

CONCURSO DE ESTANCIAS DE VERANO
EXPERIMENTALES 2026

◆ ◆

TALLER DE ANÁLISIS DE DATOS

Dra. Haydee Hernández Arellano

hydhernandez@gmail.com

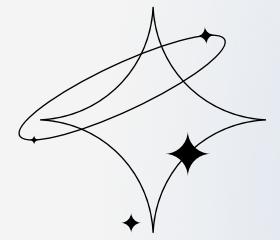
División de Partículas y Campos - Sociedad Mexicana de Física

Enero 2026

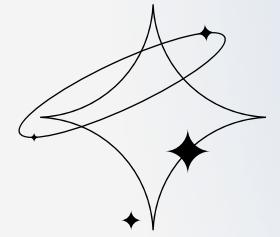


ÍNDICE

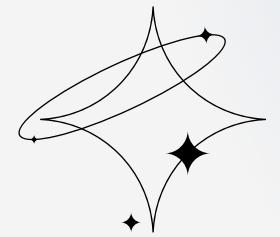
Introducción



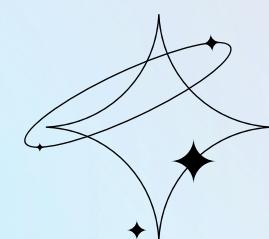
Formato de los datos



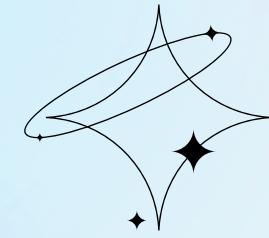
Ejercicio 1: Graficar e interpretar



Ejercicio 2: Selección básica de datos

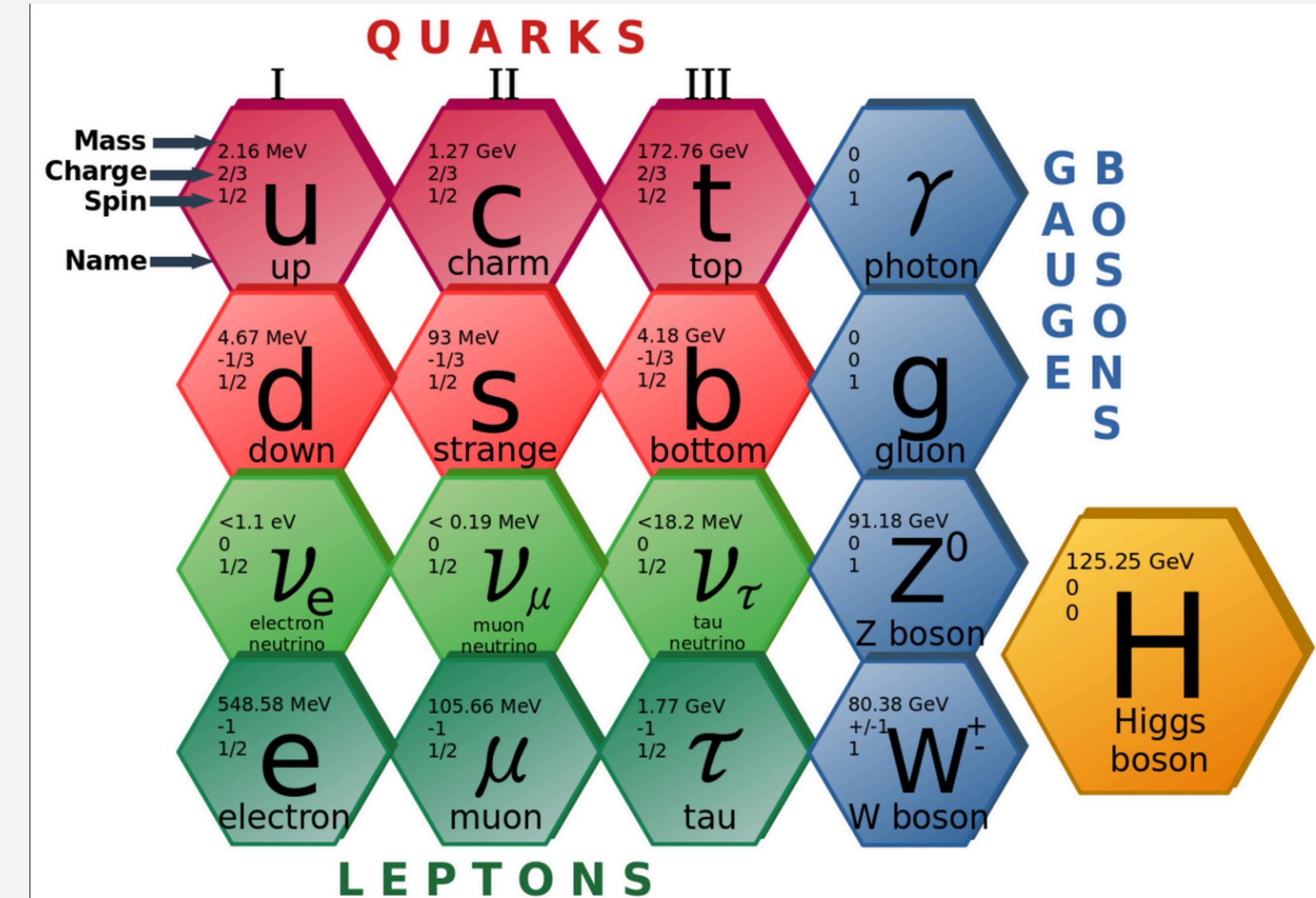


Ejercicio 3: Análisis de datos abiertos del CMS con UpROOT



Modelo Estándar de las Partículas Elementales

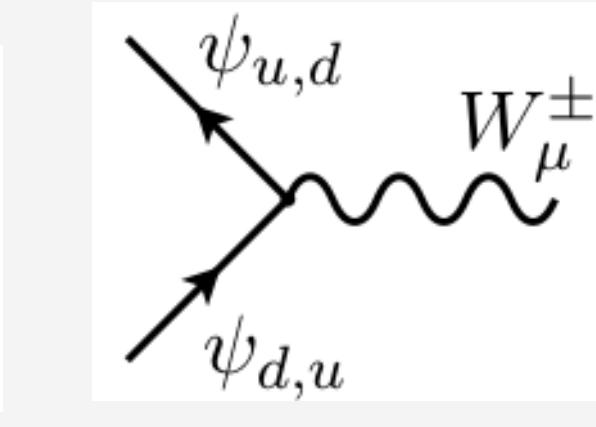
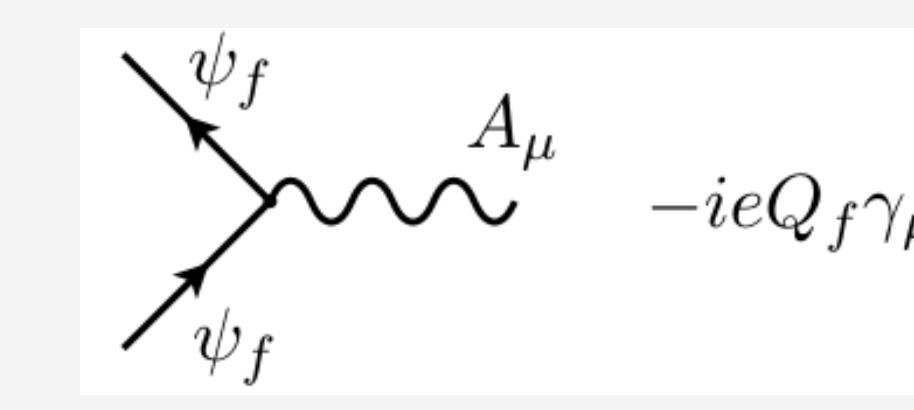
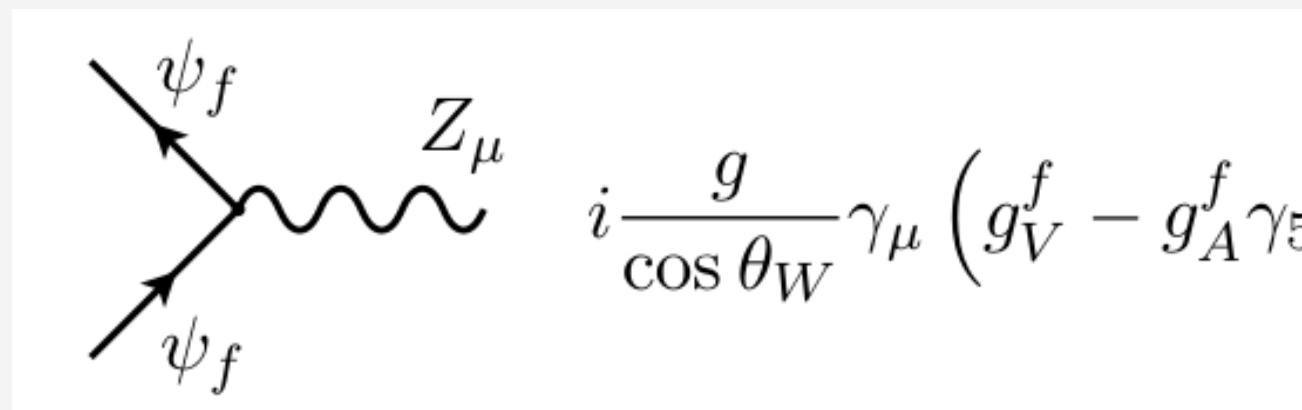
- Teoría Cuántica de Campos:
Fuerza Electromagnética + Débil
+ Fuerte
- Grupo de simetría:
 $SU(3)_C \times SU(2)_L \times U(1)_Y$
- Auto-consistente: **libre de anomalías**. Consistente en el esquema perturbativo a energías \sim TeV.
- **Higgs** explica la masa de los bosones de norma, $m_H=125$ GeV.



Interacciones entre partículas

$$\mathcal{L}_{\text{Fermion}} = \sum_{\text{quarks}} i\bar{q}\gamma^\mu D_\mu q + \sum_{\psi_L} i\bar{\psi}_L \gamma^\mu D_\mu \psi_L + \sum_{\psi_R} i\bar{\psi}_R \gamma^\mu D_\mu \psi_R$$

$$\begin{aligned} D_\mu \psi_L &= \left[\partial_\mu - i \frac{g}{\sqrt{2}} (\tau^+ W_\mu^+ + \tau^- W_\mu^-) - i \frac{g}{2} \tau_3 W_\mu^3 + ig' Y B_\mu \right] \psi_L \\ &= \left[\partial_\mu - i \frac{g}{\sqrt{2}} (\tau^+ W_\mu^+ + \tau^- W_\mu^-) + ie Q A_\mu - i \frac{g}{\cos \theta_W} \left(\frac{\tau_3}{2} - Q \sin^2 \theta_W \right) Z_\mu \right] \psi_L \end{aligned} \quad (\text{D.14})$$



$$i \frac{g}{\sqrt{2}} \gamma_\mu \frac{1 - \gamma_5}{2}$$

Particle Data Group: <https://pdg.lbl.gov/>

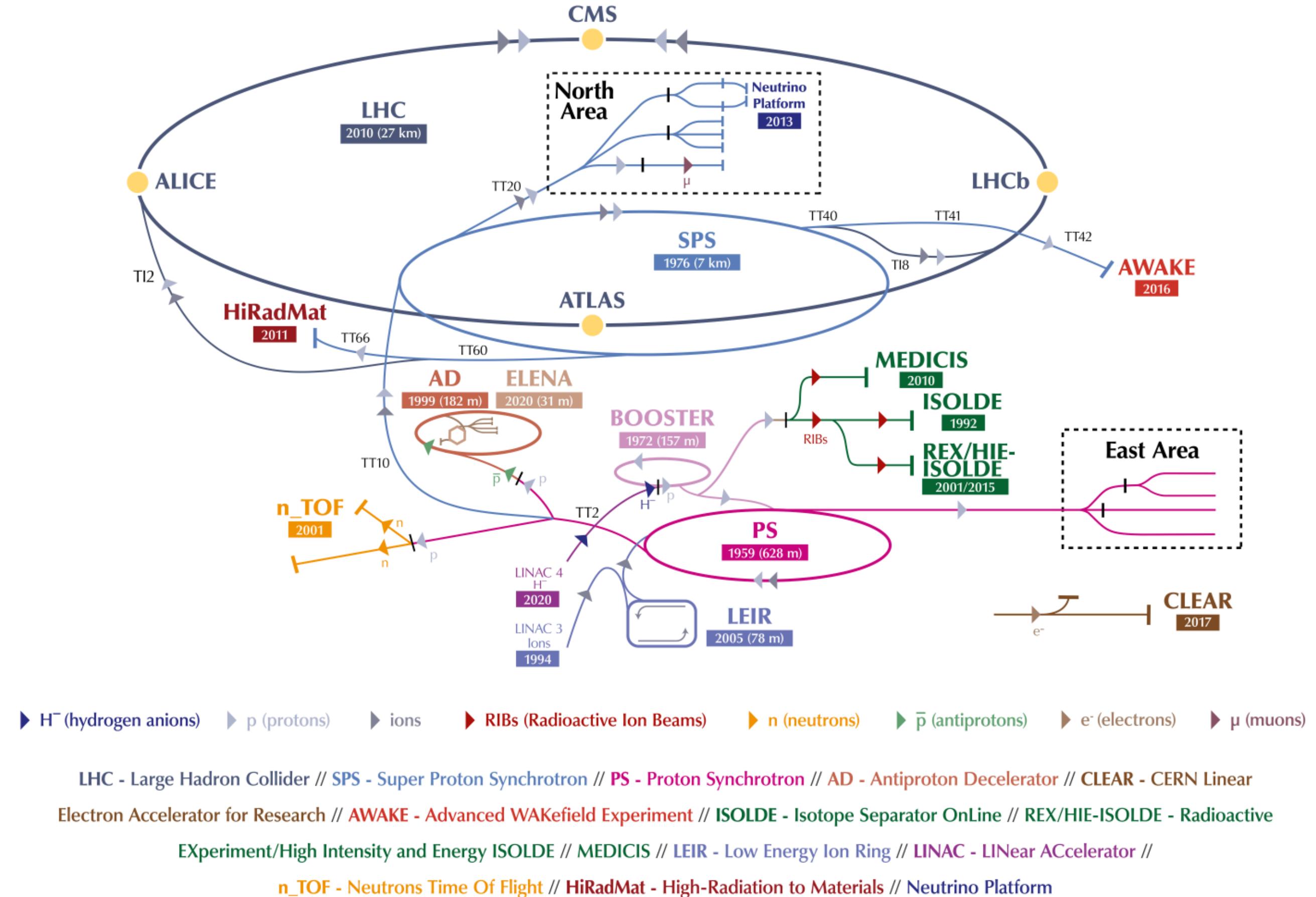
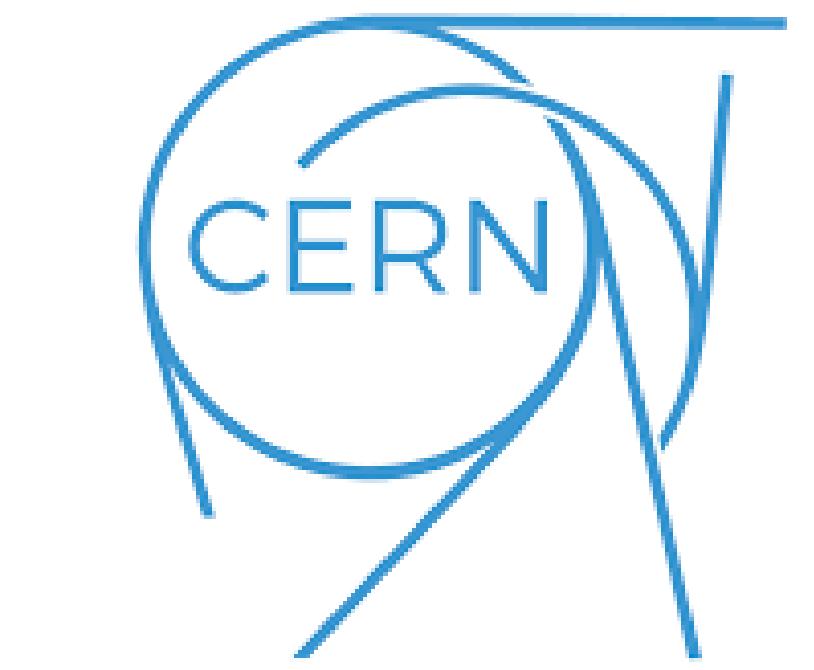
Z DECAY MODES	Fraction (Γ_i/Γ)	Scale factor/ Confidence level	p (MeV/c)
$e^+ e^-$	(3.3632 ± 0.0042) %	45594	
$\mu^+ \mu^-$	(3.3662 ± 0.0066) %	45594	
$\tau^+ \tau^-$	(3.3696 ± 0.0083) %	45559	
$\ell^+ \ell^-$	[c] (3.3658 ± 0.0023) %	—	
$\ell^+ \ell^- \ell^+ \ell^-$	[i] (4.55 ± 0.17) $\times 10^{-6}$	45594	
invisible	(20.000 ± 0.055) %	—	
hadrons	(69.911 ± 0.056) %	—	

W ⁺ DECAY MODES	Fraction (Γ_i/Γ)	Confidence level	(MeV/c)
$\ell^+ \nu$	[c] (10.86 ± 0.09) %	—	
$e^+ \nu$	(10.71 ± 0.16) %	40185	
$\mu^+ \nu$	(10.63 ± 0.15) %	40185	
$\tau^+ \nu$	(11.38 ± 0.21) %	40165	
hadrons	(67.41 ± 0.27) %	—	

H DECAY MODES	Fraction (Γ_i/Γ)	Confidence level	p (MeV/c)
WW^*	(25.7 ± 2.5) %	—	
ZZ^*	(2.80 ± 0.30) %	—	
$\gamma\gamma$	(2.50 ± 0.20) $\times 10^{-3}$	62600	
$b\bar{b}$	(53 ± 8) %	—	
$e^+ e^-$	< 3.0×10^{-4}	95%	62600
$\mu^+ \mu^-$	(2.6 ± 1.3) $\times 10^{-4}$	62600	
$\tau^+ \tau^-$	($6.0^{+0.8}_{-0.7}$) %	62575	

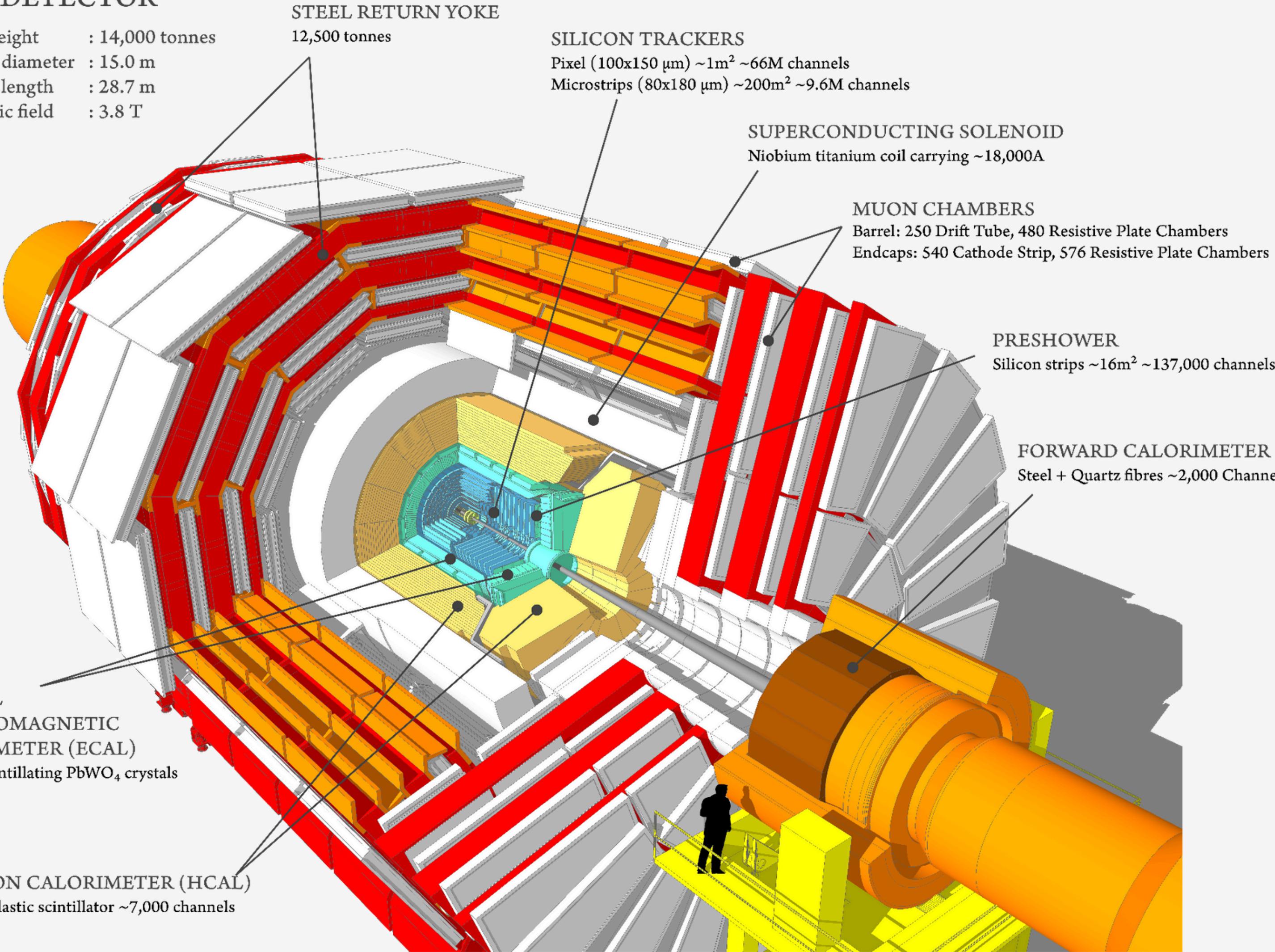
The CERN accelerator complex

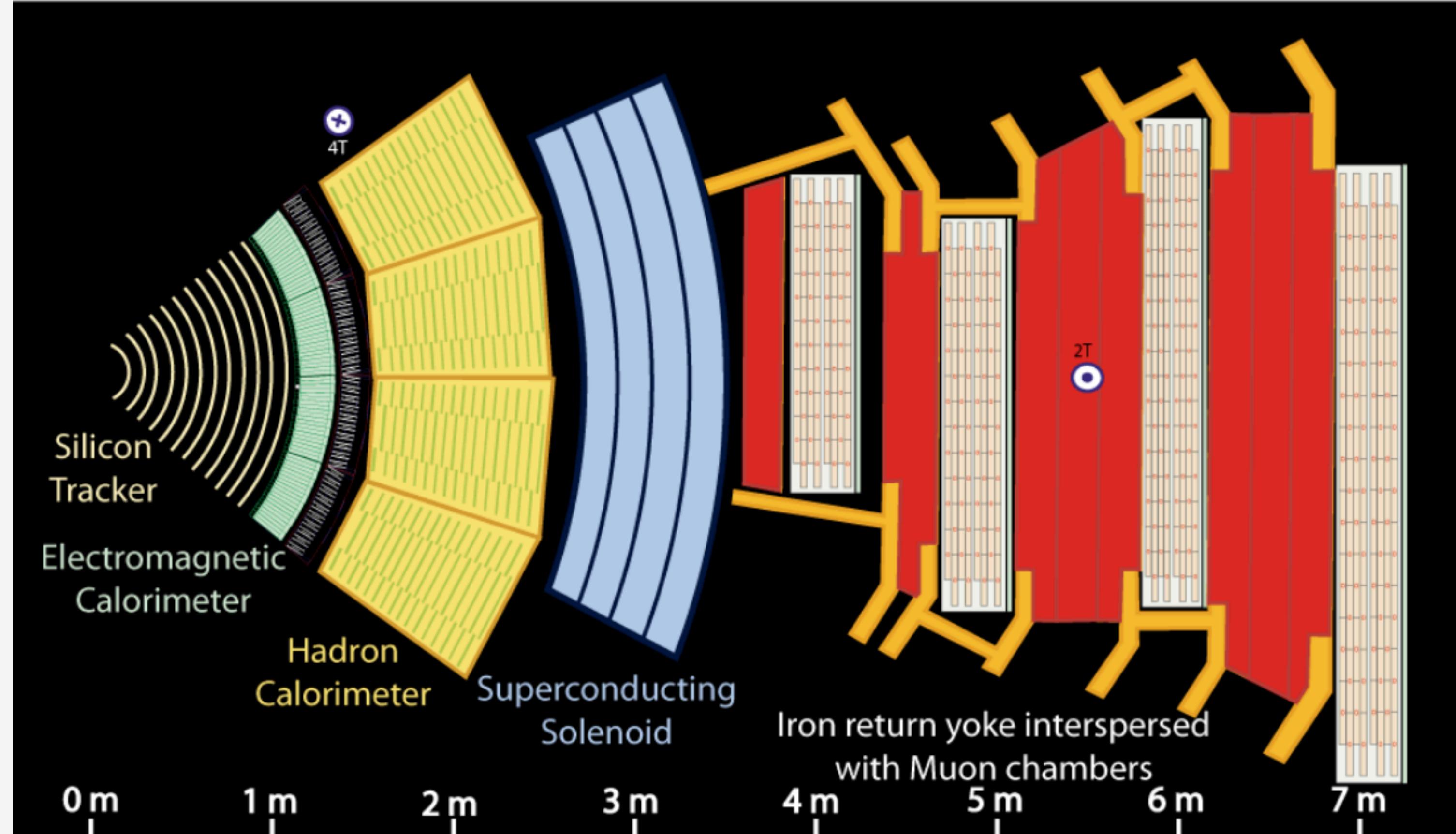
Complexe des accélérateurs du CERN



CMS DETECTOR

Total weight : 14,000 tonnes
Overall diameter : 15.0 m
Overall length : 28.7 m
Magnetic field : 3.8 T





Key:

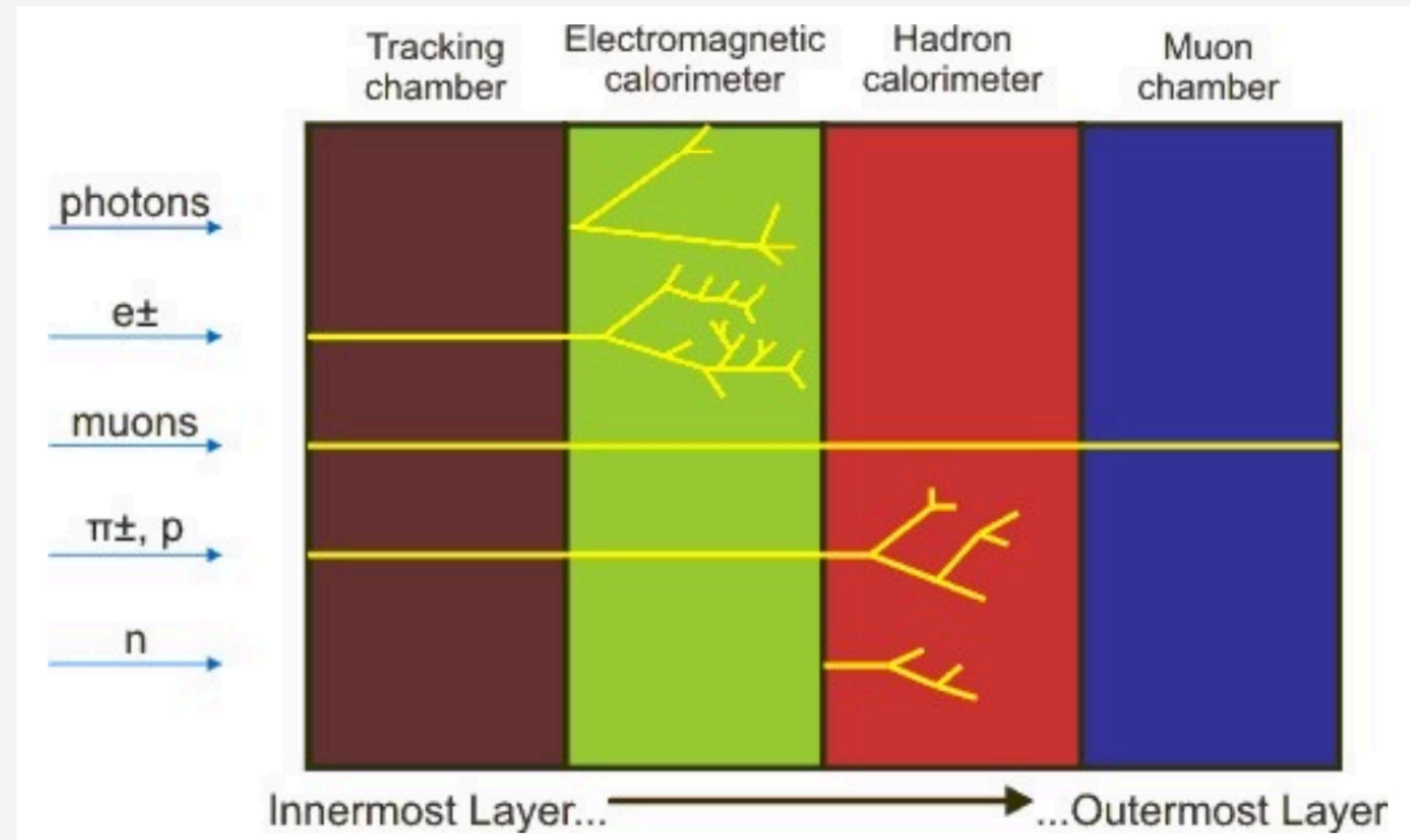
— Muon

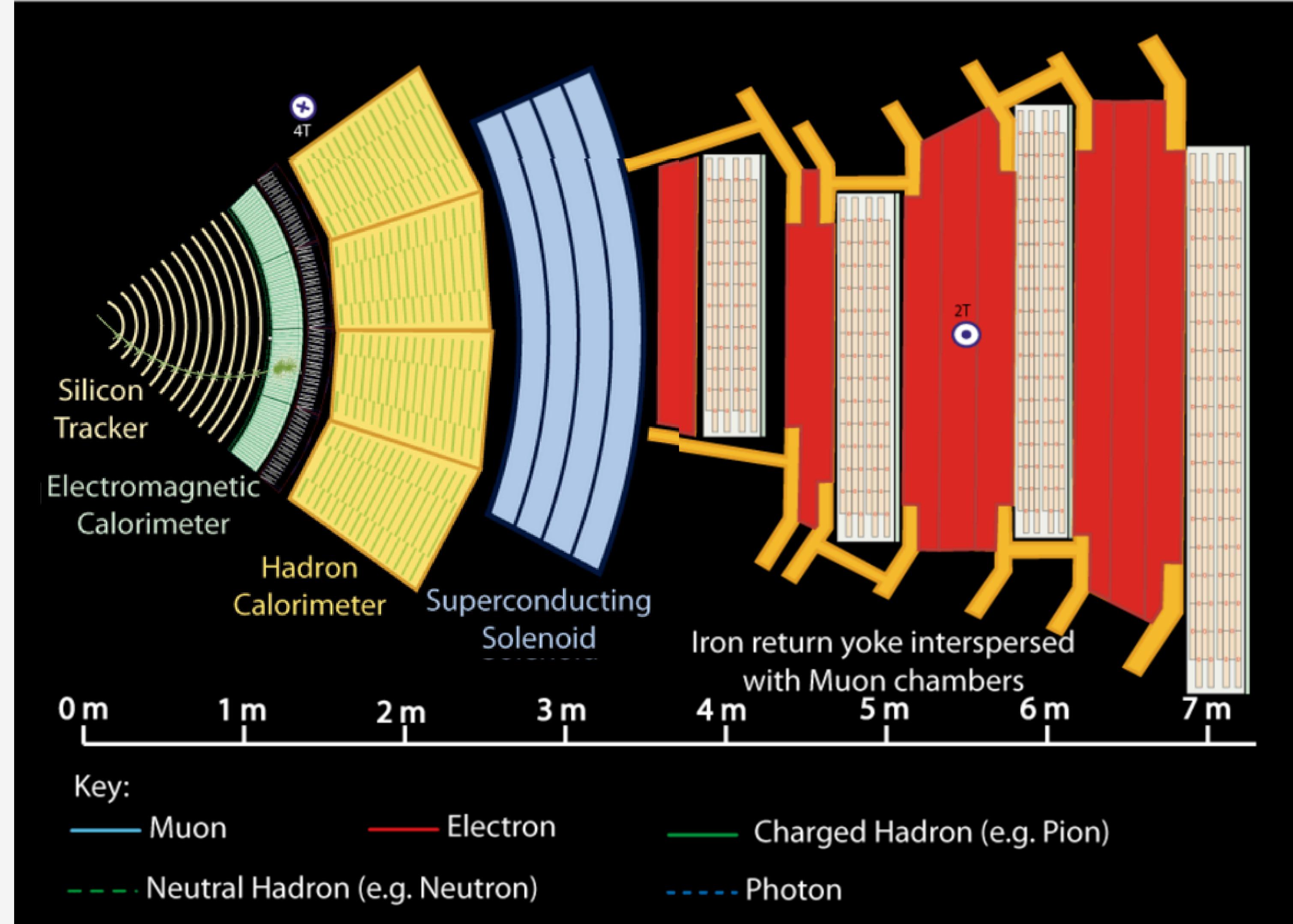
— Electron

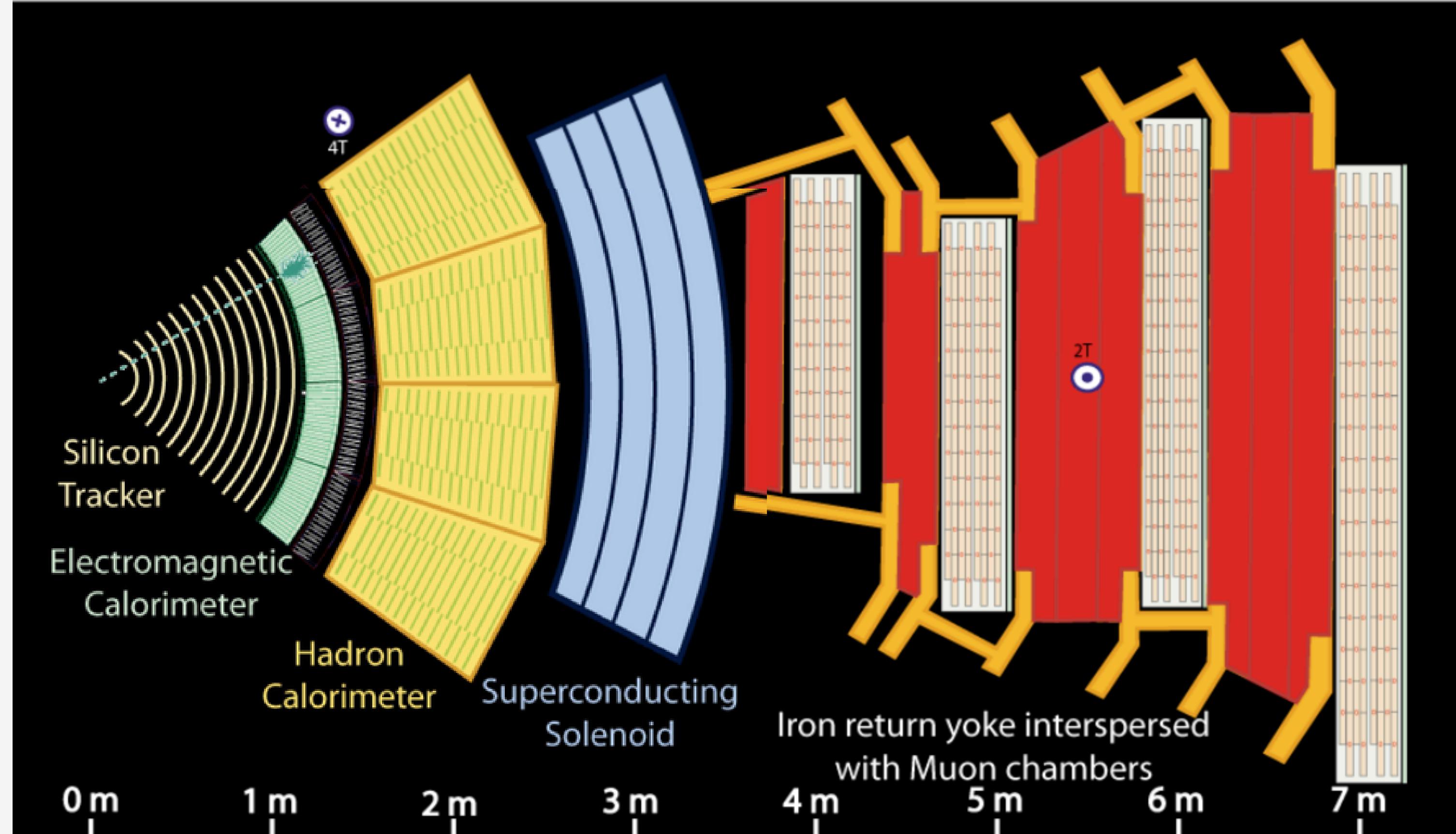
— Charged Hadron (e.g. Pion)

- - - Neutral Hadron (e.g. Neutron)

----- Photon







Key:

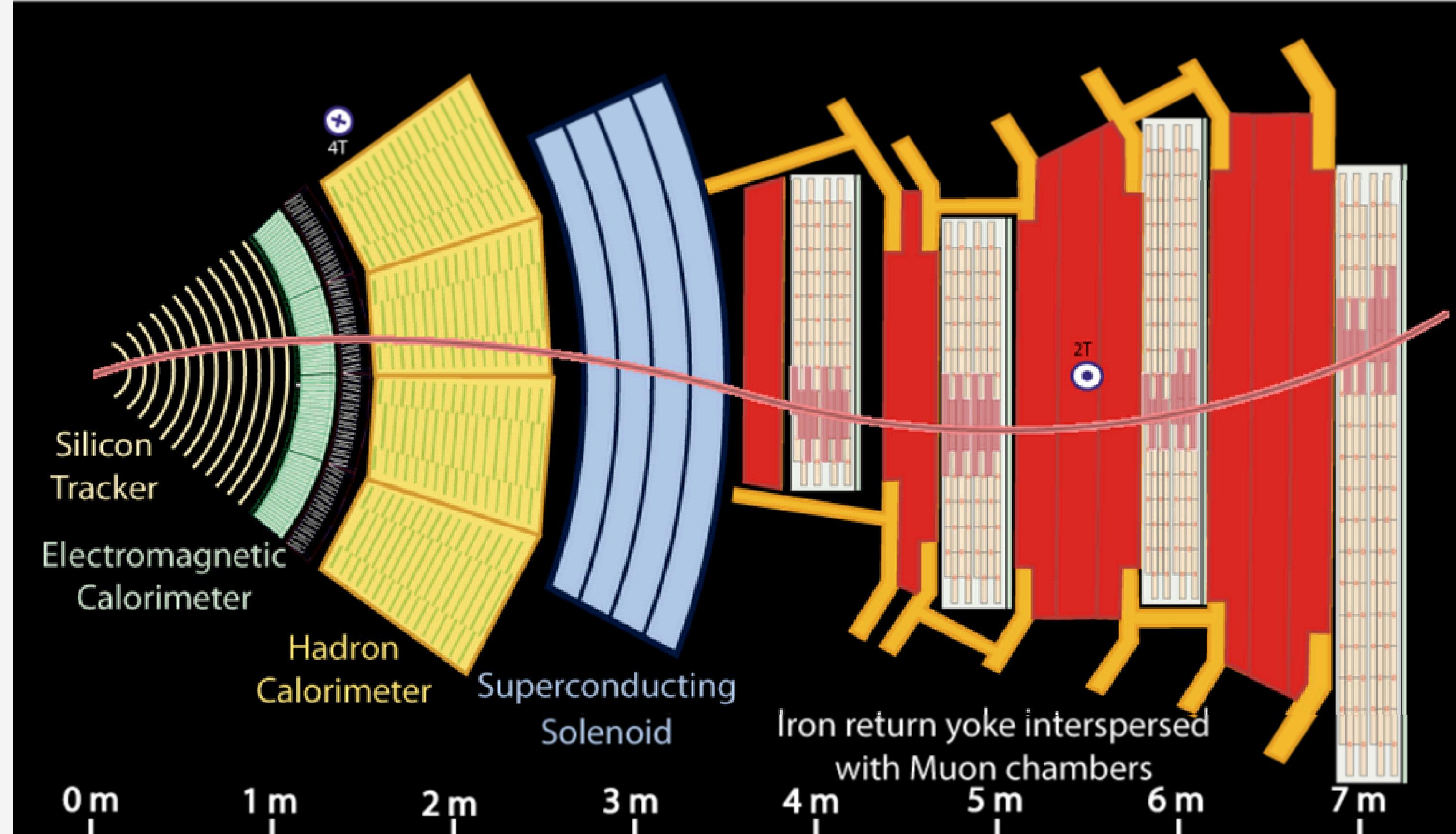
— Muon

— Electron

— Charged Hadron (e.g. Pion)

- - - Neutral Hadron (e.g. Neutron)

----- Photon



Key:

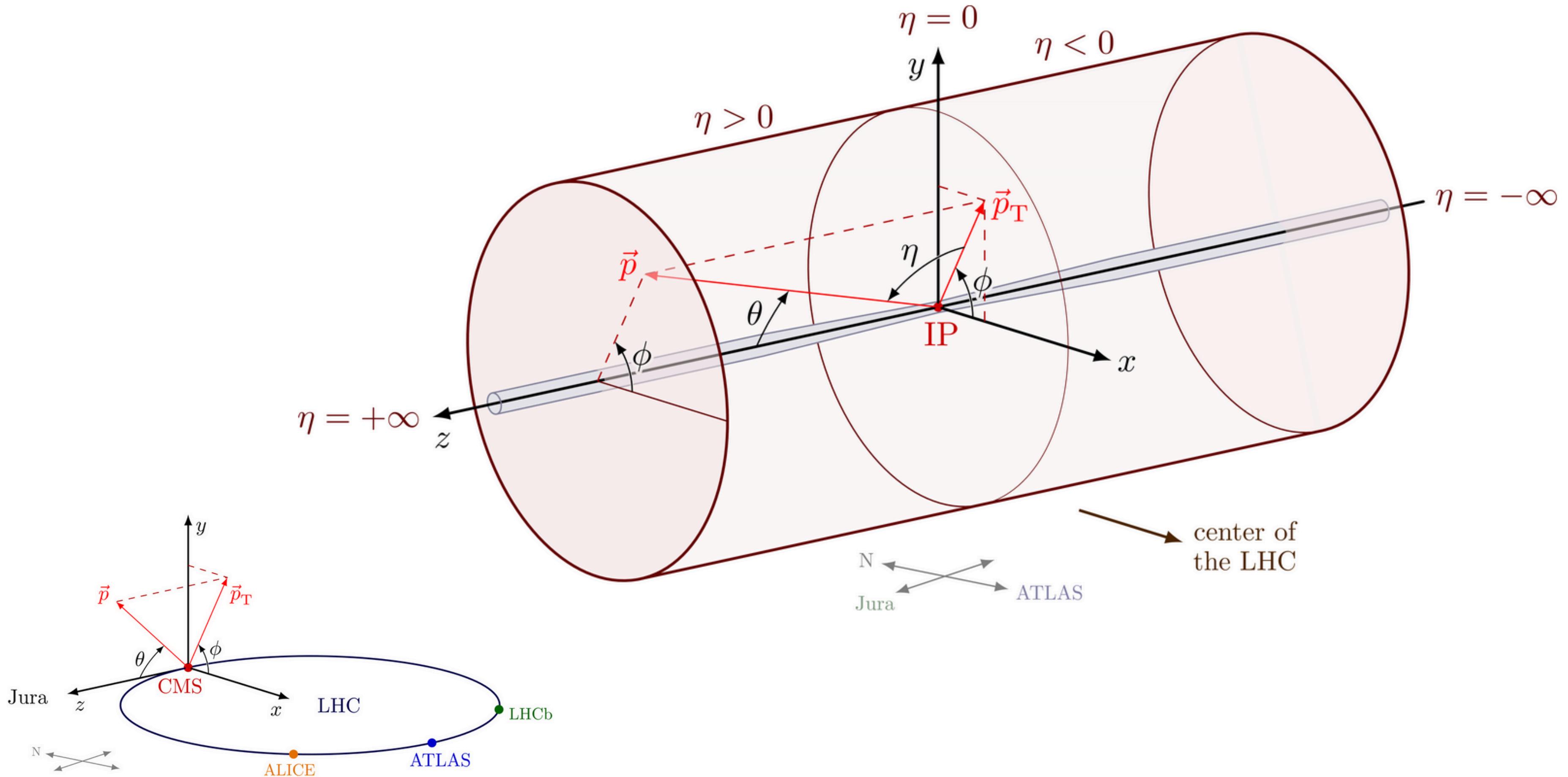
— Muon

— Electron

— Charged Hadron (e.g. Pion)

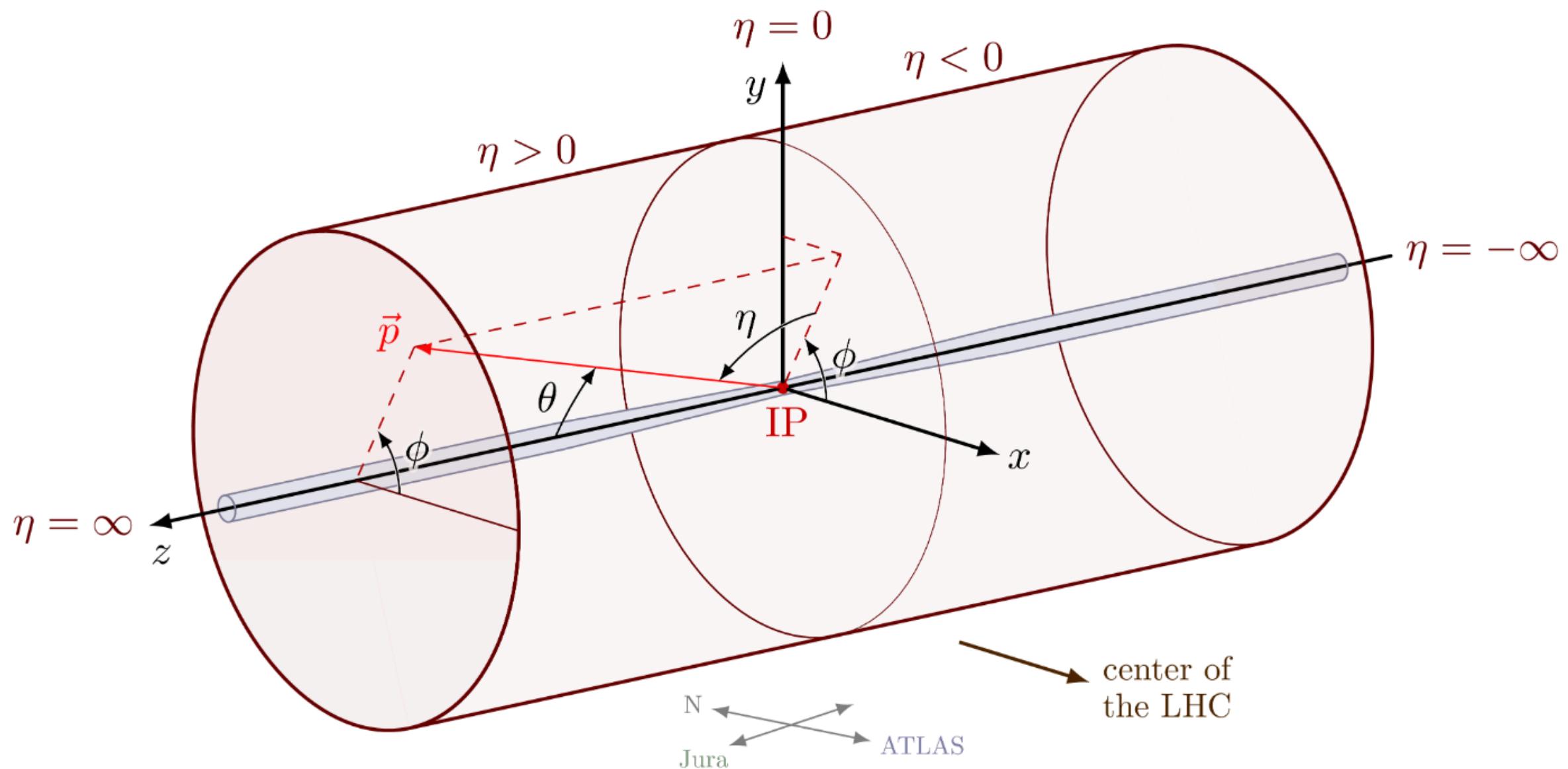
- - - Neutral Hadron (e.g. Neutron)

----- Photon



Observables

| CMS coordinate system



Momento Transverso (pt)

$$p_T = \sqrt{p_x^2 + p_y^2}$$

Pseudorapidez (eta)

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

Ángulo polar (phi)

ϕ

Rapidez y Pseudorapidez

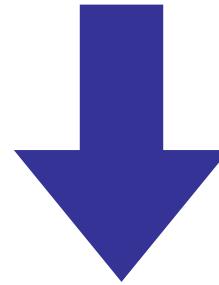
$$\begin{aligned} E &= \gamma m c^2 \\ |\mathbf{p}| &= \gamma m \beta c \end{aligned} \longrightarrow \begin{aligned} E &= m c^2 \cosh \omega \\ |\mathbf{p}| &= m c \sinh \omega \end{aligned}$$

$$\tanh \omega = \frac{\sinh \omega}{\cosh \omega} = \frac{|\mathbf{p}| c}{E}$$

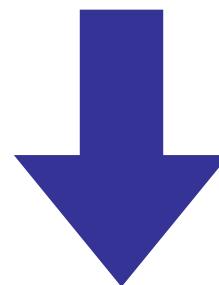
$$\omega = \arctan\left(\frac{|\mathbf{p}| c}{E}\right) = \frac{1}{2} \ln \left[\frac{E + |\mathbf{p}| c}{E - |\mathbf{p}| c} \right]$$

Rapidez y Pseudorapidez

$$\omega = \arctan\left(\frac{|\vec{p}_\perp| c}{E}\right) = \frac{1}{2} \ln\left[\frac{E + |\vec{p}| c}{E - |\vec{p}| c}\right]$$

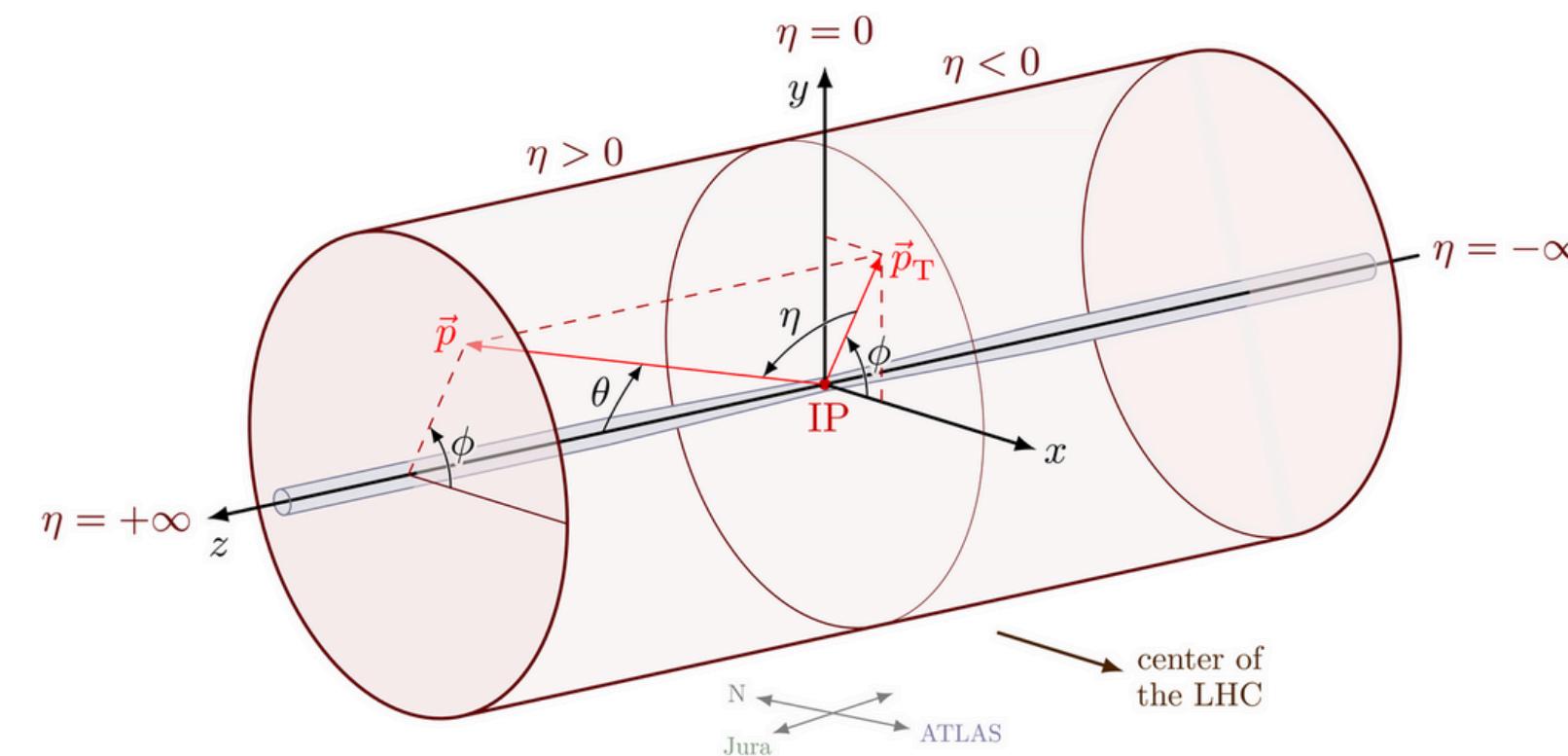
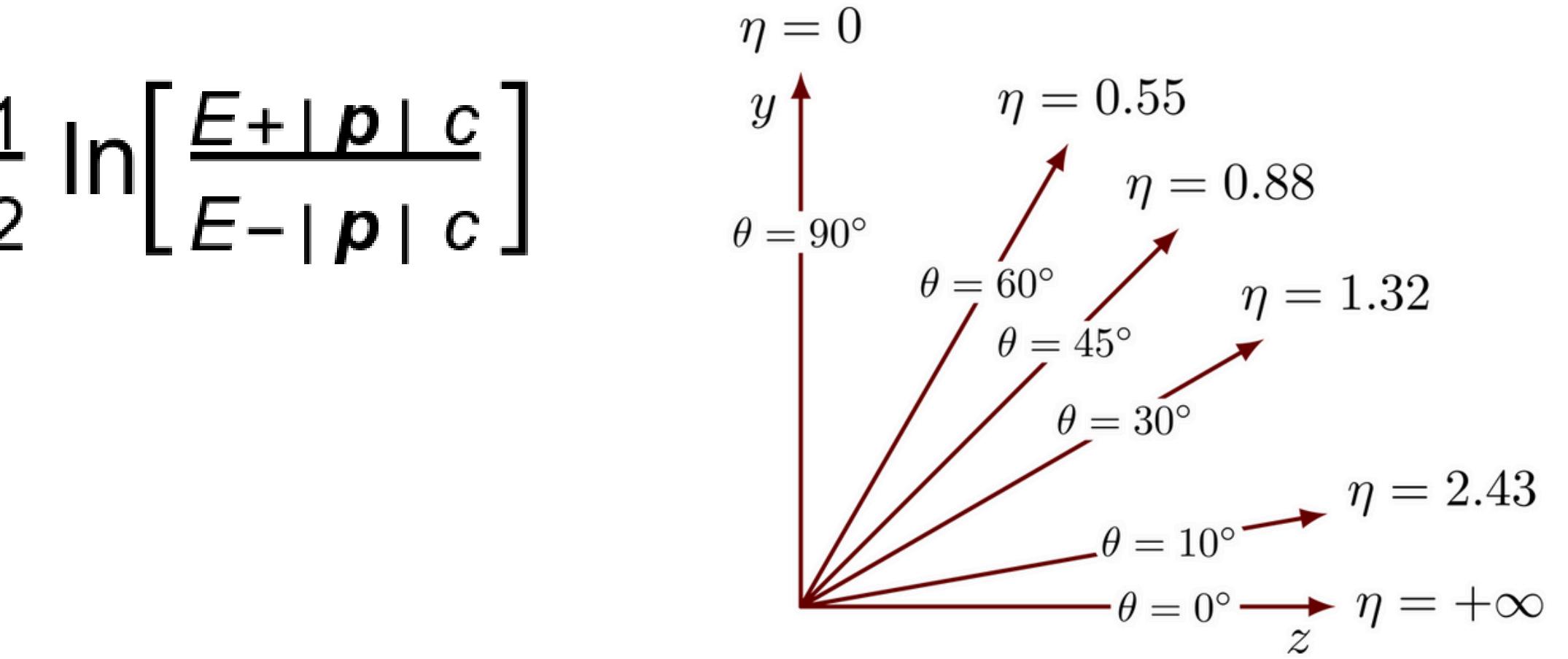


$$y = \frac{1}{2} \ln\left[\frac{E + p_z c}{E - p_z c}\right]$$



$$\eta = \frac{1}{2} \ln\left[\frac{|\vec{p}| + p_z}{|\vec{p}| - p_z}\right]$$

Pseudorapidez



Masa Invariante vs Masa Transversa

Masa invariante

Es la *masa* característica de un cuerpo o sistema que no cambia sin importar el sistema de referencia.

Para un sistema de dos partículas:

$$\begin{aligned}M^2 &= (E_1 + E_2)^2 - |\mathbf{p}_1 + \mathbf{p}_2|^2 \\&= m_1^2 + m_2^2 + 2(E_1 E_2 - \mathbf{p}_1 \cdot \mathbf{p}_2)\end{aligned}$$

En función de las coordenadas transversales:

$$M^2 = 2 p_{T,1} p_{T,2} (\cosh(\eta_1 - \eta_2) - \cos(\phi_1 - \phi_2))$$

Masa transversa

Masa característica de un cuerpo o sistema, invariante ante Boosts de Lorentz en z. En colisionadores, se usa cuando una de las partículas no puede ser detectada (e.j. neutrinos).

$$\begin{aligned}M_T^2 &= (E_{T,1} + E_{T,2})^2 - (\mathbf{p}_{T,1} + \mathbf{p}_{T,2})^2 \\&= m_1^2 + m_2^2 + 2(E_{T,1} E_{T,2} - \mathbf{p}_{T,1} \cdot \mathbf{p}_{T,2})\end{aligned}$$

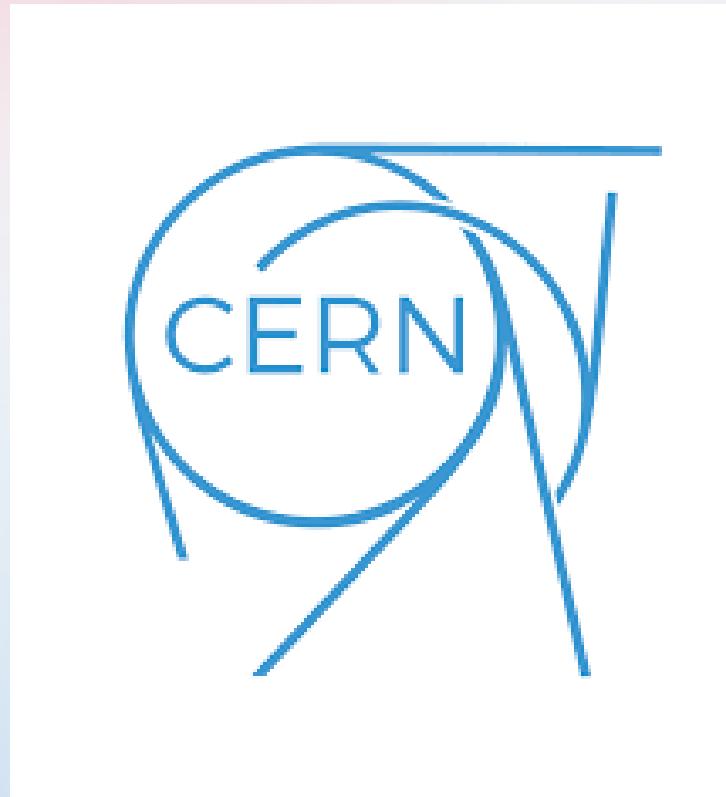
$$E_T^2 = E^2 - p_z^2$$

CMS OPEN DATA



Datos a utilizar en este taller:

- CMS Collaboration (2016). DoubleMu primary dataset in AOD format from RunA of 2011 (/DoubleMu/Run2011A-12Oct2013-v1/AOD). CERN Open Data Portal. DOI:[10.7483/OPENDATA.CMS.RZ34.QR6N](https://doi.org/10.7483/OPENDATA.CMS.RZ34.QR6N)
- CMS Collaboration, McCauley, Thomas (s.f.). Datasets derived from the Run2011A SingleElectron, SingleMu, DoubleElectron, and DoubleMu primary datasets. Data recorded in 2011 and published in 2017. <https://opendata.cern.ch/record/545>
- CMS Collaboration (2024). DoubleMuon primary dataset in NANOAOD format from RunG of 2016 (/DoubleMuon/Run2016G-UL2016_MiniAODv2_NanoAODv9-v2/NANOAOD). CERN Open Data Portal. DOI:[10.7483/OPENDATA.CMS.ZQS3.LGLP](https://doi.org/10.7483/OPENDATA.CMS.ZQS3.LGLP)



Formato de los Datos

TIPO	CONTENIDO	PROPÓSITO	TAMAÑO/EVENTO (MB)
DAQ-RAW	Datos del detector provenientes de la electrónica frontal + resultado del disparador de nivel 1 (L1).	Registro primario del evento de física. Entrada para el HLT (High Level Trigger) en línea.	1-1.5
RAW	"Datos del detector después del formateo en línea, el resultado de L1, el resultado de HLT, potencialmente algunas de las cantidades de alto nivel calculadas durante el procesamiento del HLT.	Entrada para la reconstrucción en el *Tier-0*. Archivo primario de eventos en el CERN.	0.70-0.75
RECO	Objetos reconstruidos (trazas, vértices, jets, electrones, muones, etc.) e impactos/clusters reconstruidos.	Salida de la reconstrucción del *Tier-0* y de las posteriores pasadas de reconstrucción. Permite la rebúsqueda de trazas, etc.	1.3-1.4
AOD	Subconjunto de RECO. Objetos reconstruidos (trazas, vértices, jets, electrones, muones, etc.). Posiblemente pequeñas cantidades de información de impactos muy localizada.	Análisis de física, reajuste limitado de trazas y clusters.	0.05
MiniAOD	Subconjunto de la información del AOD, con variables adicionales y calibraciones actualizadas (siempre que sea posible).	Análisis de física.	-
NanoAOD	Información reducida del AOD, reempaquetada en un formato de Ntuple simple (sin clases EDM, solo arrays de variables).	Análisis de física.	-

<https://github.com/HydHdz/TallerDatosCEVE2026.git>

Ejercicio 1:

Graficar e Interpretar

1. Extraer los datos de una muestra seleccionada de eventos con dos muones (dos electrones).
2. Visualizar el contenido del archivo. Obtener los estadísticos de algunas de las variables. ¿Qué puede concluir?
3. Graficar las variables “phi1”, “phi2”, “eta1” y “eta2”. ¿Qué puede concluir?
4. Obtener la masa invarianta del sistema dimuonico y graficar.

<https://github.com/HydHdz/TallerDatosCEVE2026.git>

Ejercicio 2:

Selección básica de eventos

1. Utilizar las muestras “Dimuon_DoubleMu.csv” y “Dimuon_SingleMu.csv”.
2. Combinar las muestras y eliminar los eventos repetidos.
3. Calcular y graficar la masa invariante del sistema di-muónico.
4. Realizar una selección básica de los eventos.
5. Graficar nuevamente la masa invariante y comparar con el resultado anterior.

Objetos Físicos - “*Physics Objects*”

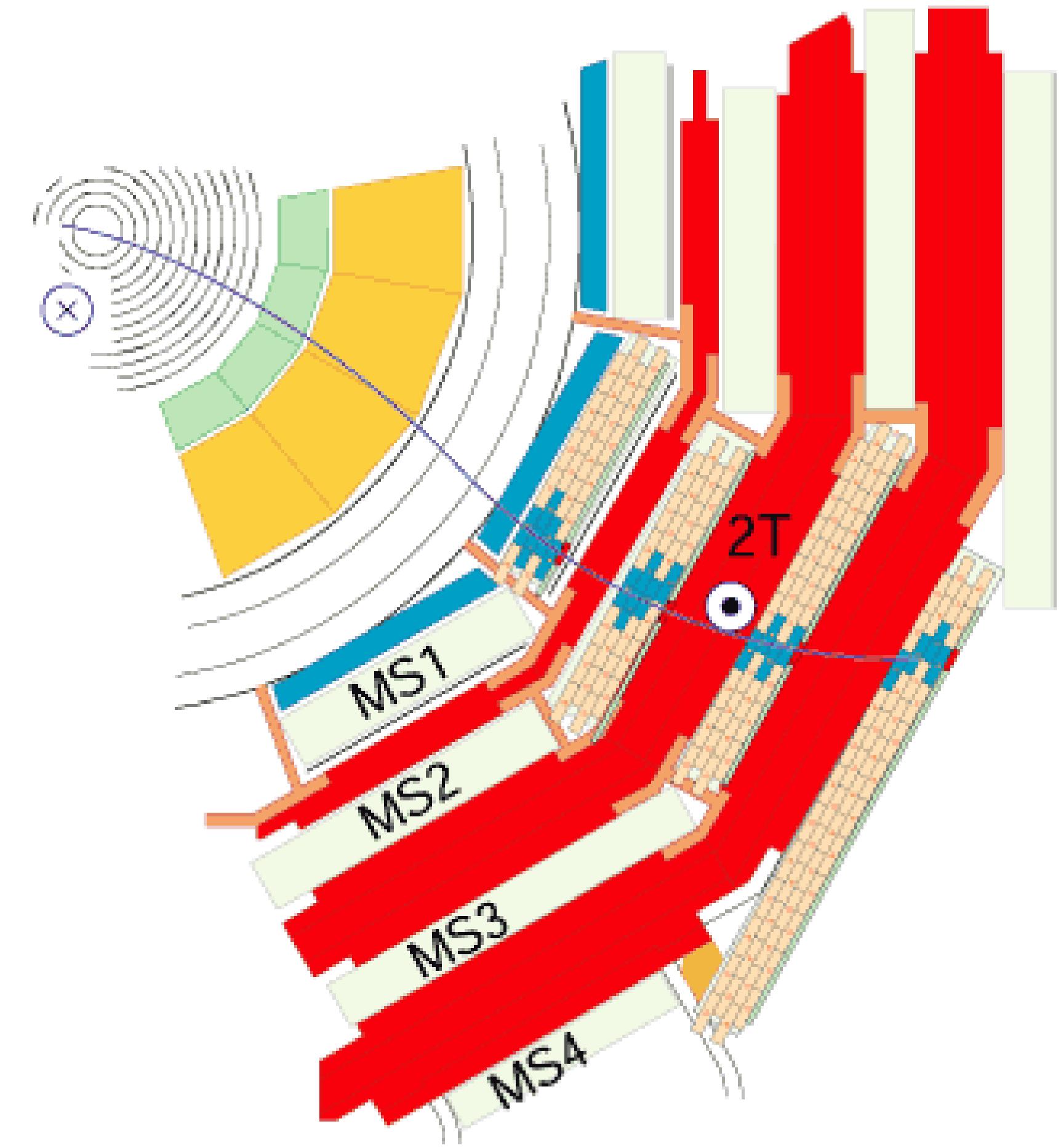
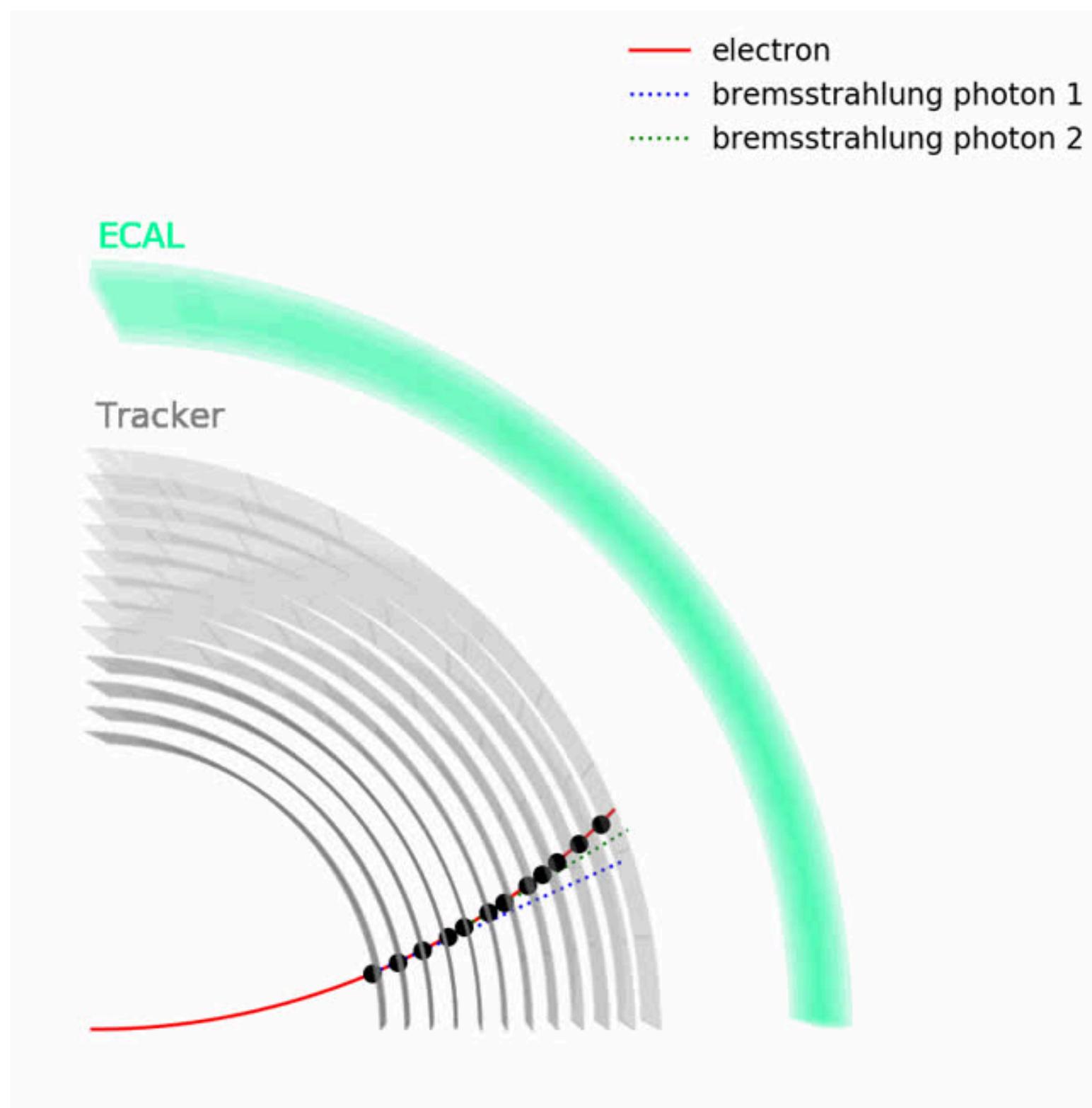
Object property	Type	Description
Electron_eta	Float_t	eta
Electron_mass	Float_t	mass
Electron_phi	Float_t	phi
Electron_pt	Float_t	$p_{\{T\}}$
electron 4-vector branches		

Object property	Type	Description
Electron_charge	Int_t	electric charge
Electron_dxy	Float_t	dxy (with sign) wrt first PV, in cm
Electron_dxyErr	Float_t	dxy uncertainty, in cm
Electron_dz	Float_t	dz (with sign) wrt first PV, in cm
Electron_dzErr	Float_t	dz uncertainty, in cm
Electron_sip3d	Float_t	3D impact parameter significance wrt first PV, in cm
Electron_tightCharge	Int_t	Tight charge criteria (0:none, 1:isGsfScPixChargeConsistent, 2:isGsfCtfScPixChargeConsistent)

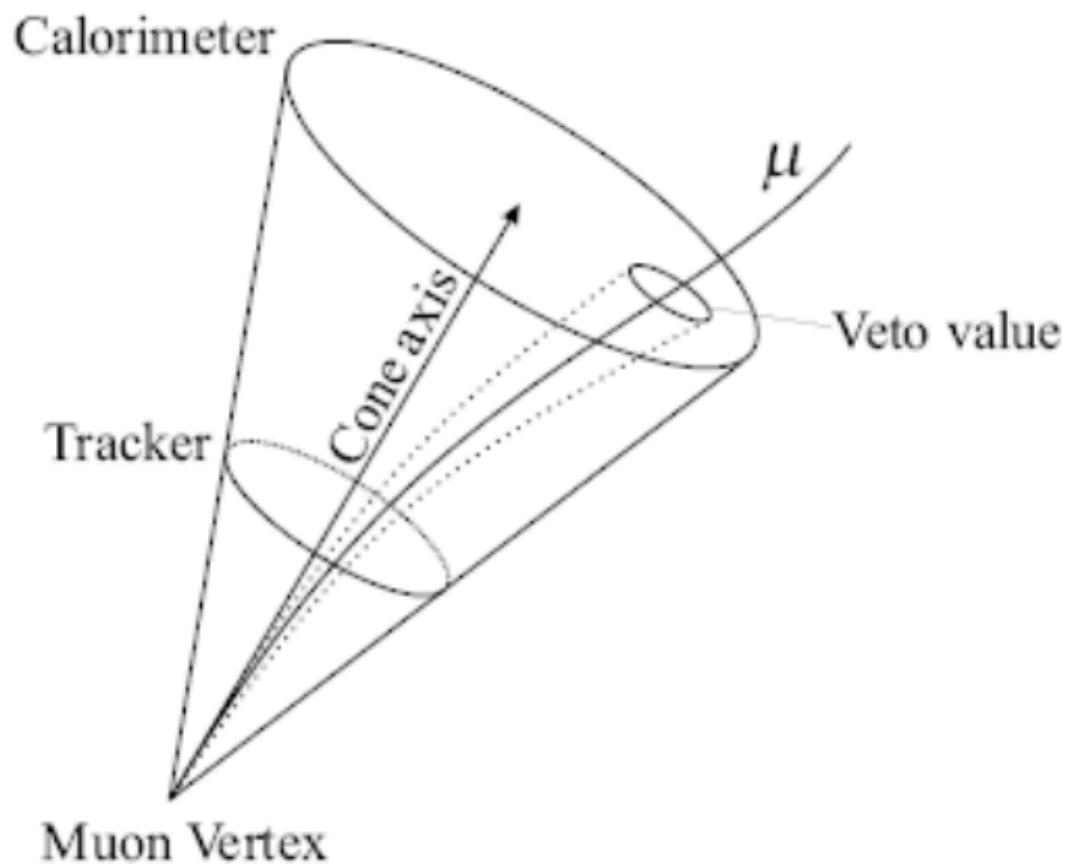
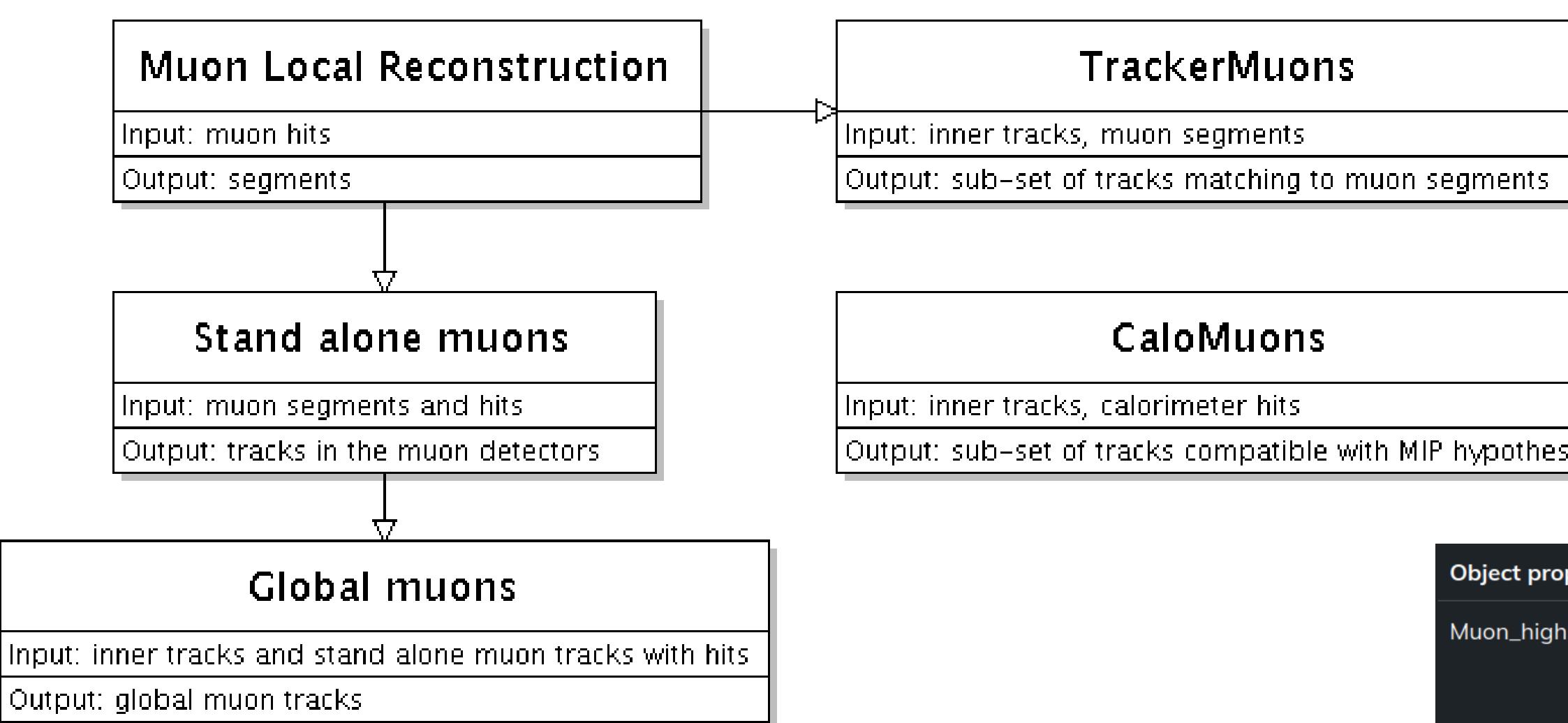
Muon_charge	Int_t	electric charge
Muon_dxy	Float_t	dxy (with sign) wrt first PV, in cm
Muon_dxyErr	Float_t	dxy uncertainty, in cm
Muon_dxybs	Float_t	dxy (with sign) wrt the beam spot, in cm
Muon_dz	Float_t	dz (with sign) wrt first PV, in cm
Muon_dzErr	Float_t	dz uncertainty, in cm
Muon_eta	Float_t	eta
Muon_ip3d	Float_t	3D impact parameter wrt first PV, in cm
Muon_mass	Float_t	mass
Muon_phi	Float_t	phi
Muon_pt	Float_t	pt
Muon_sip3d	Float_t	3D impact parameter significance wrt first PV
Muon_tightCharge	Int_t	Tight charge criterion using pterr/pt of muonBestTrack (0:fail, 2:pass)

<https://cms-opendata-workshop.github.io/workshop2024-lesson-physics-objects/index.html>





Clasificación y Tipos de Muones



Object property	Type	Description
Muon_highPtId	UChar_t	high-pT cut-based ID (1 = tracker high pT, 2 = global high pT, which includes tracker high pT)
Muon_looseId	Bool_t	muon is loose muon
Muon_mediumId	Bool_t	cut-based ID, medium WP
Muon_mediumPromptId	Bool_t	cut-based ID, medium prompt WP
Muon_mvald	UChar_t	Mva ID from miniAOD selector (1=MvaLoose, 2=MvaMedium, 3=MvaTight, 4=MvaVTight, 5=MvaVVTight)
Muon_mvaLowPtId	UChar_t	Low Pt Mva ID from miniAOD selector (1=LowPtMvaLoose, 2=LowPtMvaMedium)
Muon_mvaTTH	Float_t	TTH MVA lepton ID score
Muon_softId	Bool_t	soft cut-based ID
Muon_softMvald	Bool_t	soft MVA ID
Muon_tightId	Bool_t	cut-based ID, tight WP



Trigger

Al hacer nuestro análisis, una tarea crucial es elegir los eventos de interés de acuerdo a cierto número de criterios.

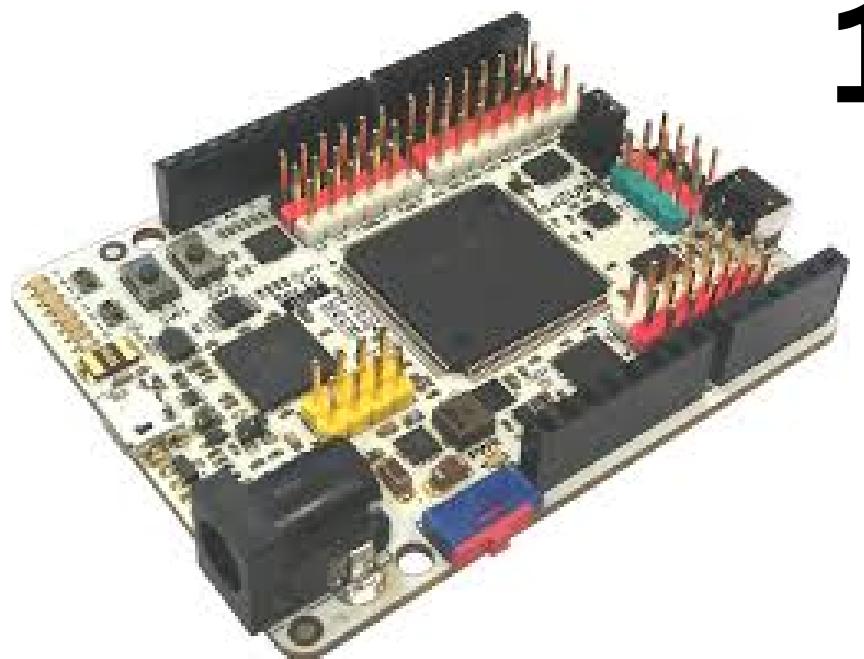
En el LHC ocurren 40 millones de colisiones por segundo (40 MHz) !

No todos los eventos son de interés científico. Guardar solo los más importantes es crucial.

Sistema de Trigger

**Level 1 Trigger
(L1)**

40 MHz



**High Level Trigger
(HLT)**

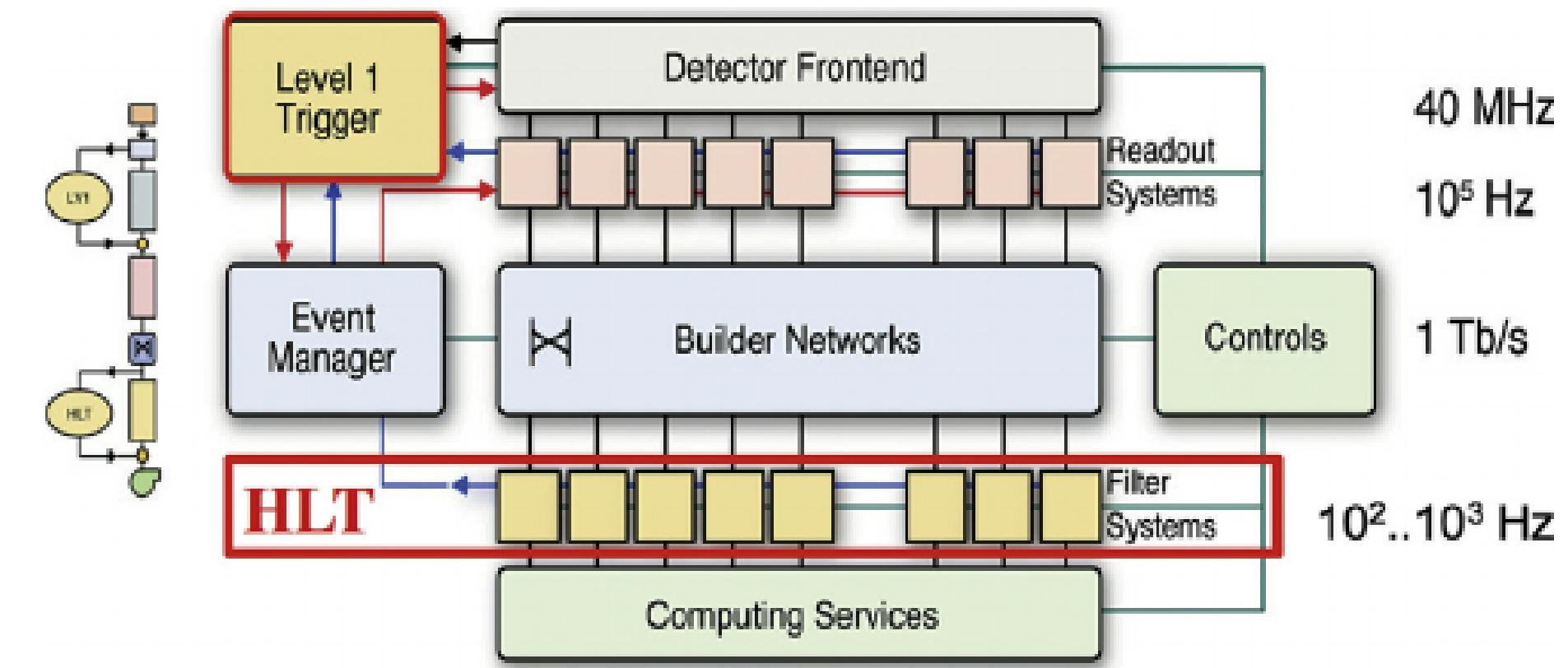
100 KHz



2 KHz

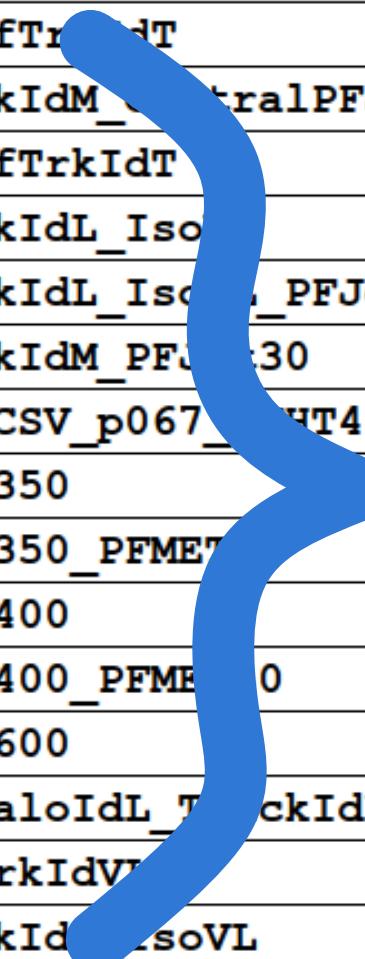
Puntos a considerar

- Los triggers va **actualizándose** incluso de corrida a corrida (e.j. HLT_Mu15_v1, HLT_Mu15_v2...).
- Pre-escalas (**Prescales**). Para ahorrar espacio, para algunos triggers solo se guardan 1 evento de 100, 1 de 200...
- Un **mismo evento** puede *pasar* la prueba de **dos o más triggers**.
- Un evento puede aparecer en más de una muestra. ¡Cuidado con **repetir el conteo** si se usan más de una muestra en un análisis!



HLT_DoubleMu3_PFMET50
HLT_DoubleMu3_Trk_Tau3mu
HLT_DoubleMu4_3_Bs
HLT_DoubleMu4_3_Jpsi_Displaced
HLT_DoubleMu4_JpsiTrk_Displaced
HLT_DoubleMu4_LowMassNonResonantTrk_Displaced
HLT_DoubleMu4_PsiPrimeTrk_Displaced
HLT_DoubleMu8_Mass8_PFHT250
HLT_DoubleMu8_Mass8_PFHT300
HLT_DoublePhoton60
HLT_DoublePhoton85
HLT_DoubleTightCombinedIsoPFTau35_Trk1_eta2p1_Reg
HLT_DoubleTightCombinedIsoPFTau40_Trk1_eta2p1
HLT_DoubleTightCombinedIsoPFTau40_Trk1_eta2p1_Reg
HLT_ECALHT800
HLT_EcalCalibration
HLT_Ele105_CaloIdVT_GsfTrkIdT
HLT_Ele10_CaloIdM_TrackIdM_CentralPFJet30_BTagCSV_p13
HLT_Ele115_CaloIdVT_GsfTrkIdT
HLT_Ele12_CaloIdL_TrackIdL_IsoVL_PFJet30
HLT_Ele12_CaloIdL_TrackIdL_IsoVL_PFJet30
HLT_Ele12_CaloIdM_TrackIdM_PFJet30
HLT_Ele15_IsoVVVL_BTagCSV_p067_PFHT400
HLT_Ele15_IsoVVVL_PFHT350
HLT_Ele15_IsoVVVL_PFHT350_PFMET50
HLT_Ele15_IsoVVVL_PFHT400
HLT_Ele15_IsoVVVL_PFHT400_PFMET50
HLT_Ele15_IsoVVVL_PFHT600
HLT_Ele16_Ele12_Ele8_CaloIdL_TrackIdL
HLT_Ele17_CaloIdL_GsfTrkIdVLPFHT400
HLT_Ele17_CaloIdL_TrackIdL_IsoVL

>=2 Muones

A large blue curly brace is positioned on the left side of the slide, spanning from the bottom of the first list of HLT triggers up to the second list. It groups the two main sections of the slide.

>=1 Electrón

Ejemplos:

HLT_Ele35_WPTight_Gsf

HLT_IsoMu27

HLT_IsoMu27_LooseChargedIsoPFTauHPS20_Trk1_eta2p1_SingleL1

“Ele” para electrón, “Mu” para muón, “Tau” para tau.

“Iso” = aislado.

“35”, “27” se refiere al mínimo en p_T o E_T.

HPS = “Hadron plus Strips”, el tau decayó a hadrones.

“eta2p1” indica $\eta < 2.1$

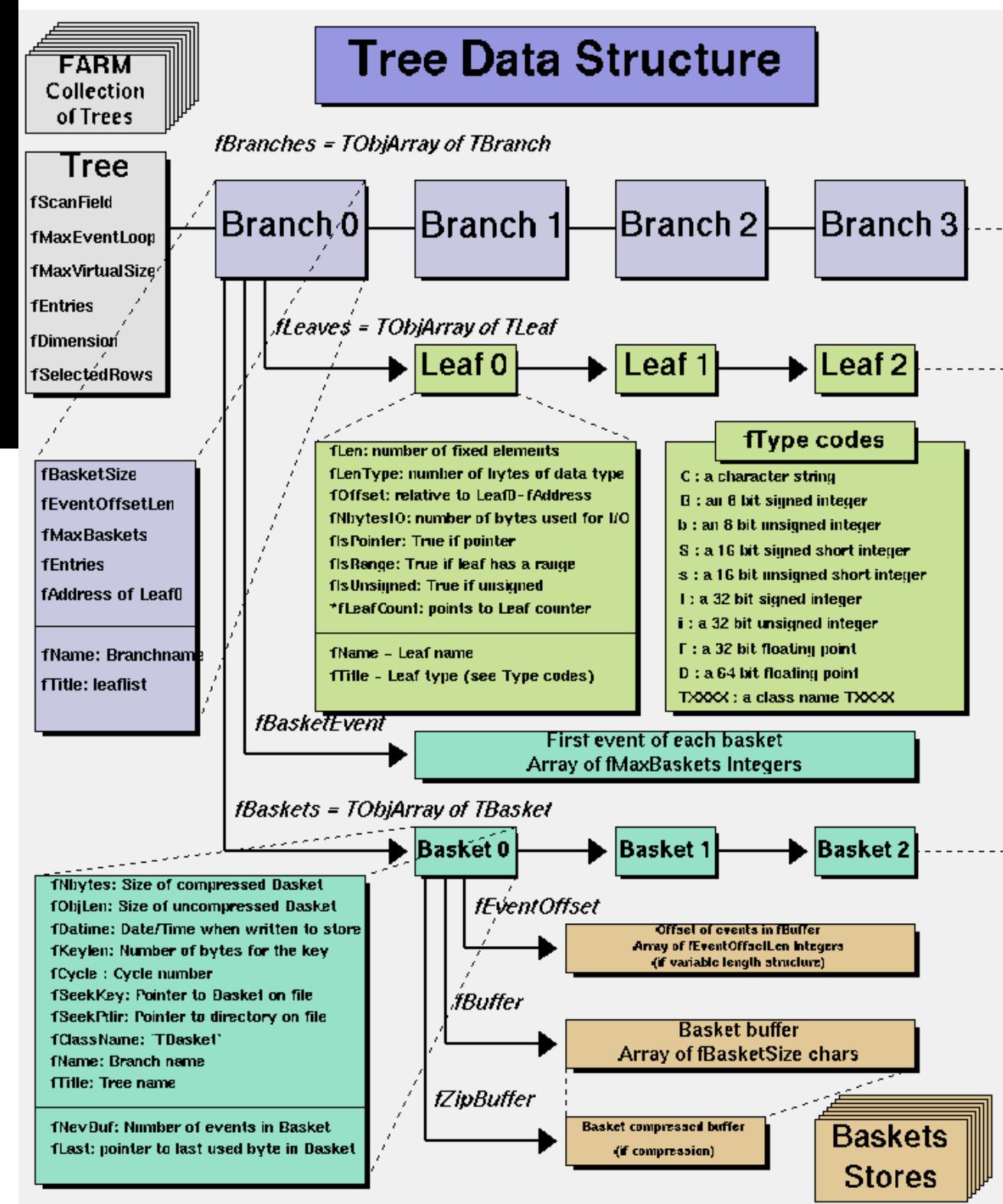
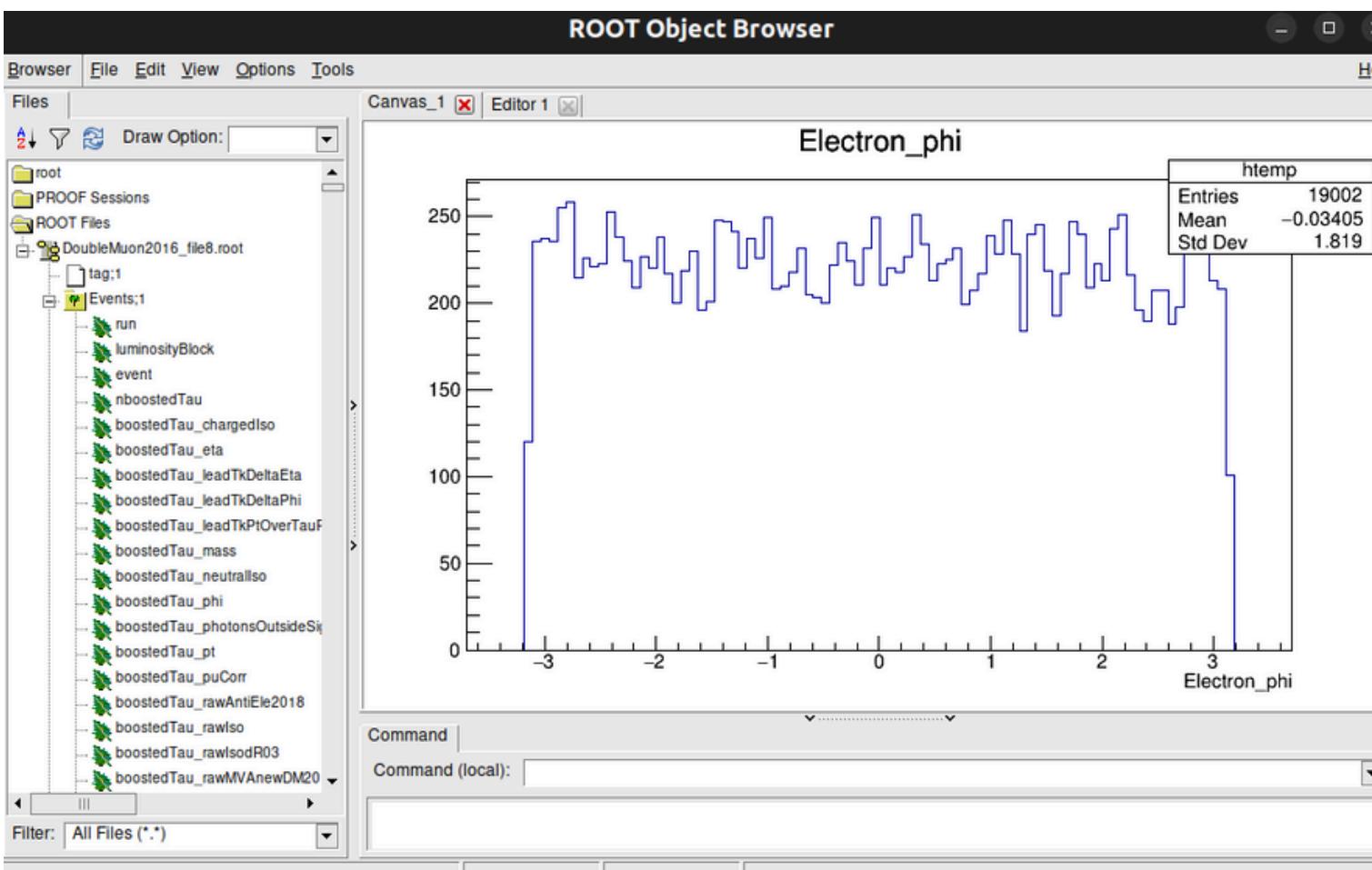
FORMATO ROOT

```

| Welcome to ROOT 6.28/06          https://root.cern
| (c) 1995-2023, The ROOT Team; conception: R. Brun, F. Rademakers
| Built for linuxx86_64-gcc on Jan 10 2024, 01:13:29
| From heads/latest-stable@7745d36d
| With c++ (Ubuntu 11.4.0-1ubuntu1~22.04) 11.4.0
| Try '.help'/'?'. 'demo', 'license', 'credits', '.quit'/.q'
-----
```

not a dynamic executable

root [0]



<https://github.com/HydHdz/TallerDatosCEVE2026.git>

Ejercicio 3:

Análisis de Datos Abiertos con UpROOT

1. Utilizar un fragmento de la muestra /DoubleMuon/Run2016G en formato NANO AOD. Abrir el archivo y leerlo con UpROOT.
2. Seleccionar las ramas relevantes para el análisis y guardarlas en formato DataFrame.
3. Realizar la selección de eventos para reconstruir la masa del bosón Z.
4. Discutir EN EQUIPO qué otras ramas y cortes realizar para filtrar los eventos. Presente sus hallazgos en una presentación corta (**7 minutos**).