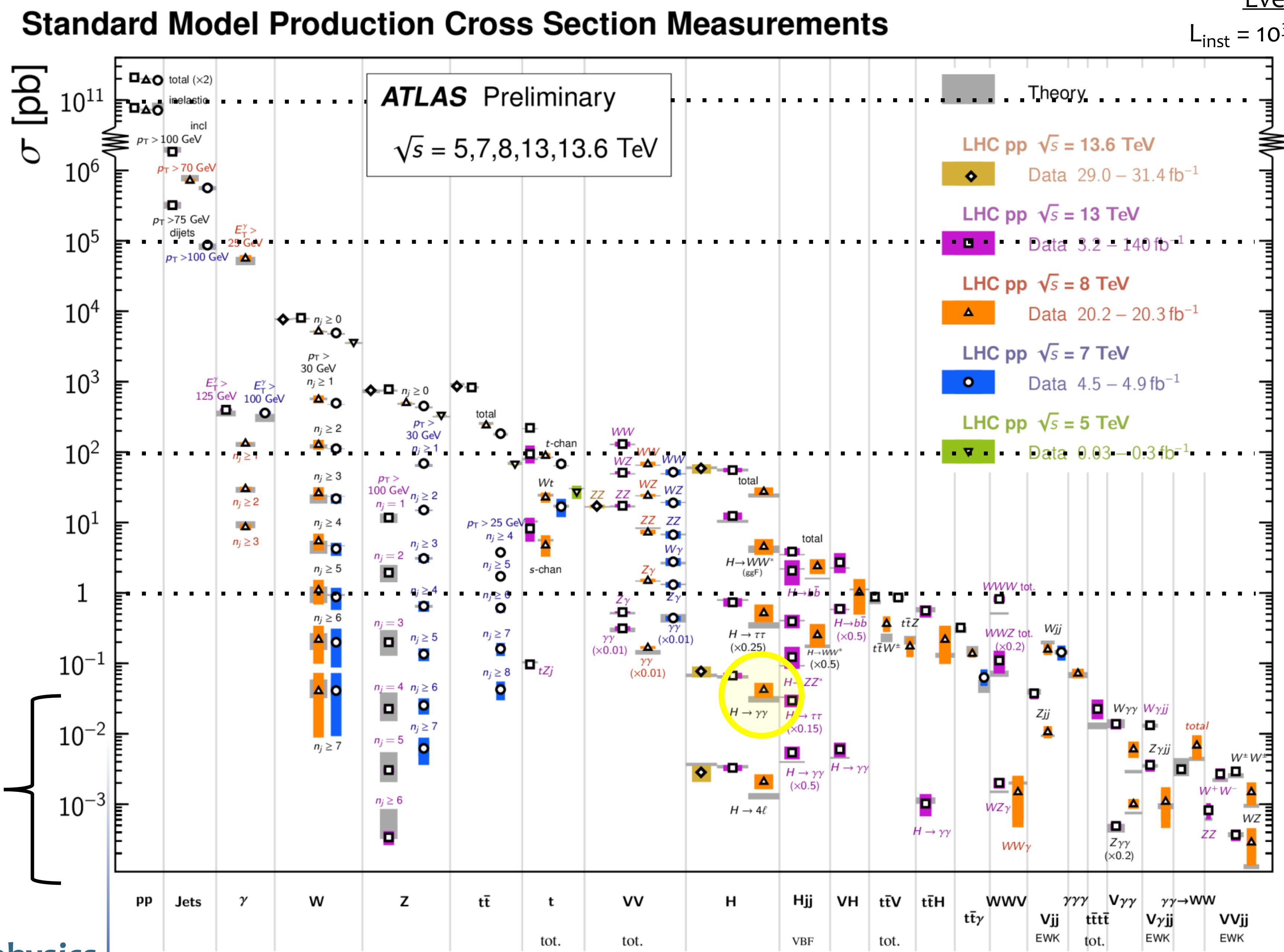


Triggering in Physics



Reminder: $\sigma = \frac{\#\ \text{events}}{L}$

Status of Standard Model



Event Rate

$L_{inst} = 10^{34}\ cm^{-2}s^{-1}$

1 GHz

1 kHz

1 Hz

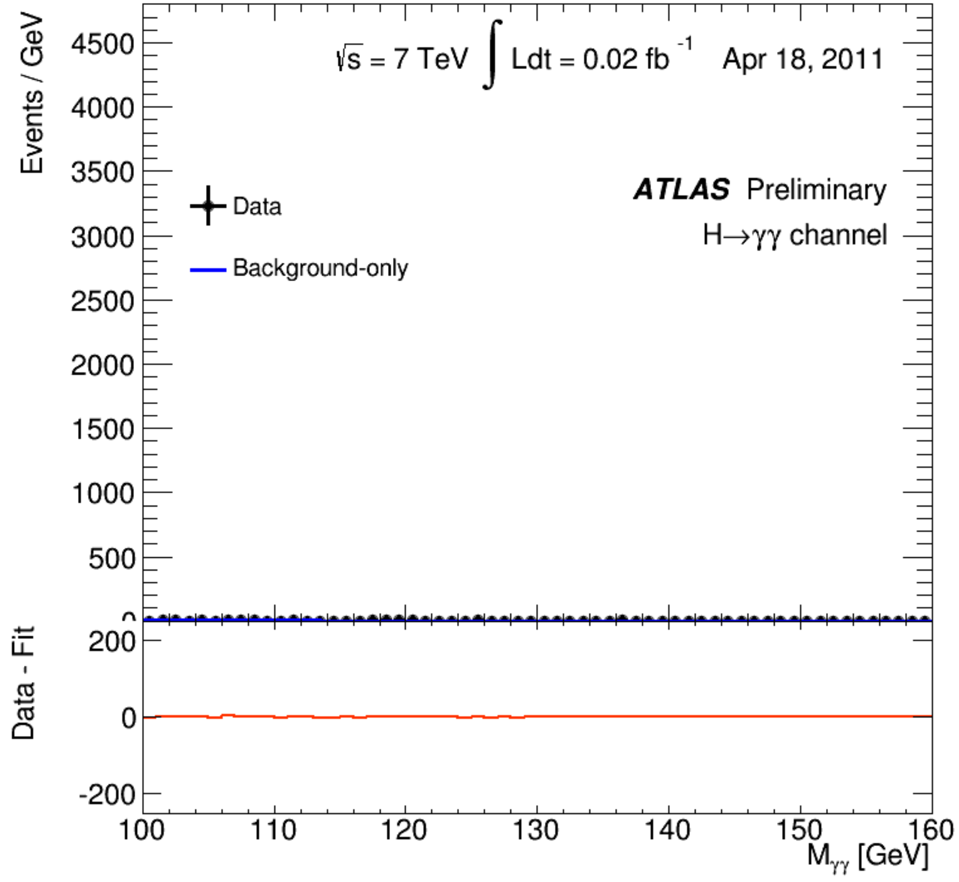
10^{-2} Hz

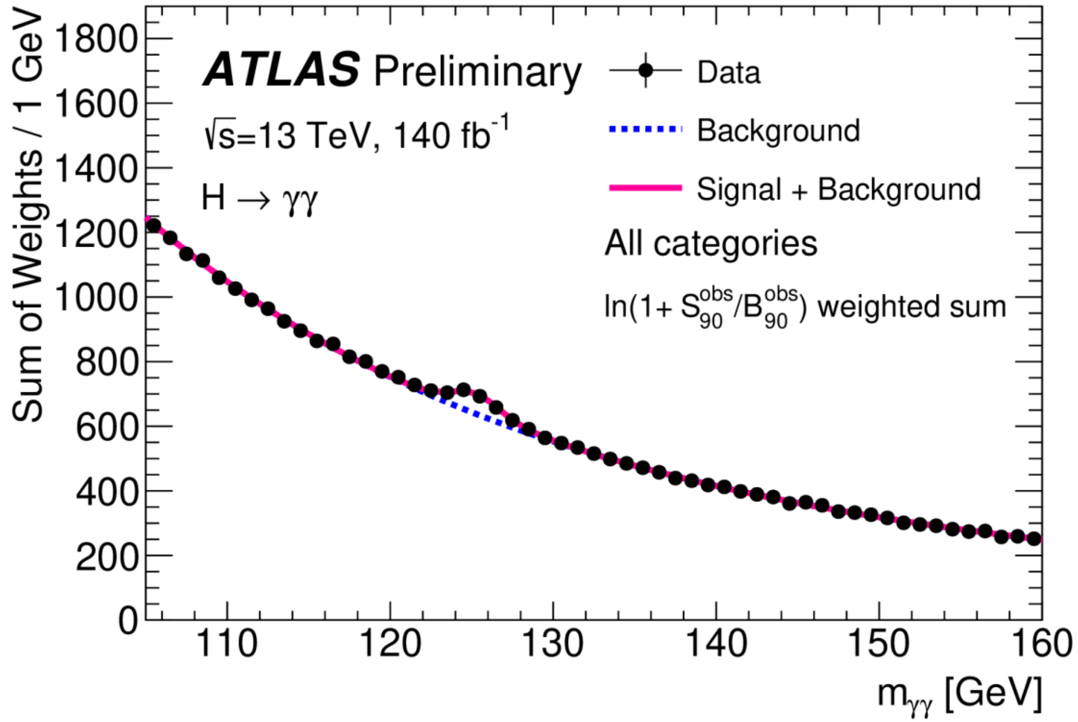
Key aspects of Standard Model still to be fully explored

New physics

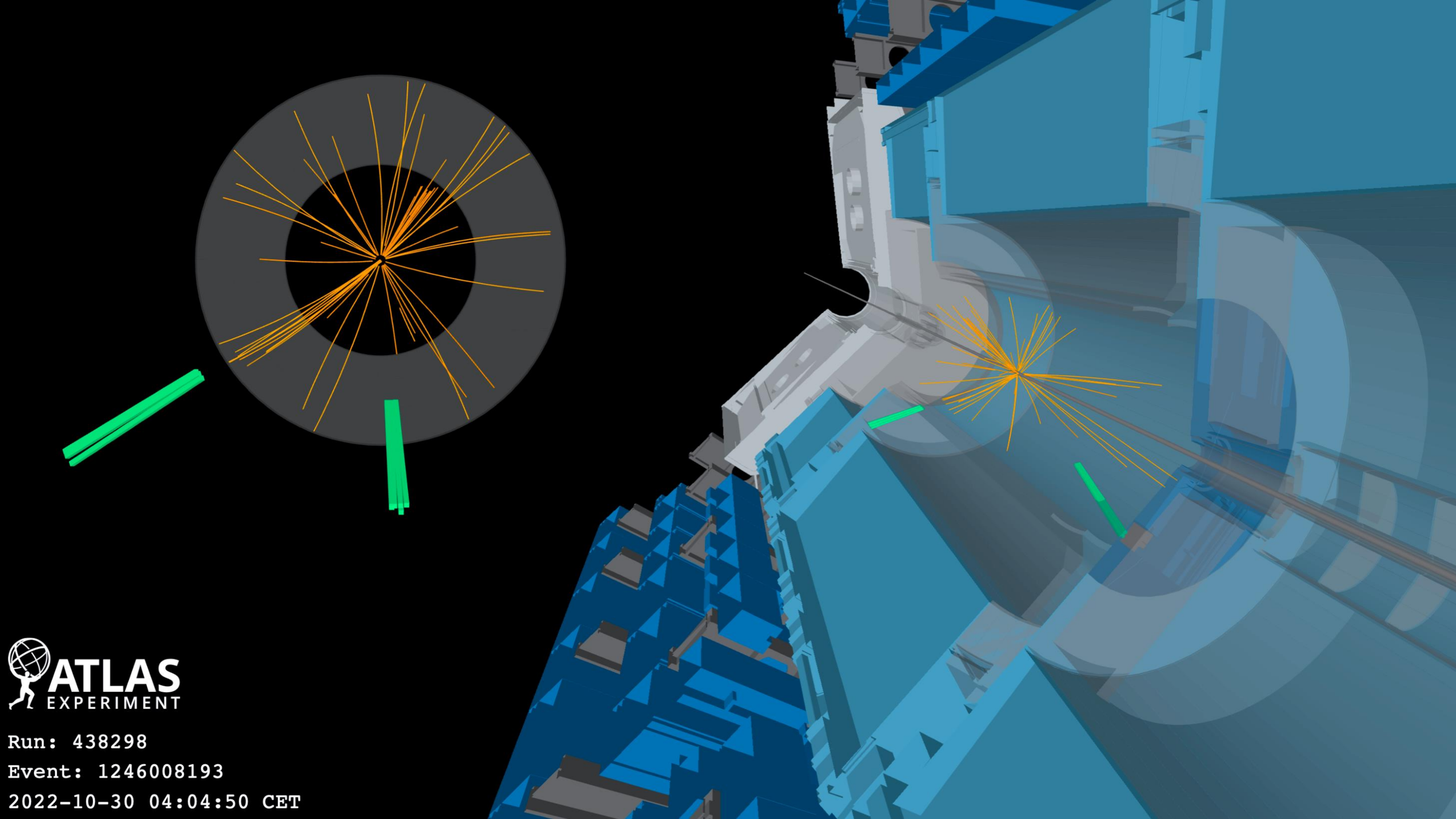
Photon triggers

among those used to discover the Higgs boson…

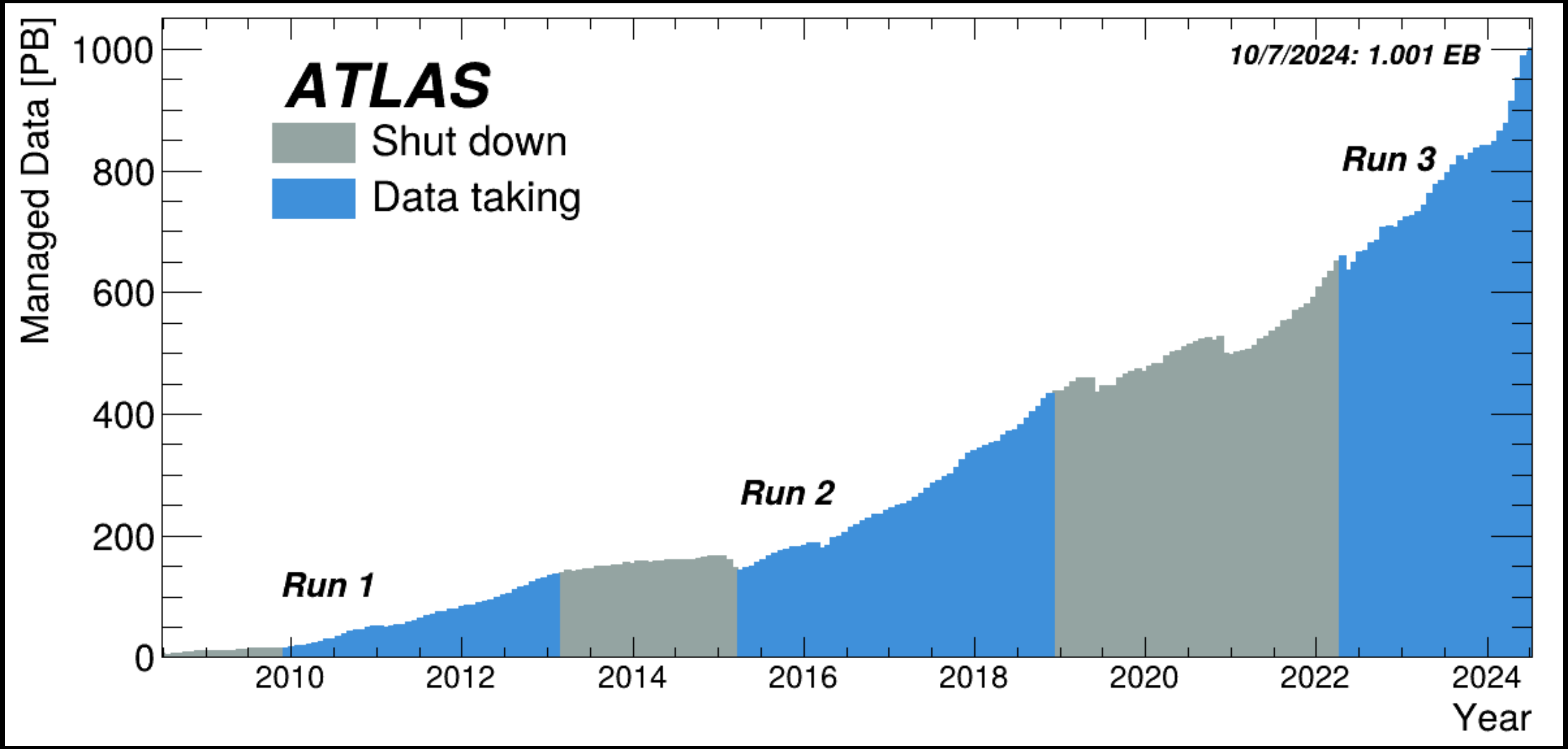




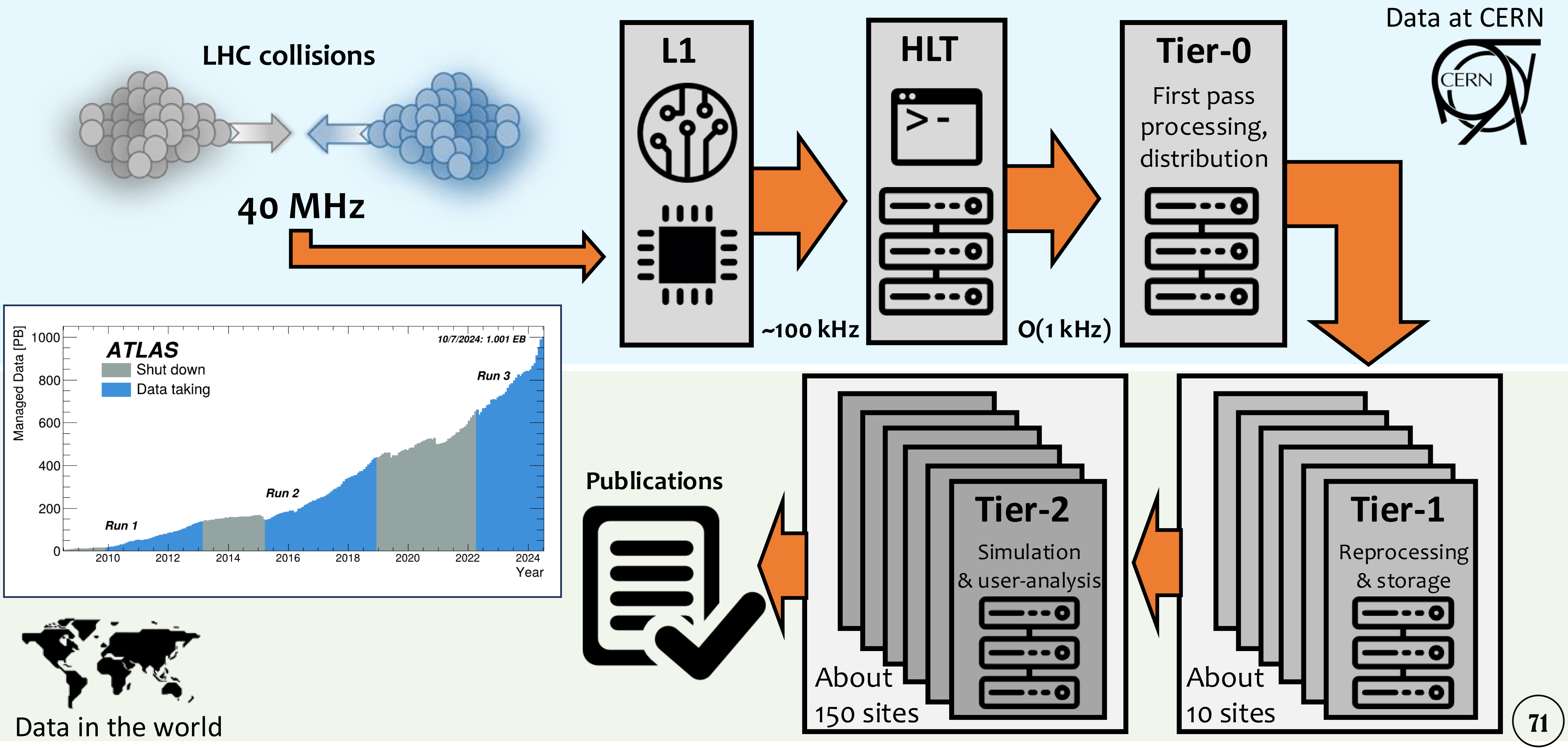
…and still used to keep on studying its properties, e.g. its mass


ATLAS
EXPERIMENT
Run: 438298
Event: 1246008193
2022-10-30 04:04:50 CET

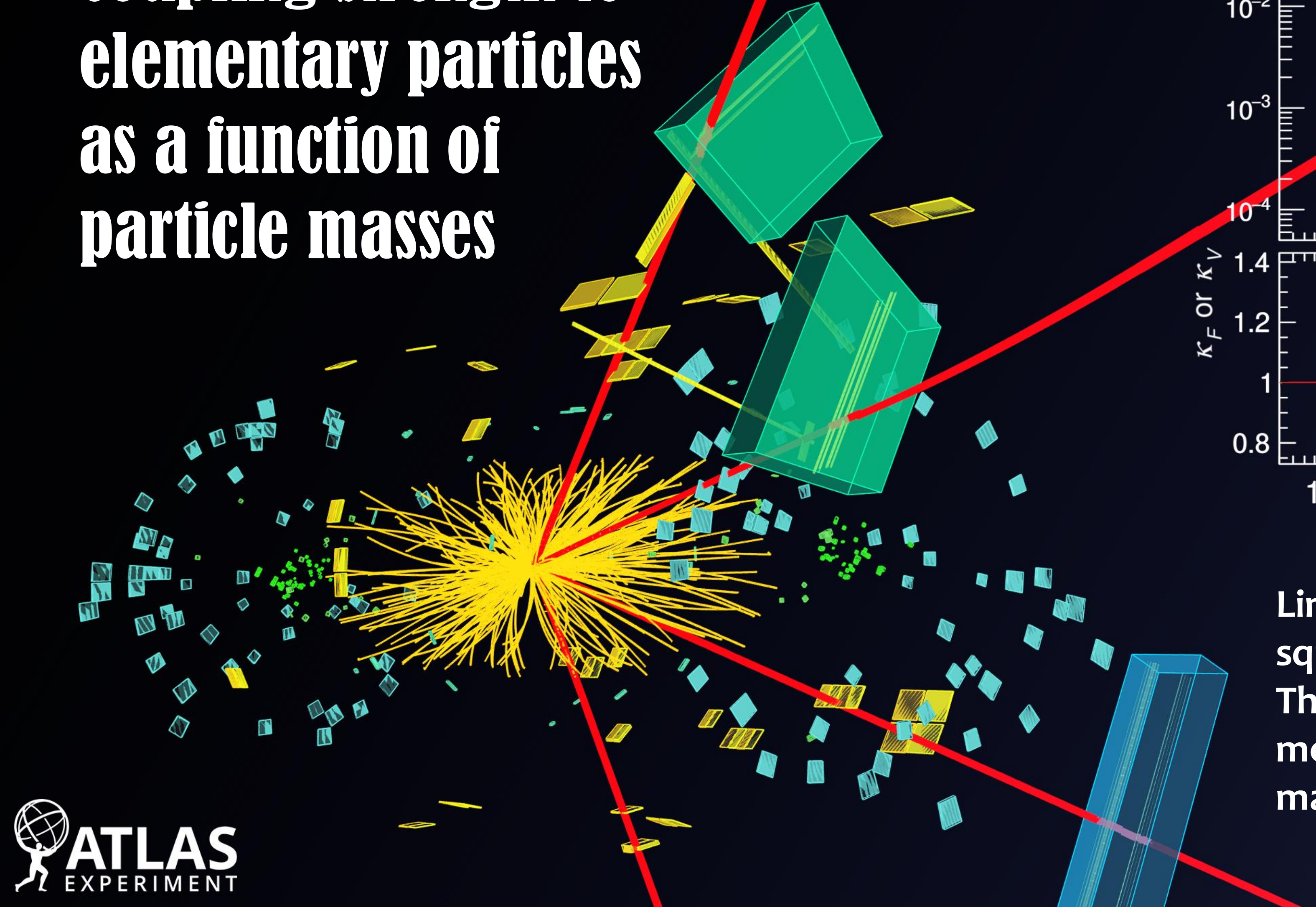
ATLAS Computing – data size

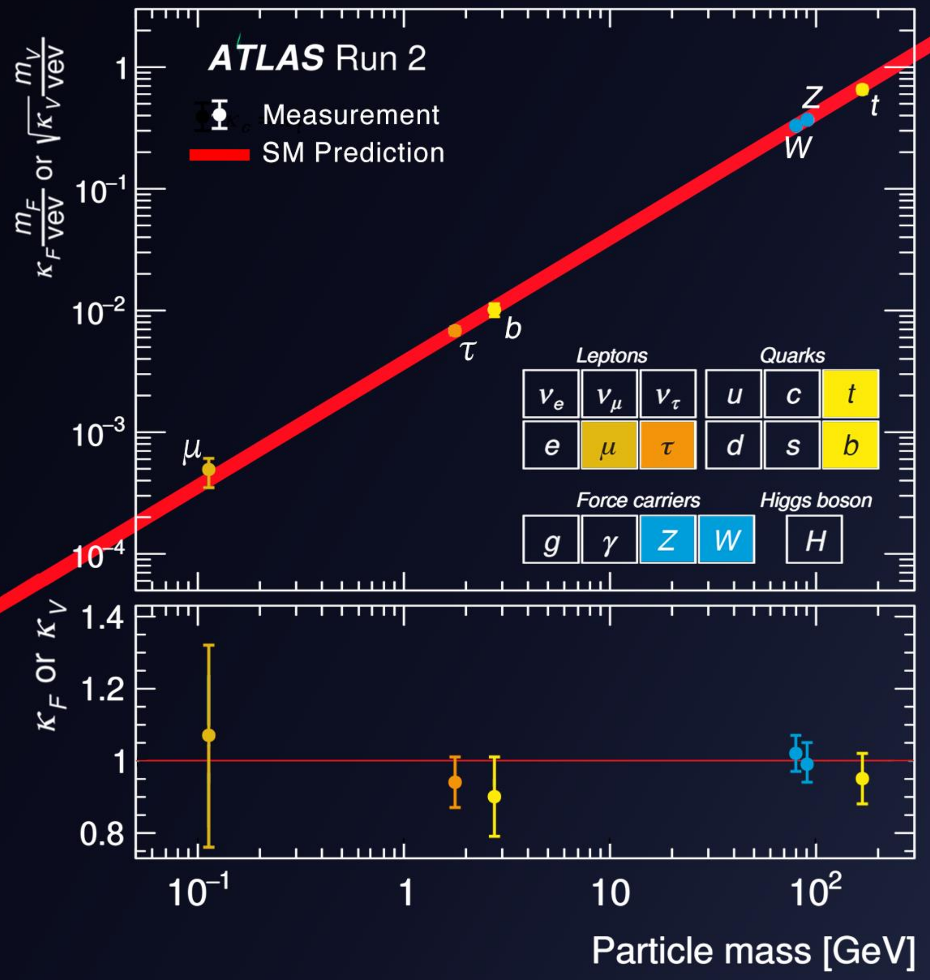


ATLAS Computing Path

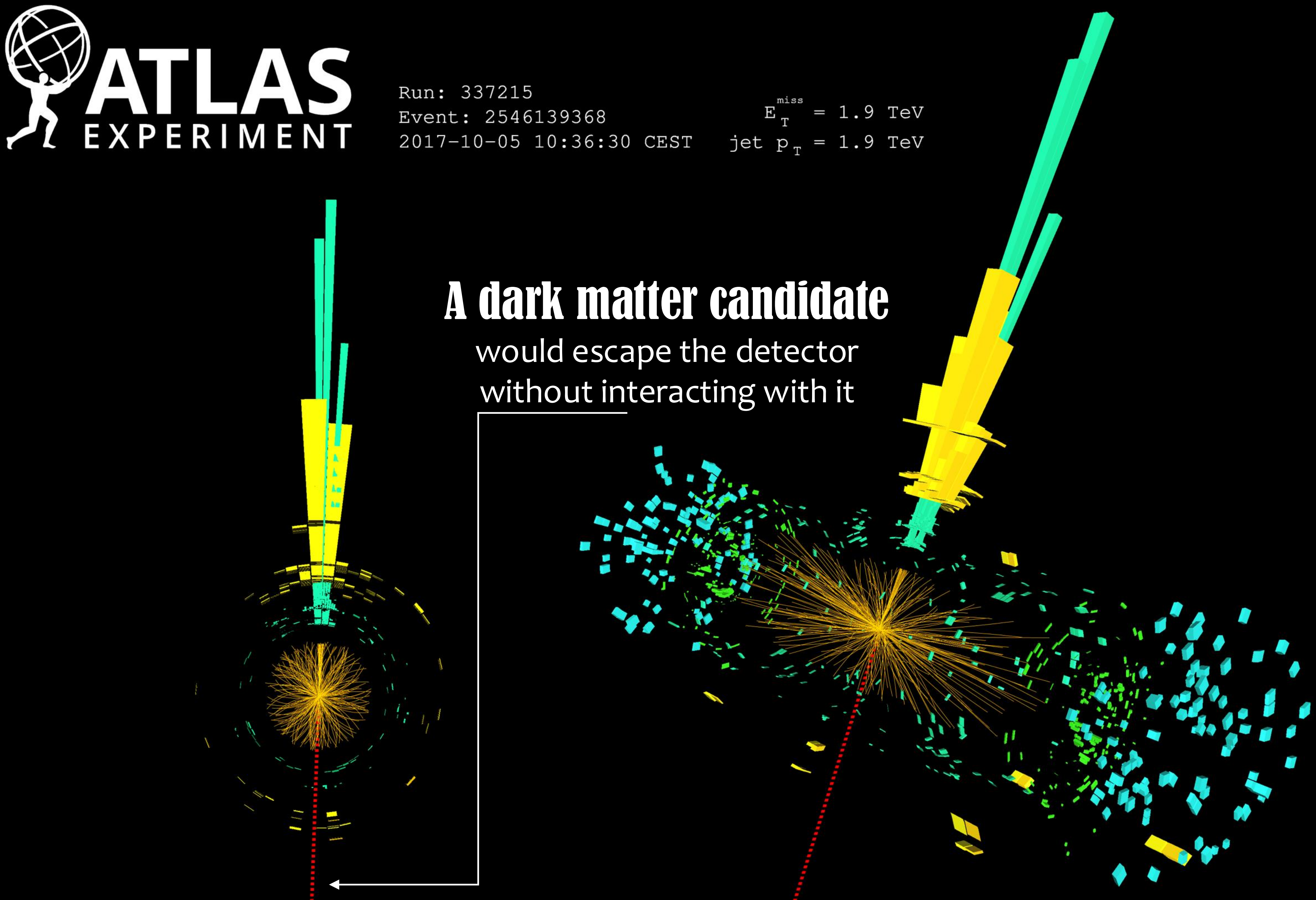


Measured modifiers of the Higgs boson coupling strength to elementary particles as a function of particle masses



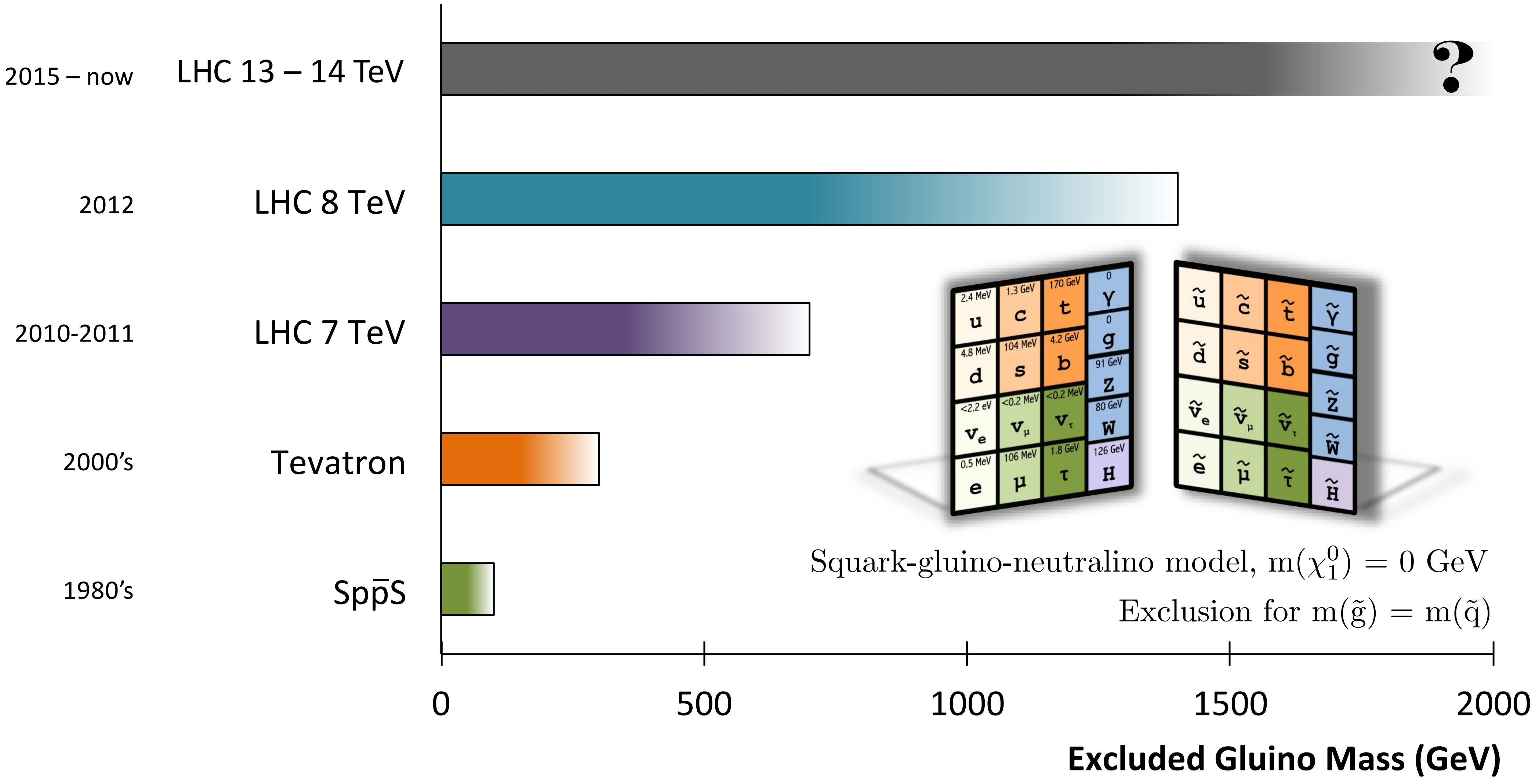


Linear dependence for fermions and square-root for weak bosons.
The diagonal red line represents the BEH mechanism's prediction of non-universal, mass-dependent interactions.



A dark matter candidate

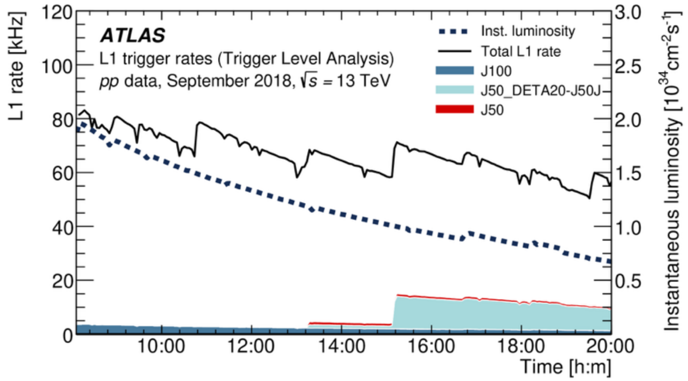
would escape the detector
without interacting with it

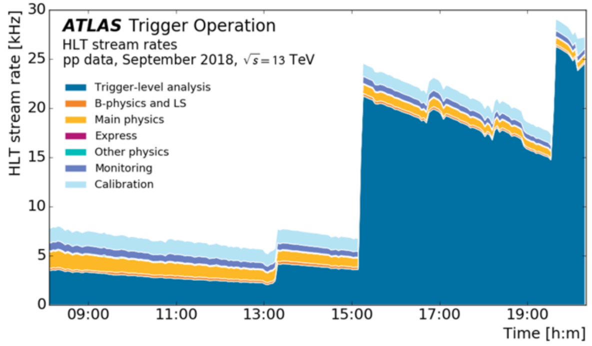


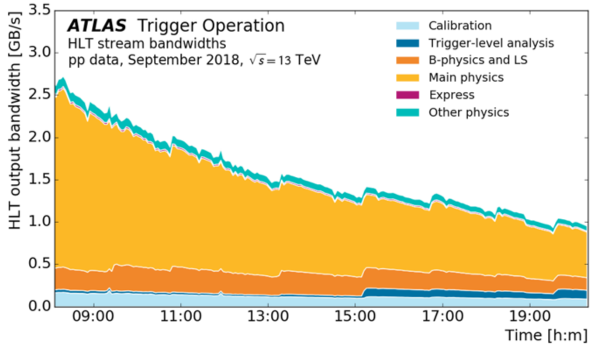
Dark matter is still one of the biggest open questions in science.

Results like this tell us where *not* to look, and bring us closer to where the answer really is.

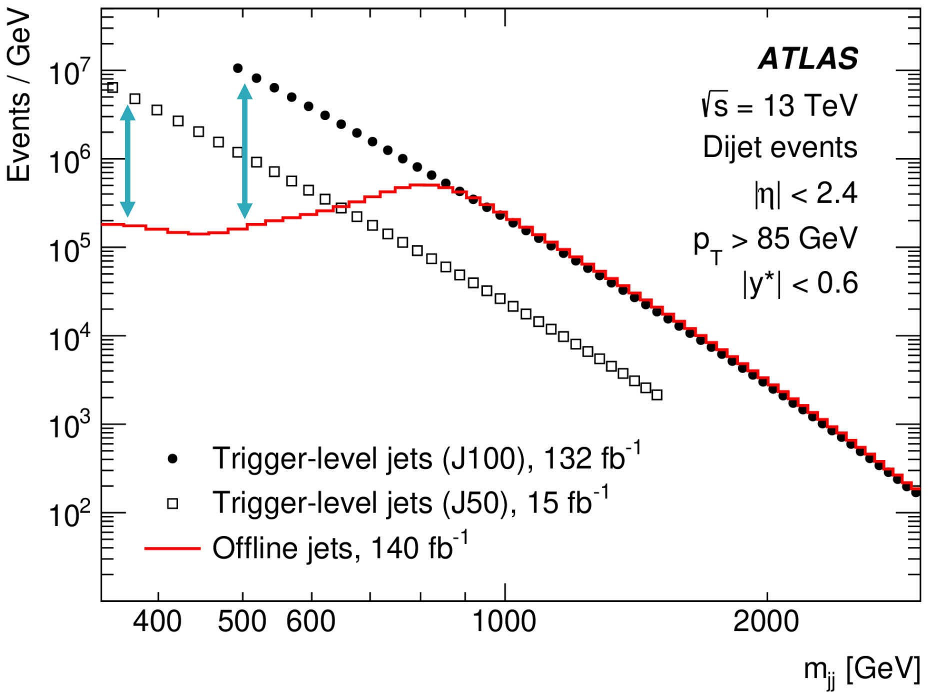
Pushing trigger-driven searches





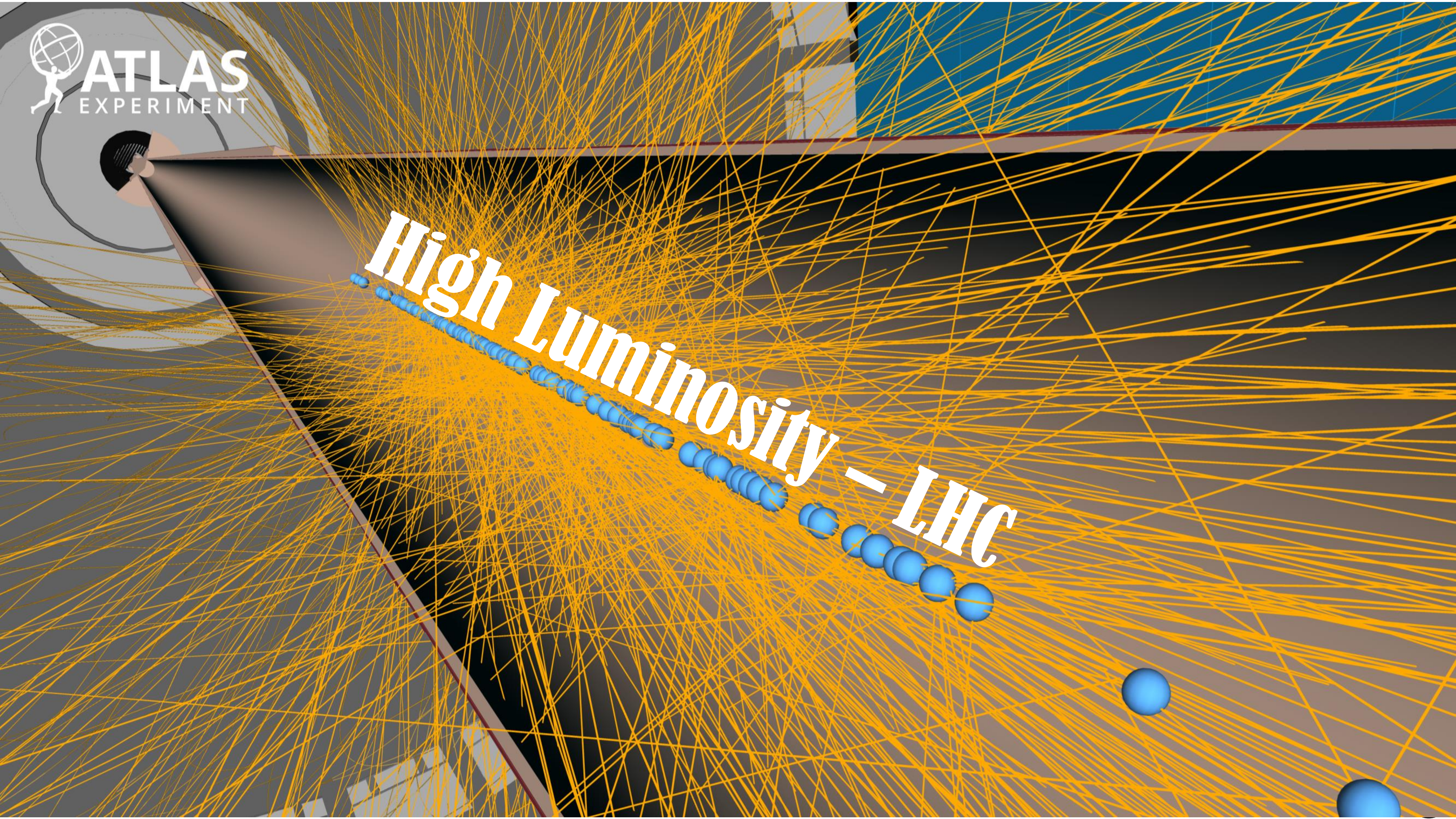


Exploiting non-standard trigger appropaches

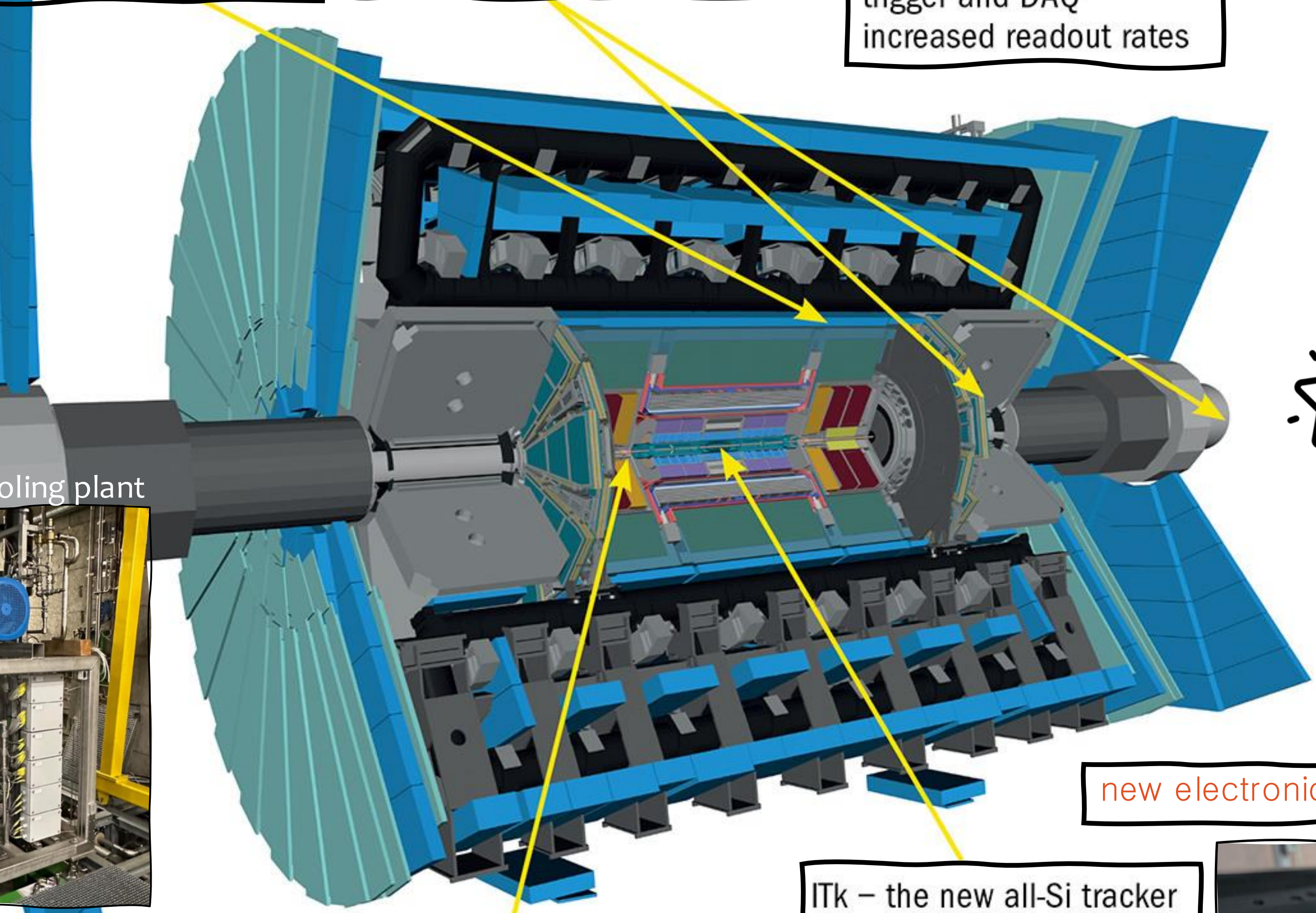


➜ **We are searching for the unknown:**
we have to be creative to make sure we capture it in our data !

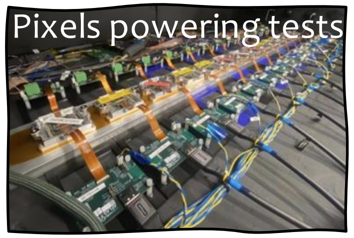
arXiv:2509.01219

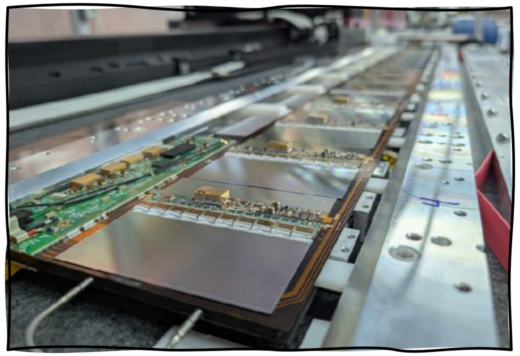

ATLAS
EXPERIMENT
High Luminosity – LHC

at HL-LHC



CO2 cooling plant



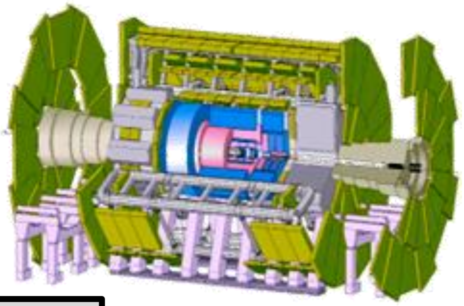


Moving from preparing to producing
The Phase-II detector is becoming real !

5 Billion channels
(reminder: 100 M now)

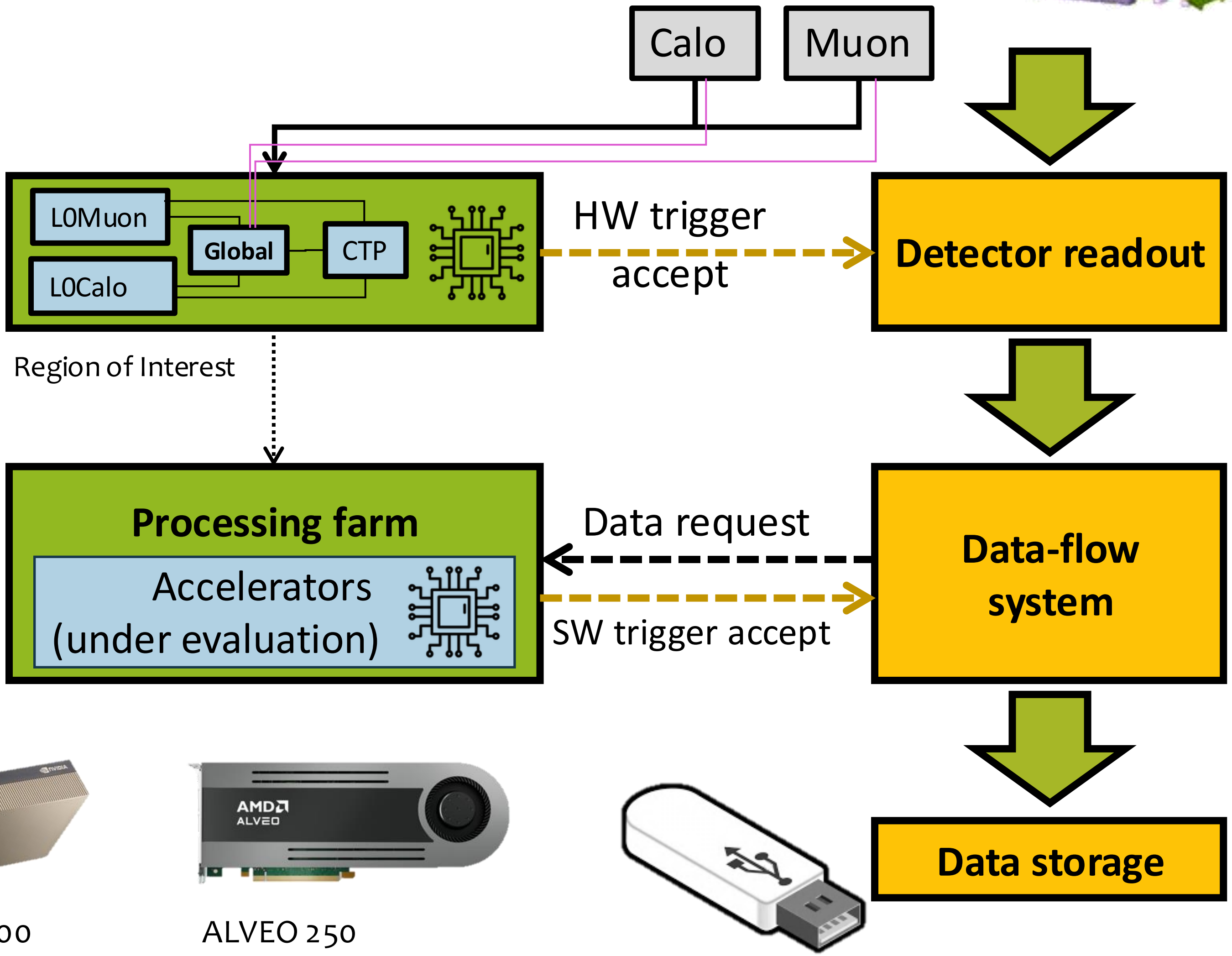
ATLAS Trigger

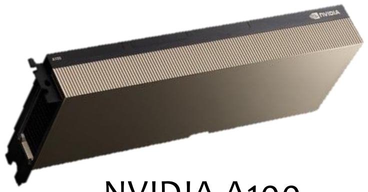
Architecture: Very simplified view



- *Higher granularity inputs to L1 trigger ('Global')*
- *Higher latency: improved algorithms*

- *Baseline: CPU farm*
- *GPUs and FPGAs being evaluated as accelerators*





NVIDIA A100

ALVEO 250

In HL-LHC

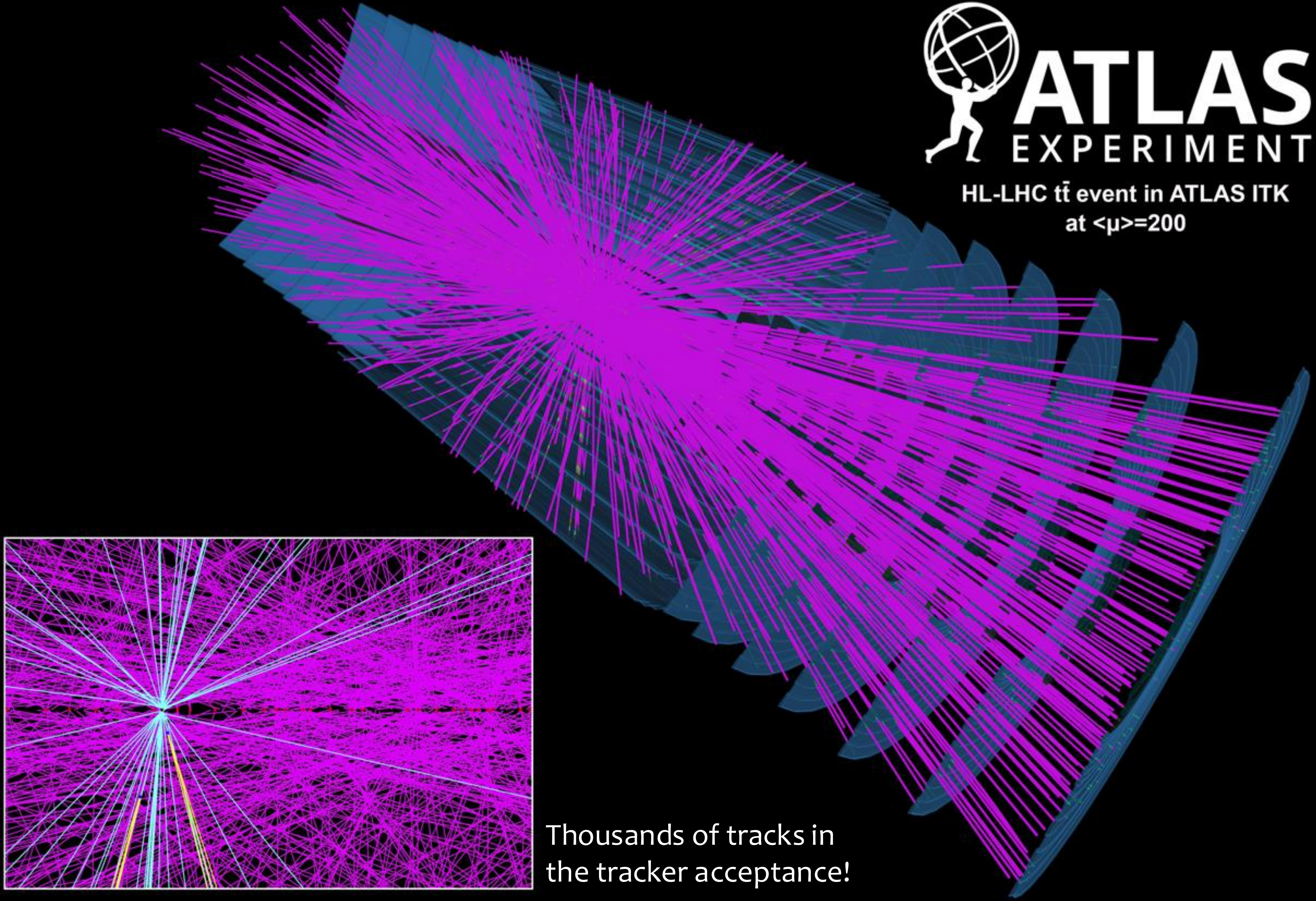
40 MHz
pile-up ~200
(now: 60)

Max 10 µs
(now: 2.5 µs)

HW trigger accept
Max 1 MHz
(now: Max 100 kHz)

O(100 ms)

Storage
O(10 kHz)
(now: O(kHz))



HL-LHC $t\bar{t}$ event in ATLAS ITK at $\langle\mu\rangle=200$

Thousands of tracks in the tracker acceptance!

ATLAS EXPERIMENT

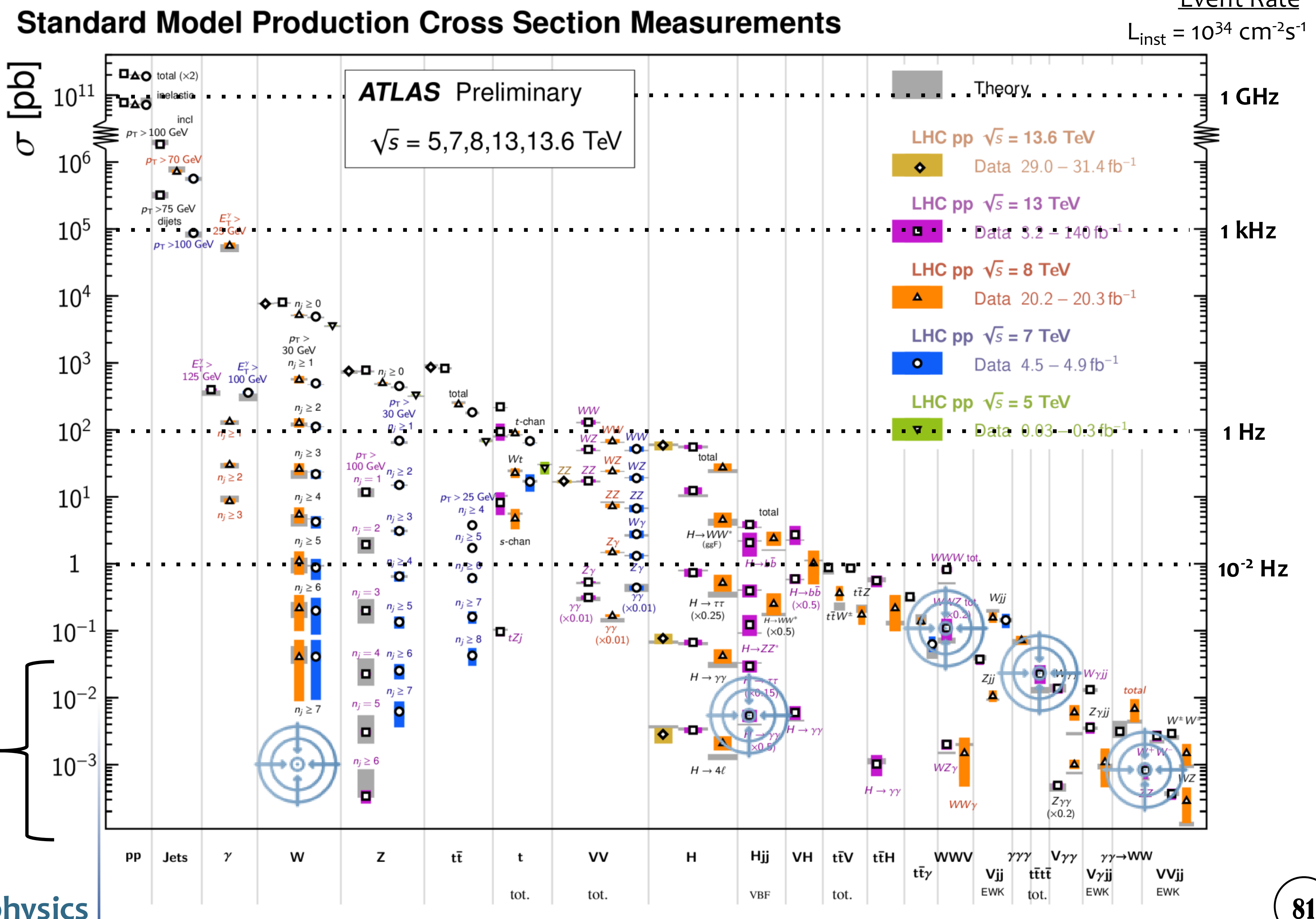
HL-LHC $t\bar{t}$ event in ATLAS ITK at $\langle\mu\rangle=200$

HL-LHC triggering challenges (and opportunities)

- Extreme pile-up and high object multiplicities drive significantly increased algorithic complexity to maintain the physics
- Heavy use of ML for reconstruction, identification, and selection
- Challenging FPGA deployement at L1 and HLT driven by algorithmic complexity and tight latency, bandwidth, and resource constraints
- R&D required on algorithm quantisation and firware-aware design

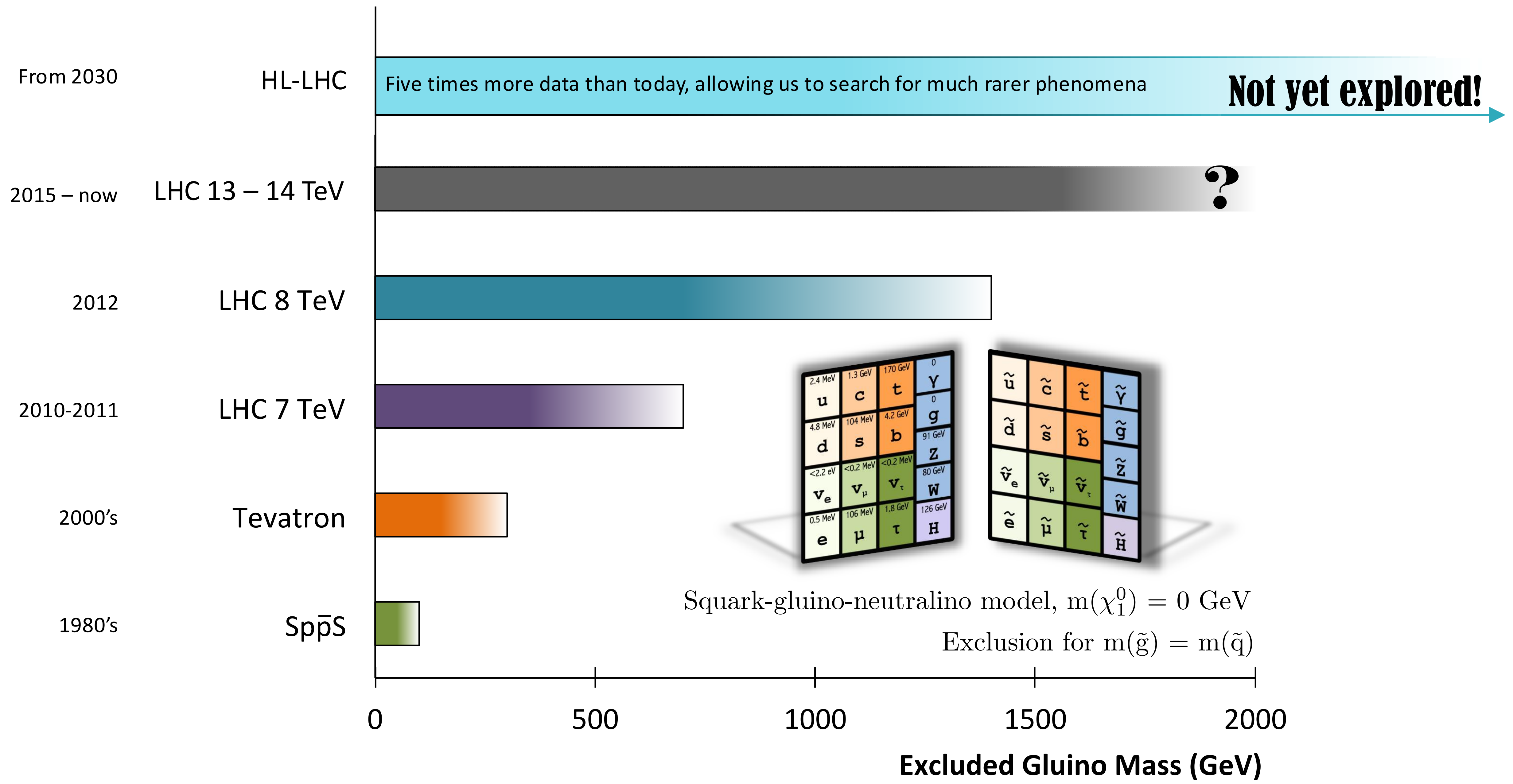
Reminder: $\sigma = \frac{\#\ \text{events}}{L}$

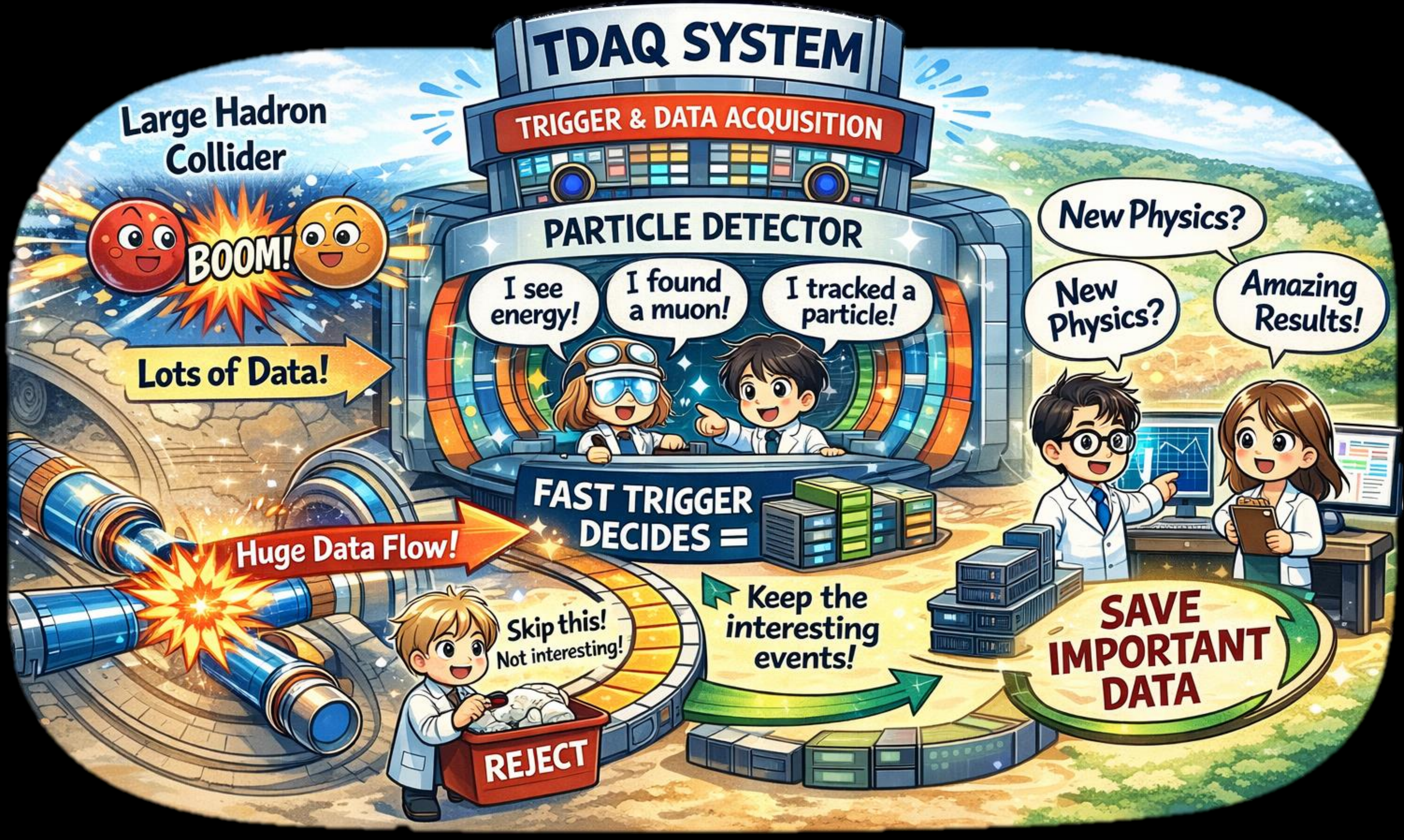
Physics objectives at HL-LHC: Precision & Exploration



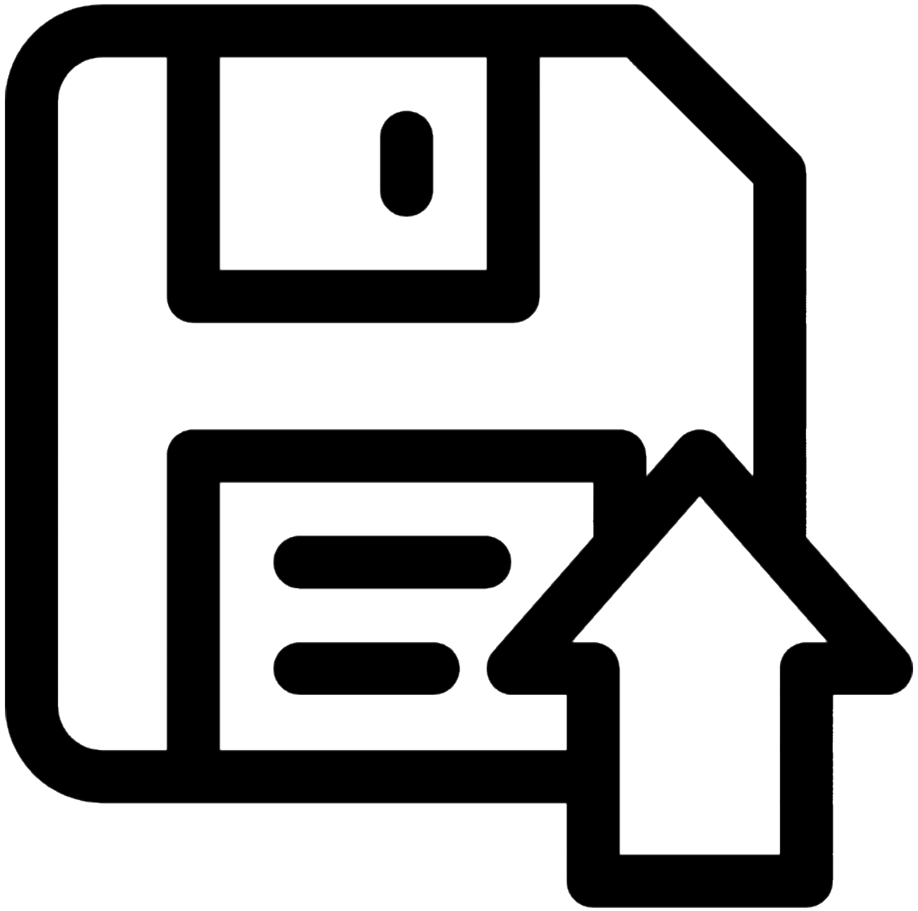
Key aspects of Standard Model still to be fully explored

New physics


From 2030
HL-LHC
Five times more data than today, allowing us to search for much rarer phenomena
Not yet explored!
2015 – now
LHC 13 – 14 TeV
?
2012
LHC 8 TeV
2010-2011
LHC 7 TeV
2000's
Tevatron
1980's
Sp$\bar{p}$S
Squark-gluino-neutralino model, $m(\chi_1^0) = 0$ GeV
Exclusion for $m(\tilde{g}) = m(\tilde{q})$
0
500
1000
1500
2000
Excluded Gluino Mass (GeV)



The TDAQ system of a collider experiment, according to ChatGPT



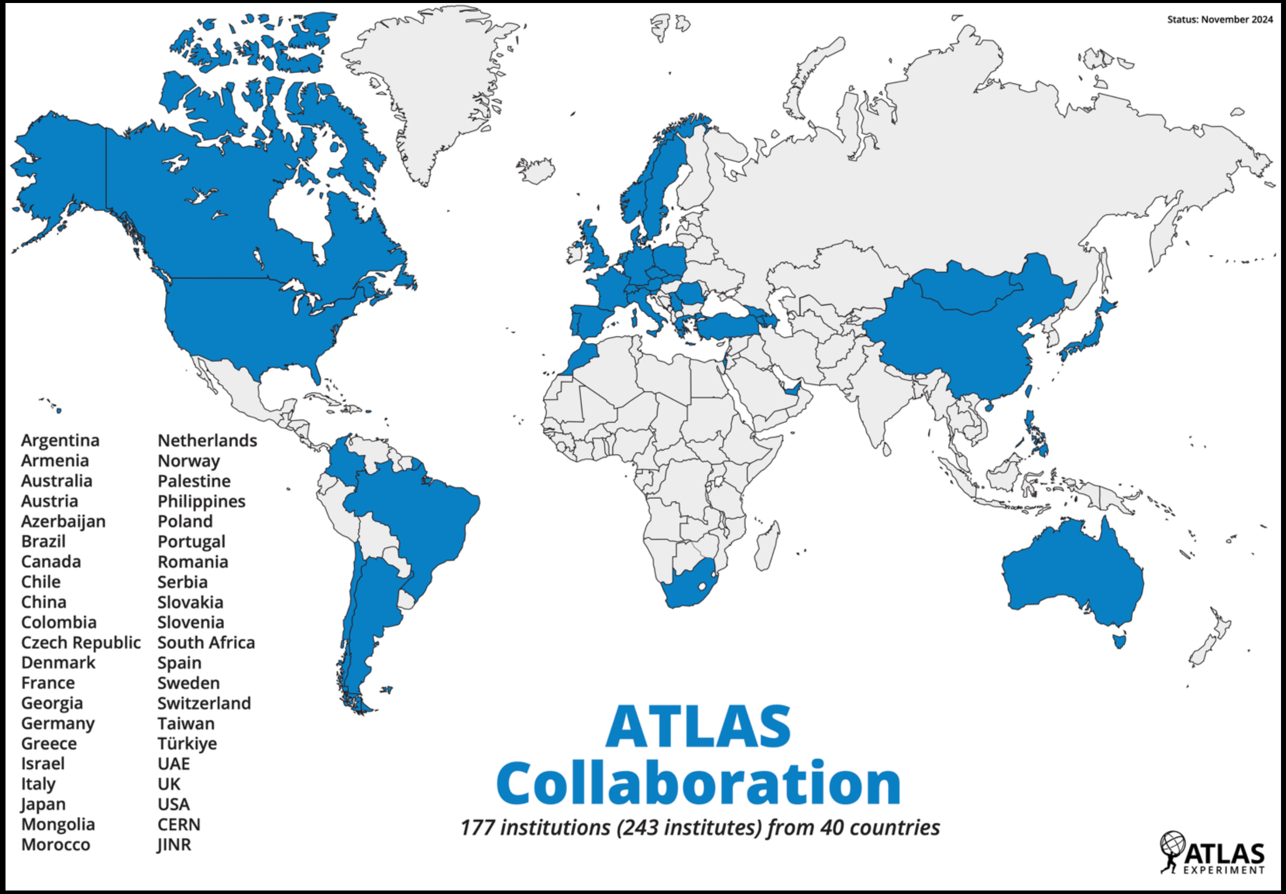
3000
Scientific authors

40
Countries

177
Institutions

1200
Doctoral students

The ATLAS Collaboration



https://atlas.cern/Discover/Collaboration
https://cds.cern.ch/record/2654110

3000

Scientific authors

40

Countries

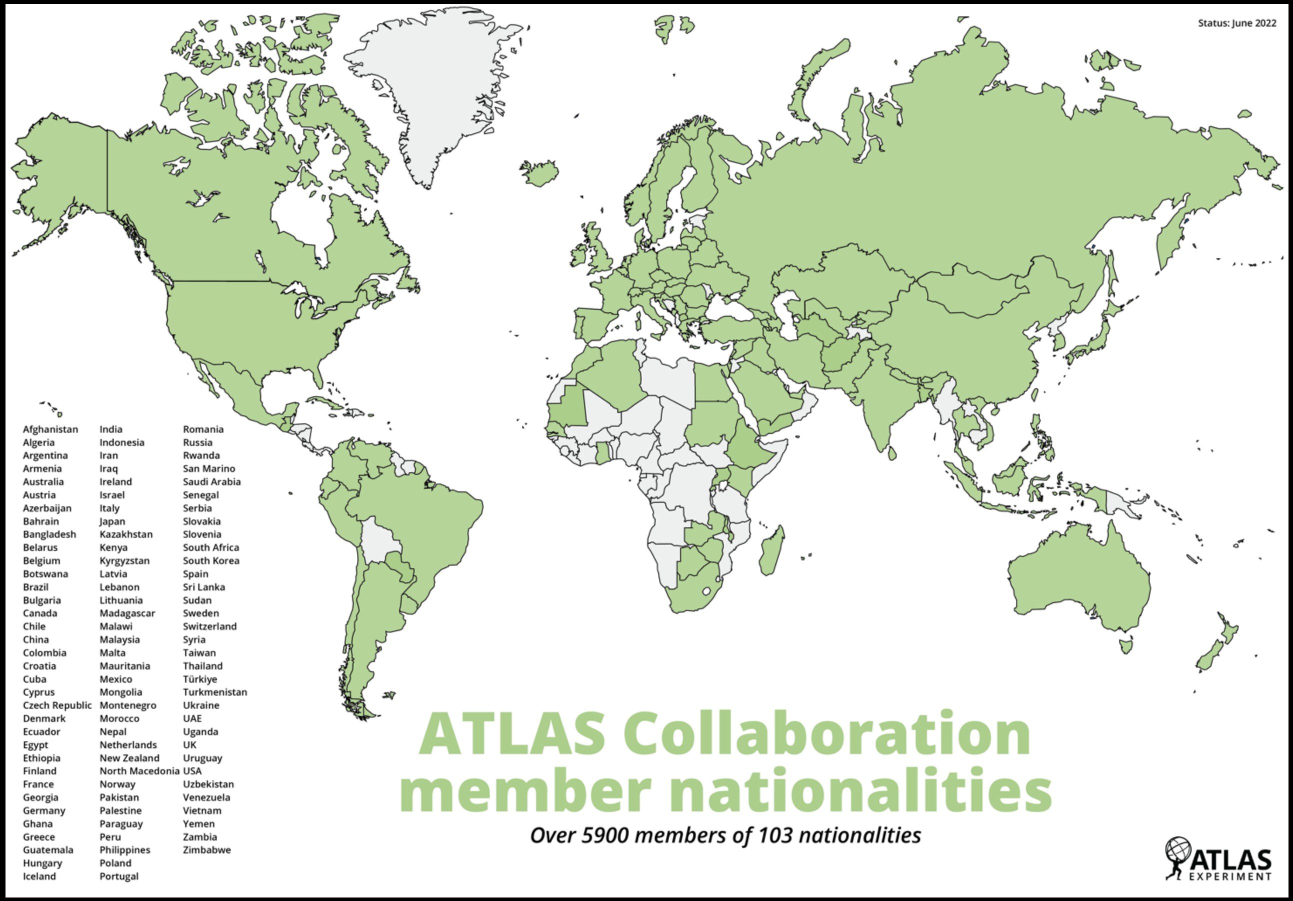
177

Institutions

1200

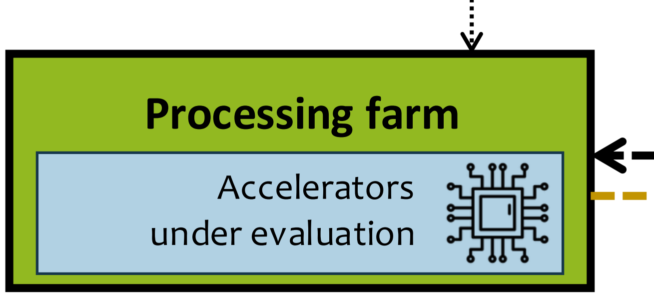
Doctoral students

The ATLAS Collaboration



https://atlas.cern/Discover/Collaboration
https://cds.cern.ch/record/2654110

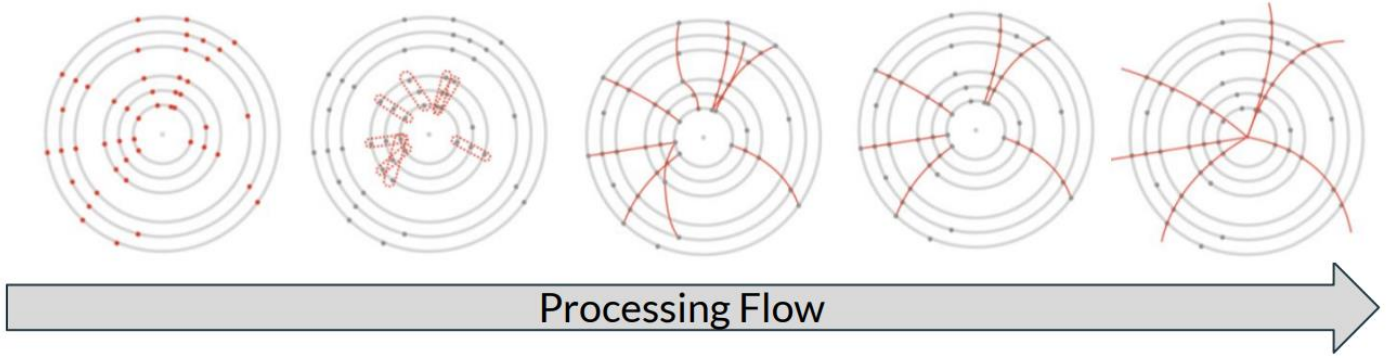
Event Filter (Tracking)



- Track reconstution: **single most computationally intensive algorithm**
  - Enormous challenge in HL-LHC: 5000 charged particles and 100,000 hits per event!
- Originally a hardware system with custom boards as co-processor to the CPU was considered
  - ATLAS decided to go with commodity solutions instead
- **Opportunities opened up for explorations of how to use FPGAs and GPUs as accelerators**
- ML algorithms for the various tracking steps under evaluation

CPU
GPU
FPGA

Mix of ML and classical algorithms



ACTS-based "Fast Tracking" Clustering + Spacepoints, Seeding, CKF

ACTS/Traccc Clustering + Spacepoints | Seeding | (C)KF

FPGA Clustering + Spacepoints | Seeding | (C)KF

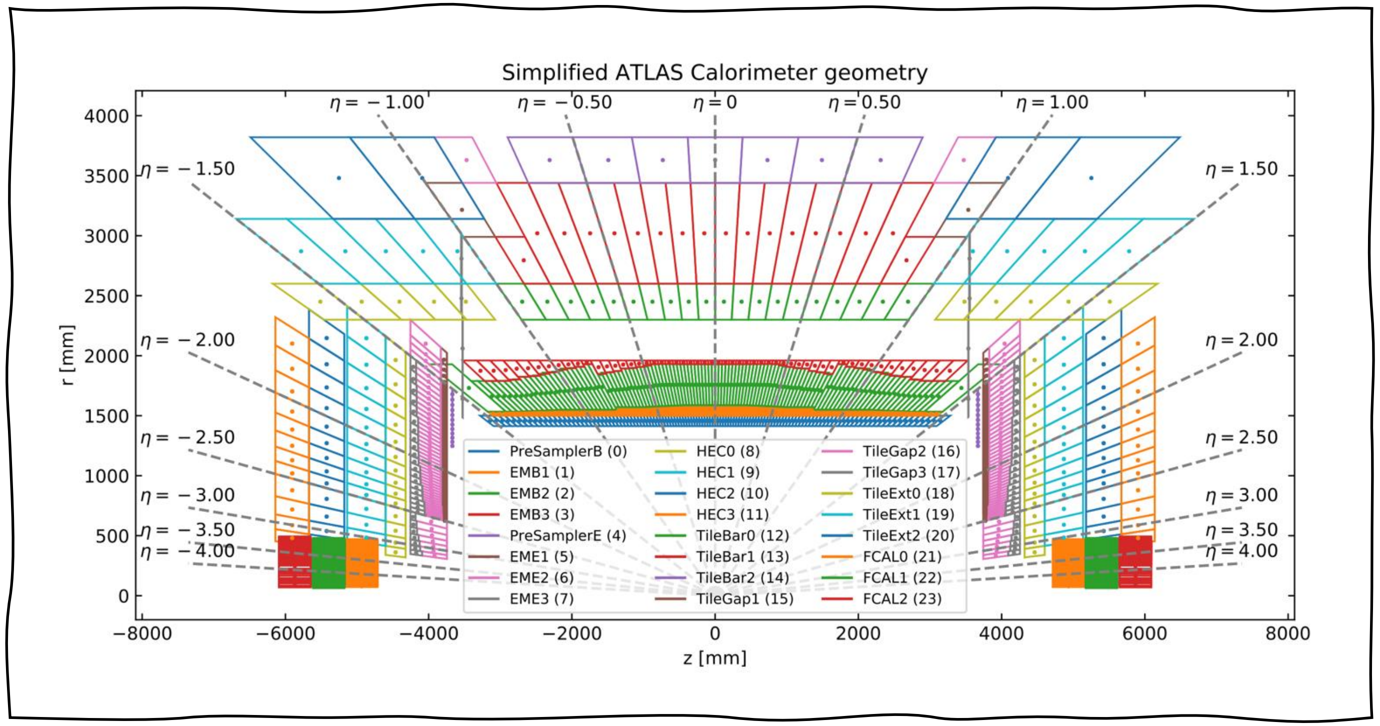
CKF: Combinatorial Kalman Filter

Next more challenging algorithm: *calorimeter clustering* – see next!

Calorimeter clustering

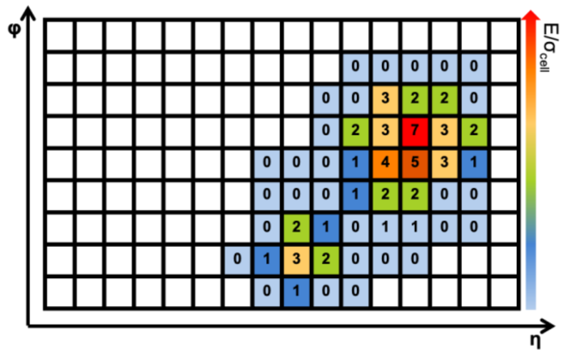
ATLAS calorimeter:

- ~200k cells
- Highly non-uniform geometry
- Clustering is executed in almost all events (L1 and HLT) arXiv:1603.02934



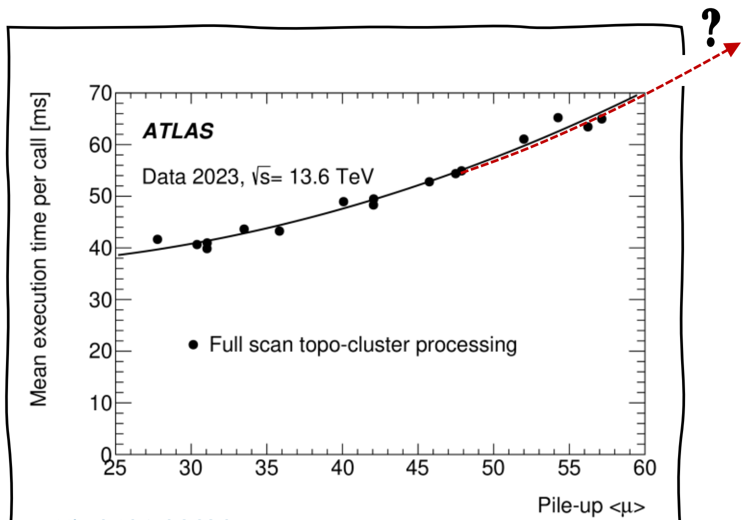
Three main steps:

- Cluster making
- Cluster splitting
- Cluster moment calculation



Challenges:

- One of the most compute-hungry algorithms
    - second only to tracking
- Default algorithm highly iterative ➜ not accelerator-friendly

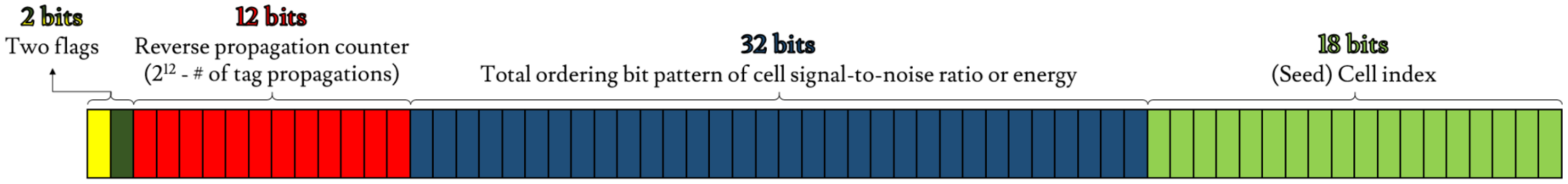


arXiv:2401.06630

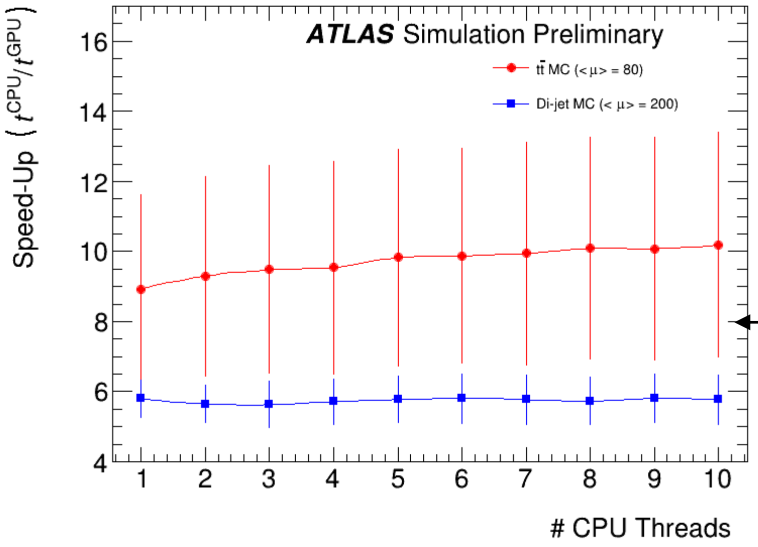
Topo-automaton clustering on GPU

A GPU alternative algorithm to standard topo-clustering

Based on the cellular automaton concept, which updates each cell in a grid by repeatedly applying simple local rules based on its neighbours, causing complex patterns to emerge over time.



Resource utilisation comparison



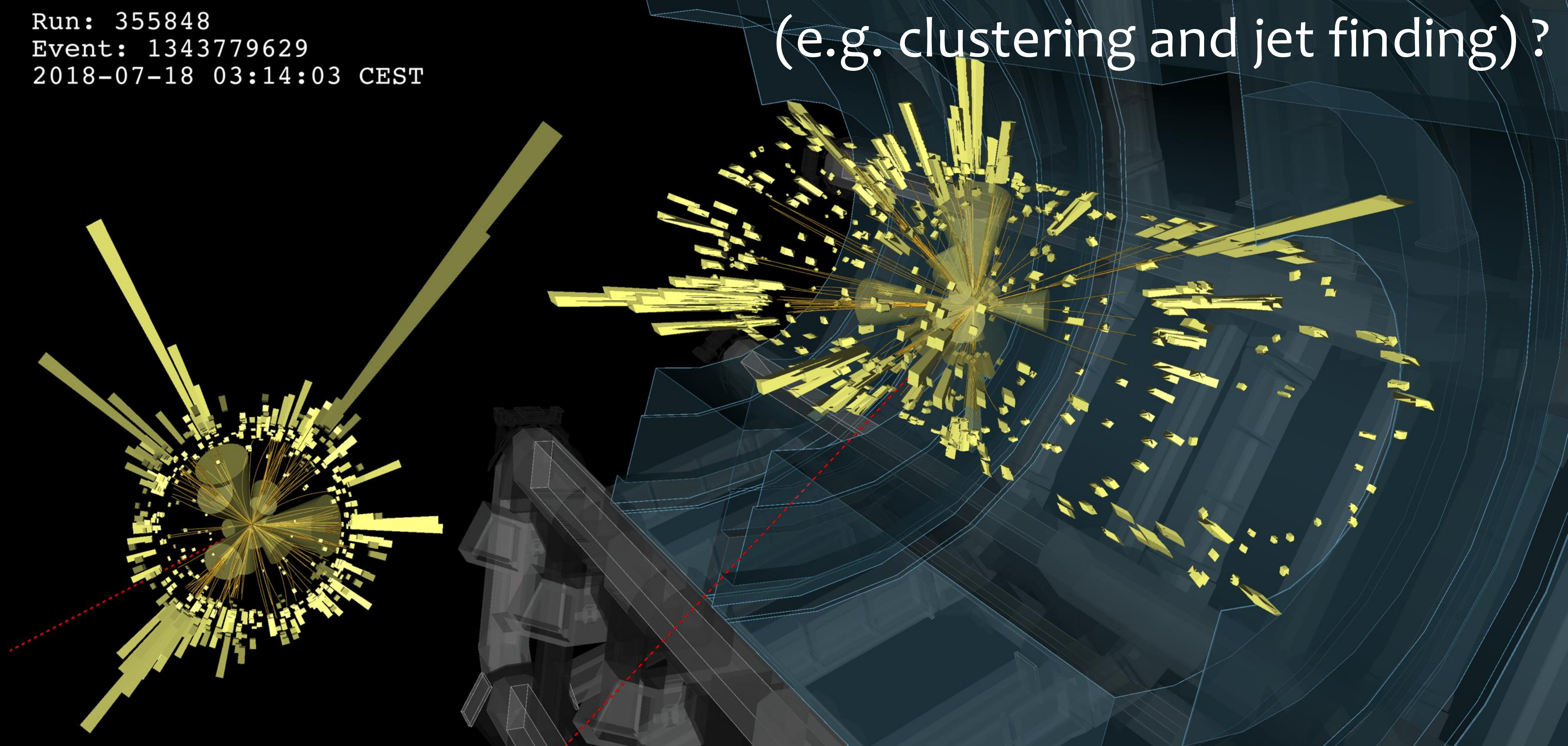
Xeon E5-2695 v4 CPU vs Tesla P100 GPU

- Algorithm fully running on GPU
- Speed-up despite a significant portion of the GPU event processing time (60~70%) is spent in data conversions
- First step (cluster making) also implemented on FPGA accelerator; resource utilisation and speed optimisations in progress

With involvement from hepia – Haute école du paysage, d'ingénierie et d'architecture de Genève

Beautiful calorimeter images

Can ML replace / accelerate calorimeter reconstruction (e.g. clustering and jet finding) ?

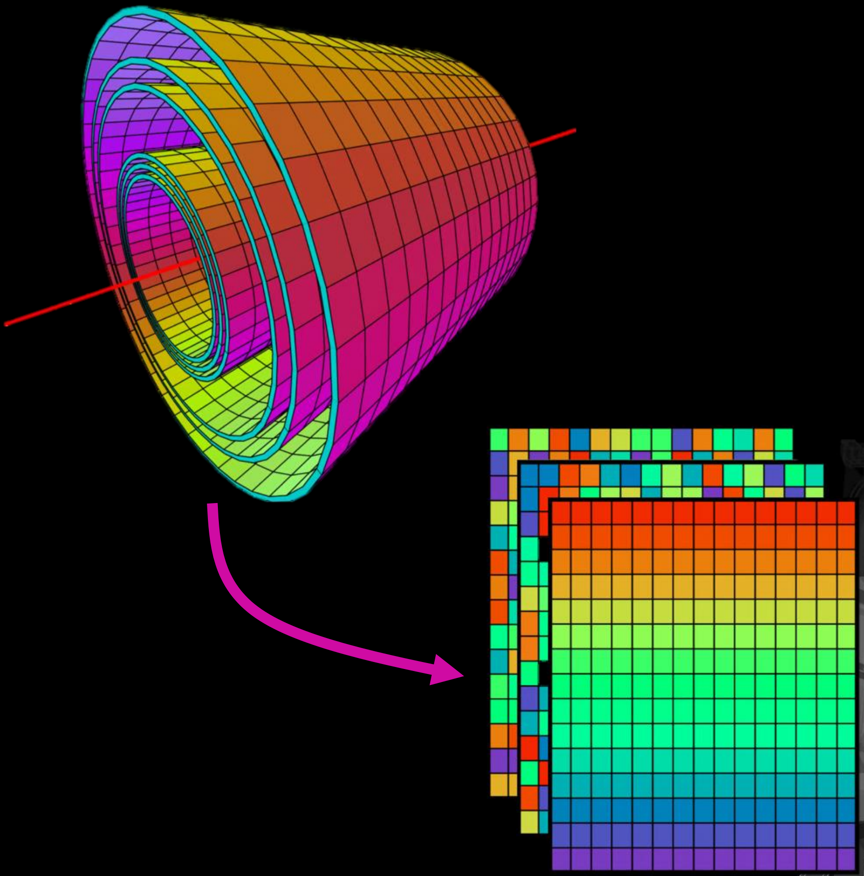


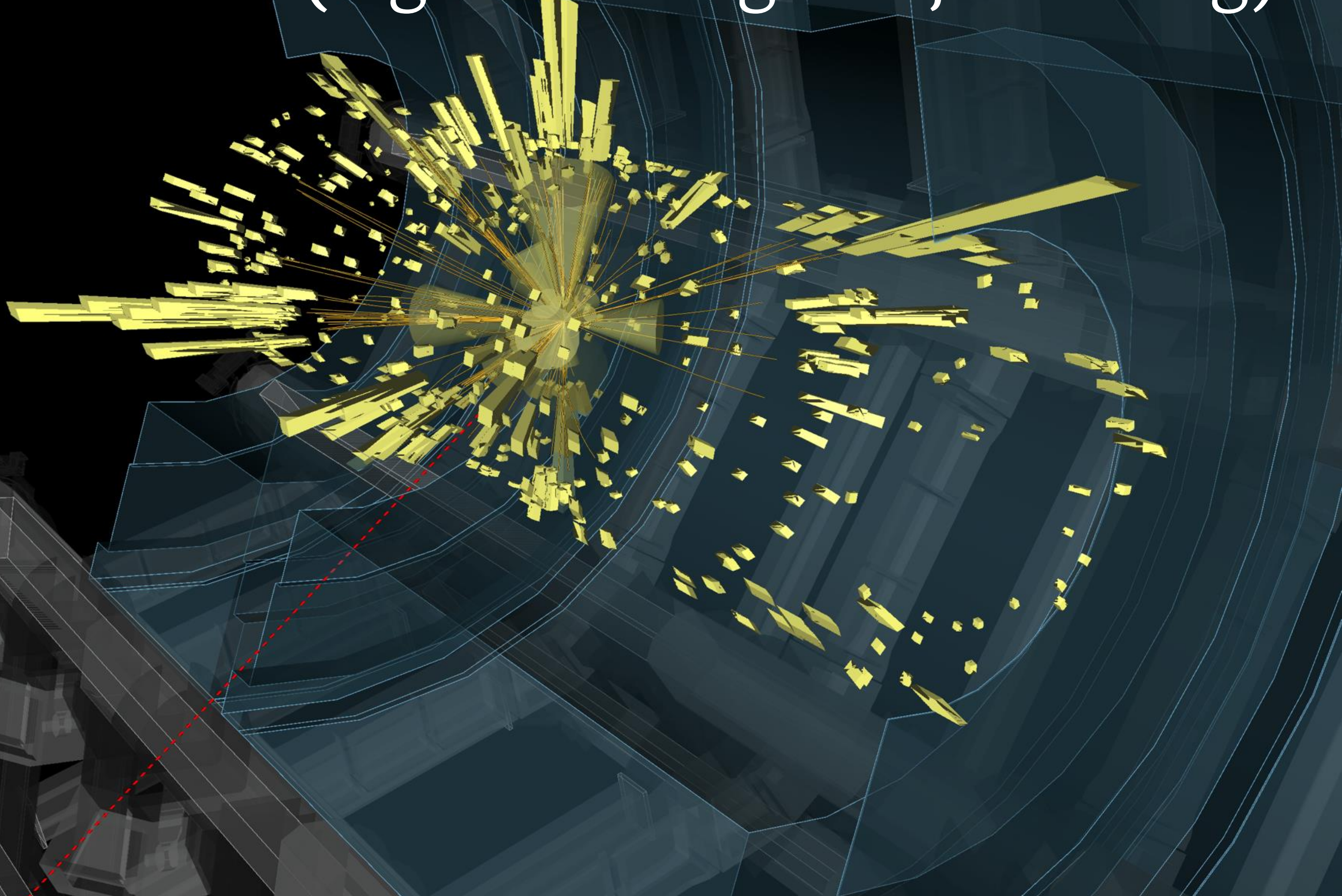
Beautiful calorimeter images

Can ML replace / accelerate calorimeter reconstruction (e.g. clustering and jet finding) ?

ATLAS EXPERIMENT

Run: 355848
Event: 1343779629
2018-07-18 03:14:03 CEST





Real-time “jet” finding using object detection

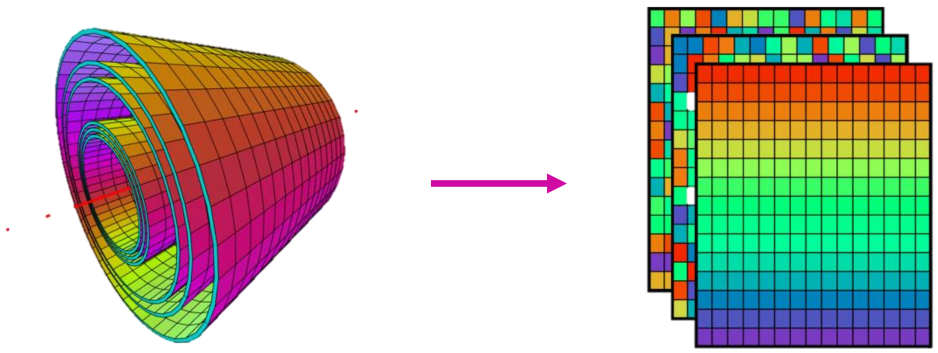
A fast alternative to a
traditional sequence of algorithms

Idea:

- Treat the calorimeter as a **2D image.**
- Use **object-detection ML models** to identify jets directly from cell energies.

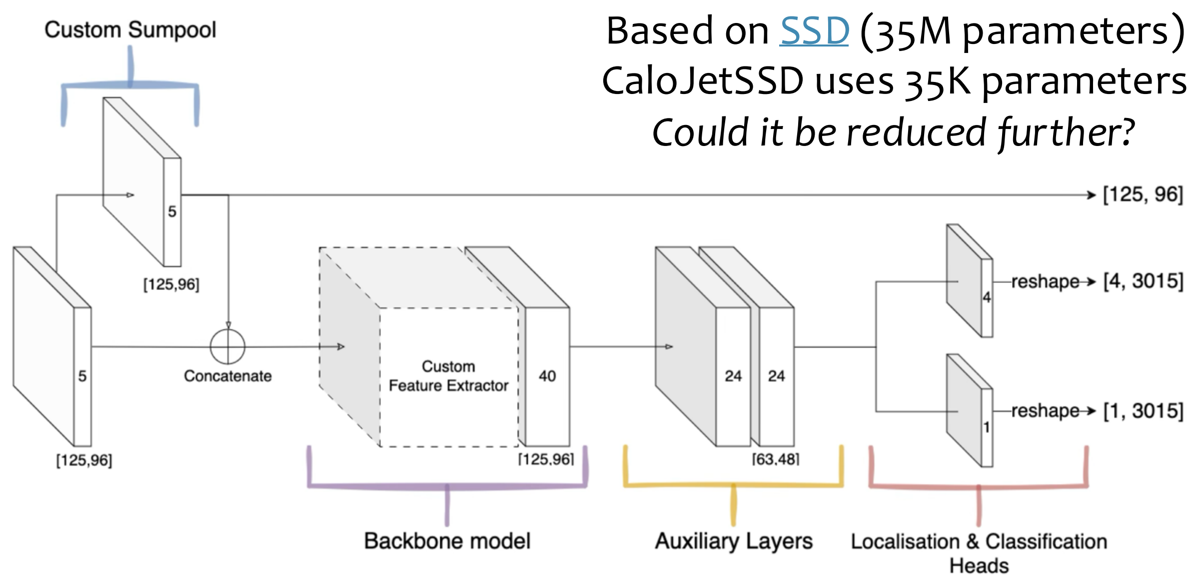
Why?

- Avoids the full chain (cells ➜ clusters ➜ calibrations ➜ anti-kt)
- Enables **much faster** jet reconstruction at trigger level



CaloJetSSD

Based on SSD (35M parameters)
CaloJetSSD uses 35K parameters
Could it be reduced further?



ATL-DAQ-PUB-2024-003, talk

Real-time "jet" finding using object detection - results

A fast alternative to a traditional sequence of algorithms

Results

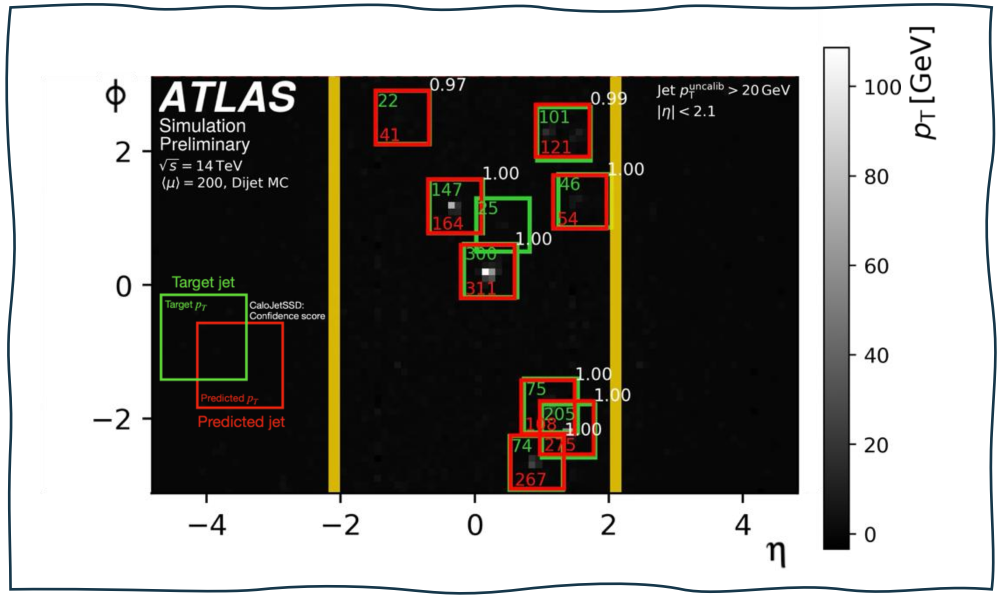
- ~×10 speed-up vs. standard jet building; even greater at high pile-up
  - Image creation runs on CPU; ML inference runs on GPU; speed-up includes both
- Robust (promising) performance, even with high pile-up
- Uses **35k parameters** (much smaller than typical vision models)

→ Promising direction for **HL-LHC real-time jet triggers** using lightweight ML

With involvement from

hepia

Haute école du paysage, d'ingénierie
et d'architecture de Genève



Could such an algorithm run on FPGAs?

ATL-DAQ-PUB-2024-003, talk