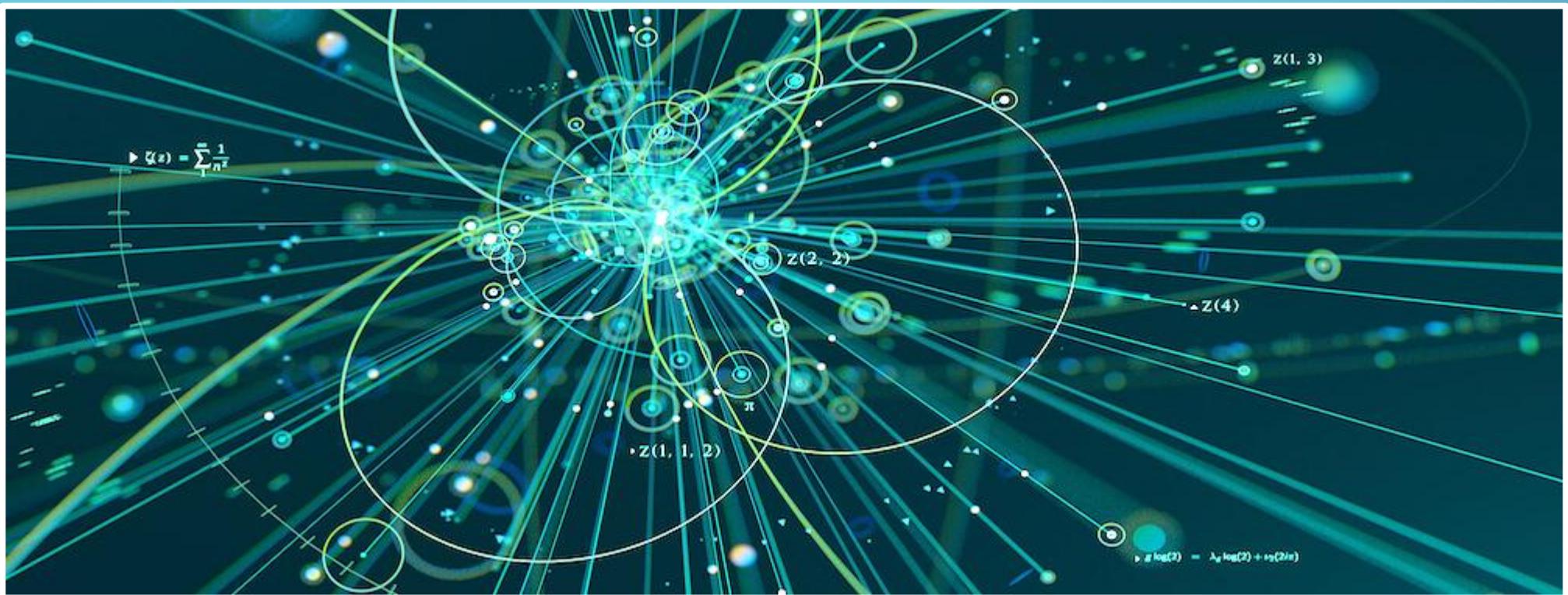


# DATALAB: FÍSICA DE PARTÍCULAS



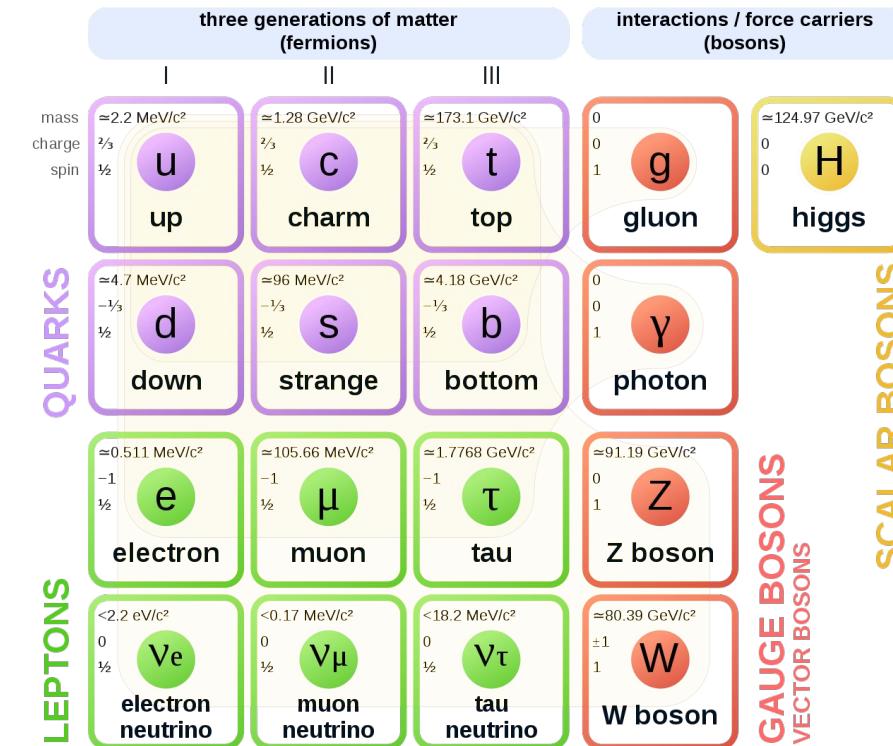
Lara Lloret Iglesias ([lloret@ifca.unican.es](mailto:lloret@ifca.unican.es))  
Celia Fernández Madrazo ([fmadrazoc@unican.es](mailto:fmadrazoc@unican.es))  
Master en Ciencia de Datos (UC/UIMP)

# EL MODELO ESTÁNDAR

El **Modelo Estándar** de partículas elementales resumen nuestro conocimiento actual del mundo microscópico:

- Describe los **constituyentes de la materia** (quarks y leptones) y sus **interacciones** (mediado por bosones)
- El éxito más reciente: descubrimiento del **bosón de Higgs** en 2012 por los experimentos ATLAS y CMS (LHC)
- Pero quedan muchas preguntas por responder!
- Tratamos de responderlas con medidas de precision y buscando “nueva física”....

Standard Model of Elementary Particles



# FUERZAS FUNDAMENTALES

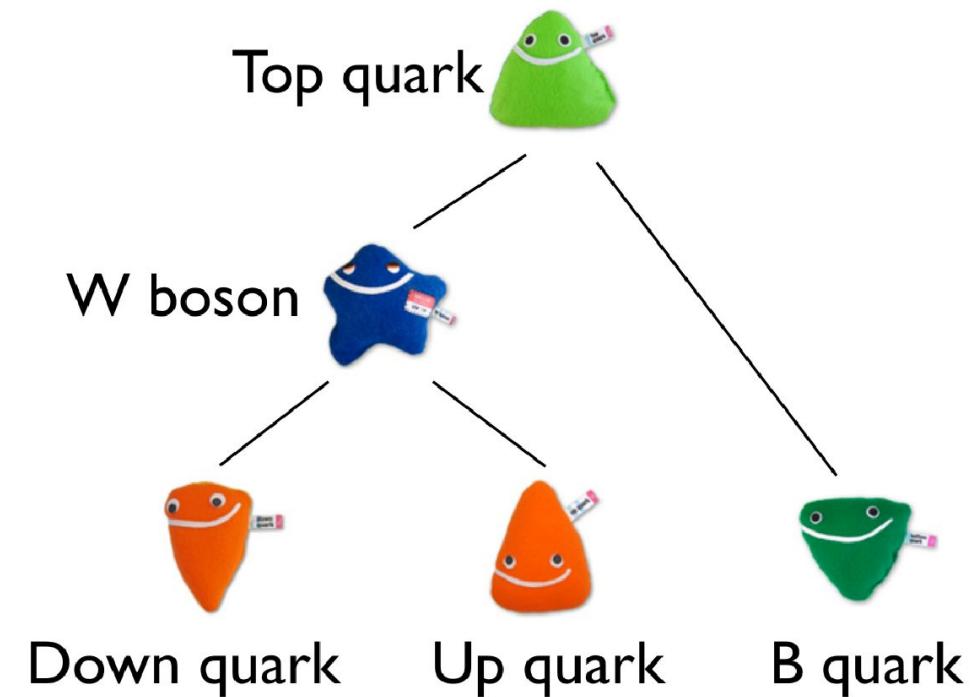
Fundamental Force Particles

Force	Particles Experiencing	Force Carrier Particle	Range	Relative Strength*
<b>Gravity</b> acts between objects with mass	all particles with mass	graviton (not yet observed)	infinity	much weaker
<b>Weak Force</b> governs particle decay	quarks and leptons	$W^+$ , $W^-$ , $Z^0$ (W and Z)	short range	
<b>Electromagnetism</b> acts between electrically charged particles	electrically charged	$\gamma$ (photon)	infinity	
<b>Strong Force**</b> binds quarks together	quarks and gluons	$g$ (gluon)	short range	much stronger

# EL QUARK TOP

- El **quark top** es la partícula fundamental más pesada conocida (172.5 GeV)
- Tiene una vida media muy corta ( $10^{-25}$  segundos): **solo podemos ver los productos de su desintegración**
- Se descubrió en 1995 en los experimentos D0 y CDF en Fermilab (Chicago)
- Es una partícula clave para buscar nueva física más allá del Modelo Estándar y para realizar medidas de precisión
- La desintegración más desafiante (e interesante) del quark top es la **desintegración hadrónica**:

$$t \rightarrow W b \rightarrow q q' b$$

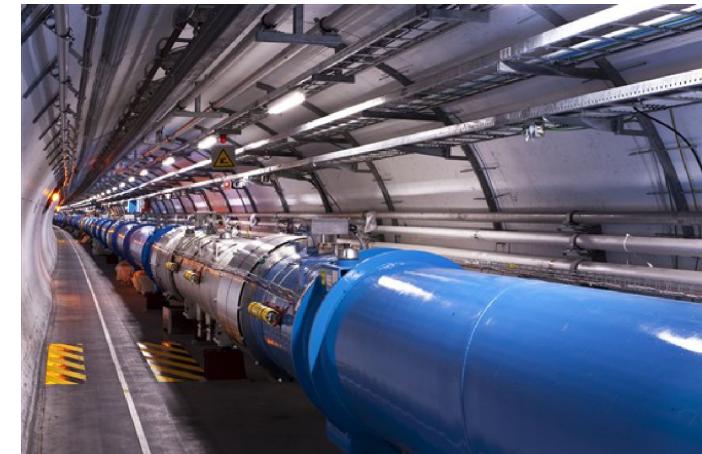
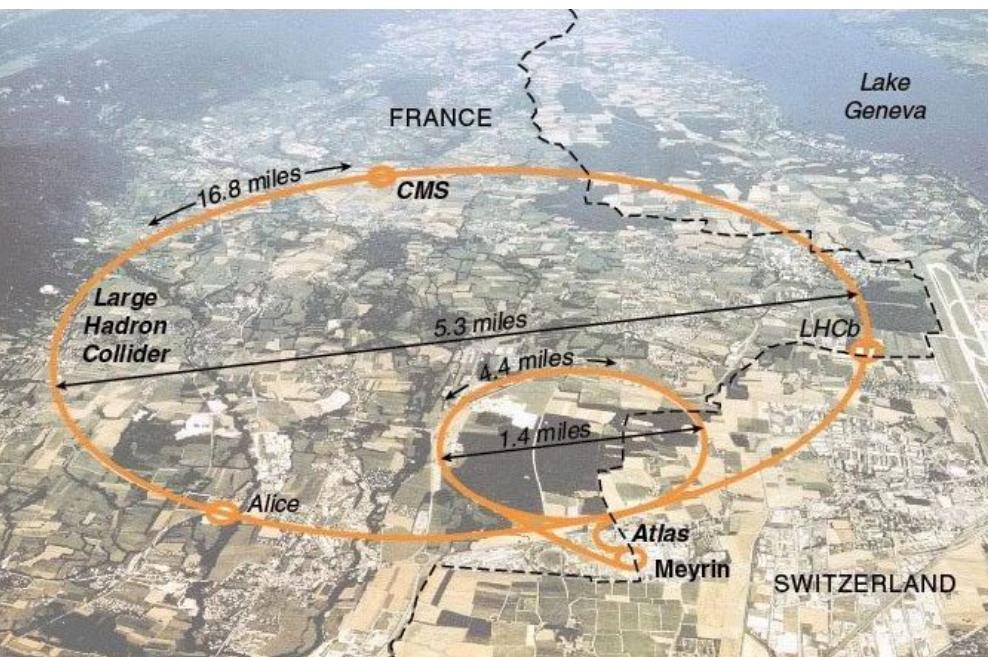


En este imagen q y q' son el quark down y up

# ¿CÓMO ENCONTRAR UN QUARK TOP?

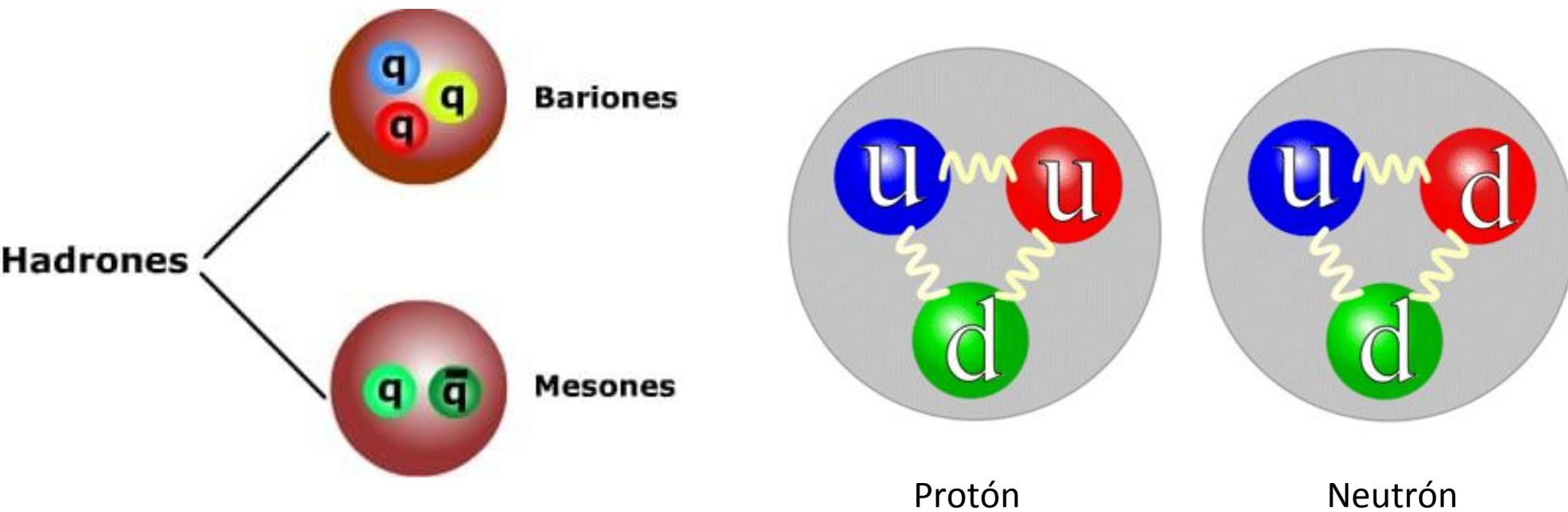
□ PRODUCIÉNDOLO: En un acelerador como el LHC

# LHC: LARGE HADRON COLLIDER



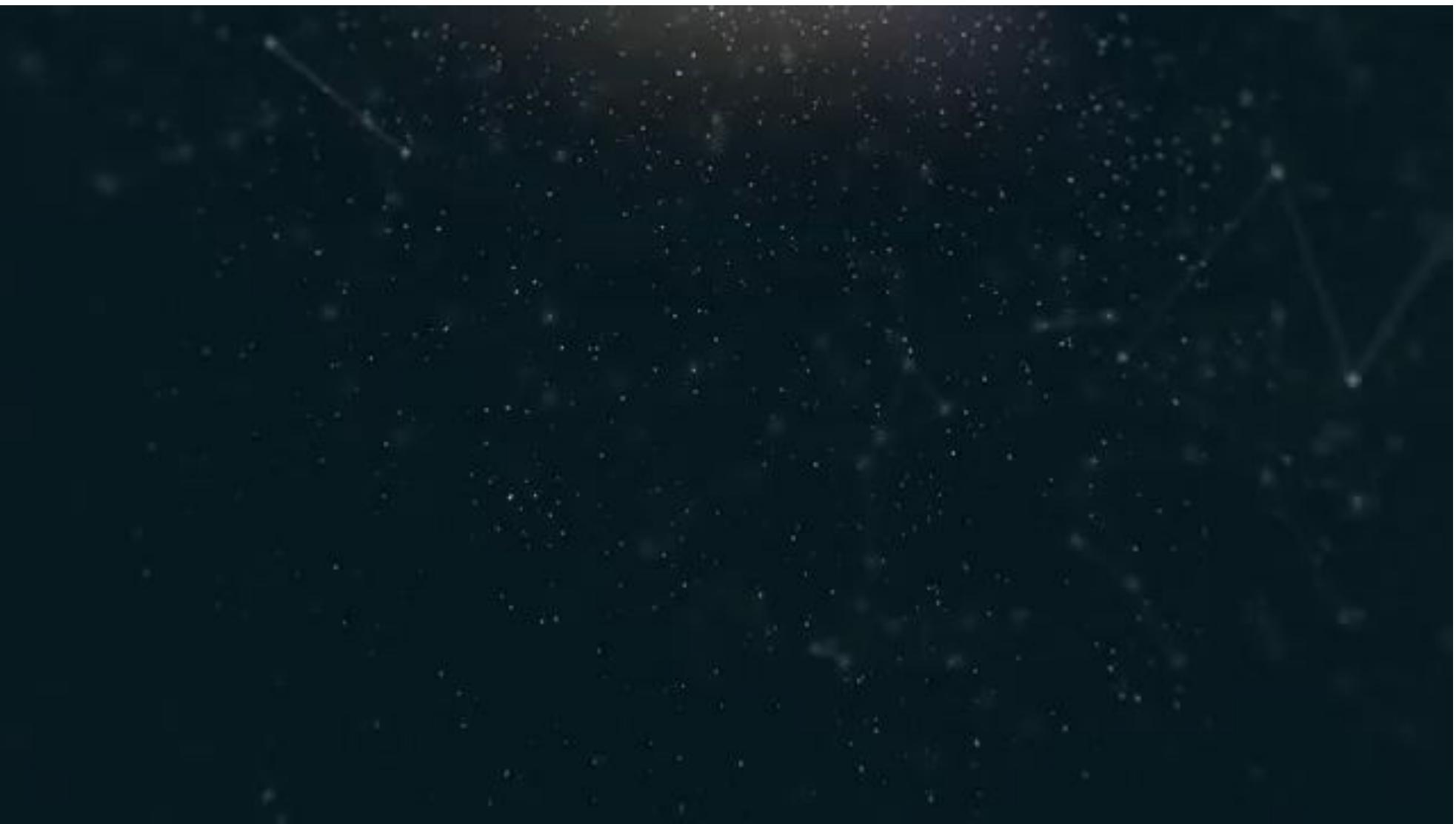
# LHC: LARGE HADRON COLLIDER

Un hadrón (del griego ἀδρός, hadrós, "denso" o "fuerte") es una partícula subatómica formada por quarks que permanecen unidos debido a la interacción nuclear fuerte entre ellos



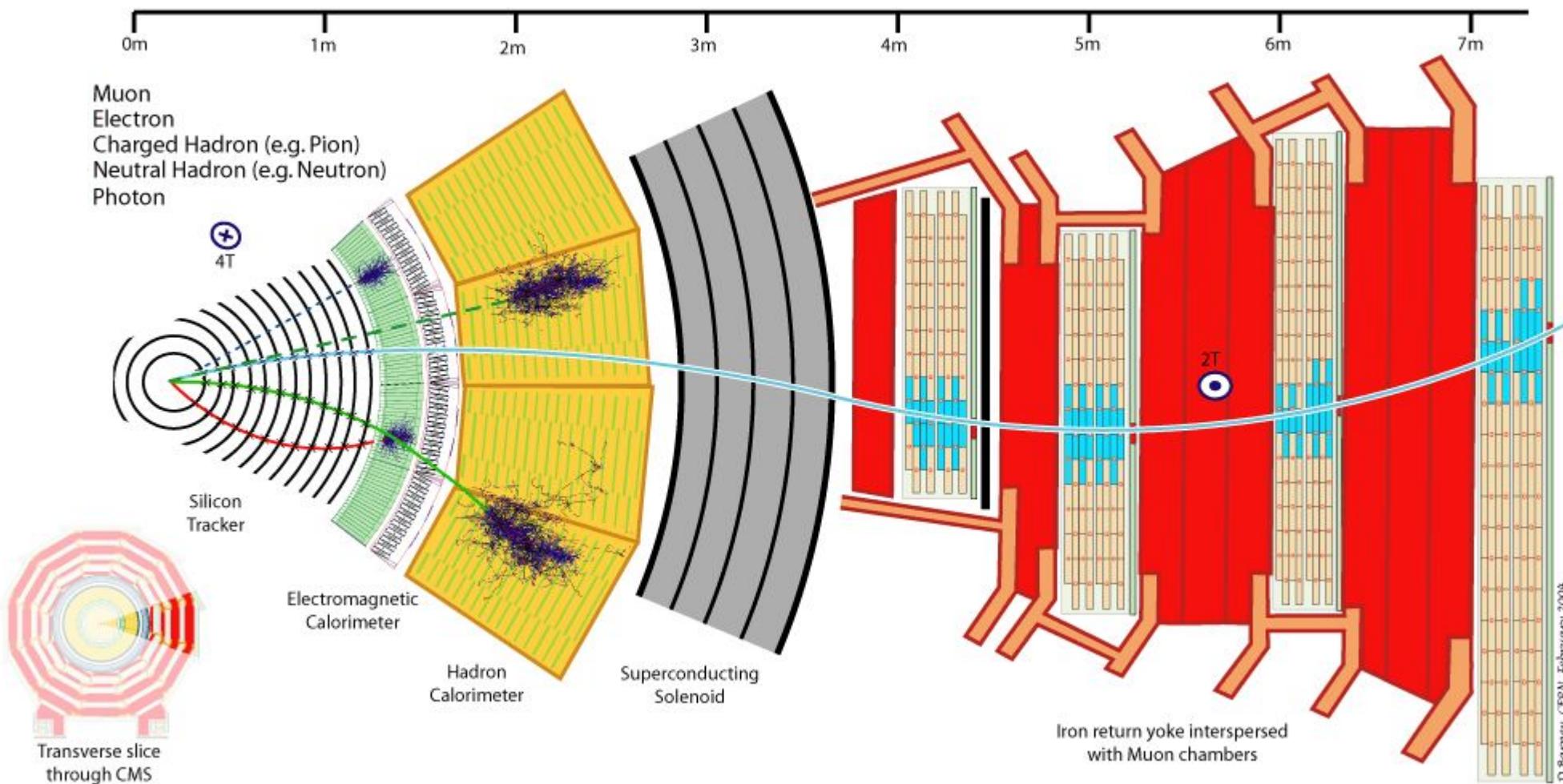
# ¿CÓMO ENCONTRAR UN QUARK TOP?

- PRODUCIÉNDOLO: En un acelerador como el LHC
- DETECTÁNDOLO: Con un detector (como CMS) que reconstruya la energía y la posición de cada partícula.

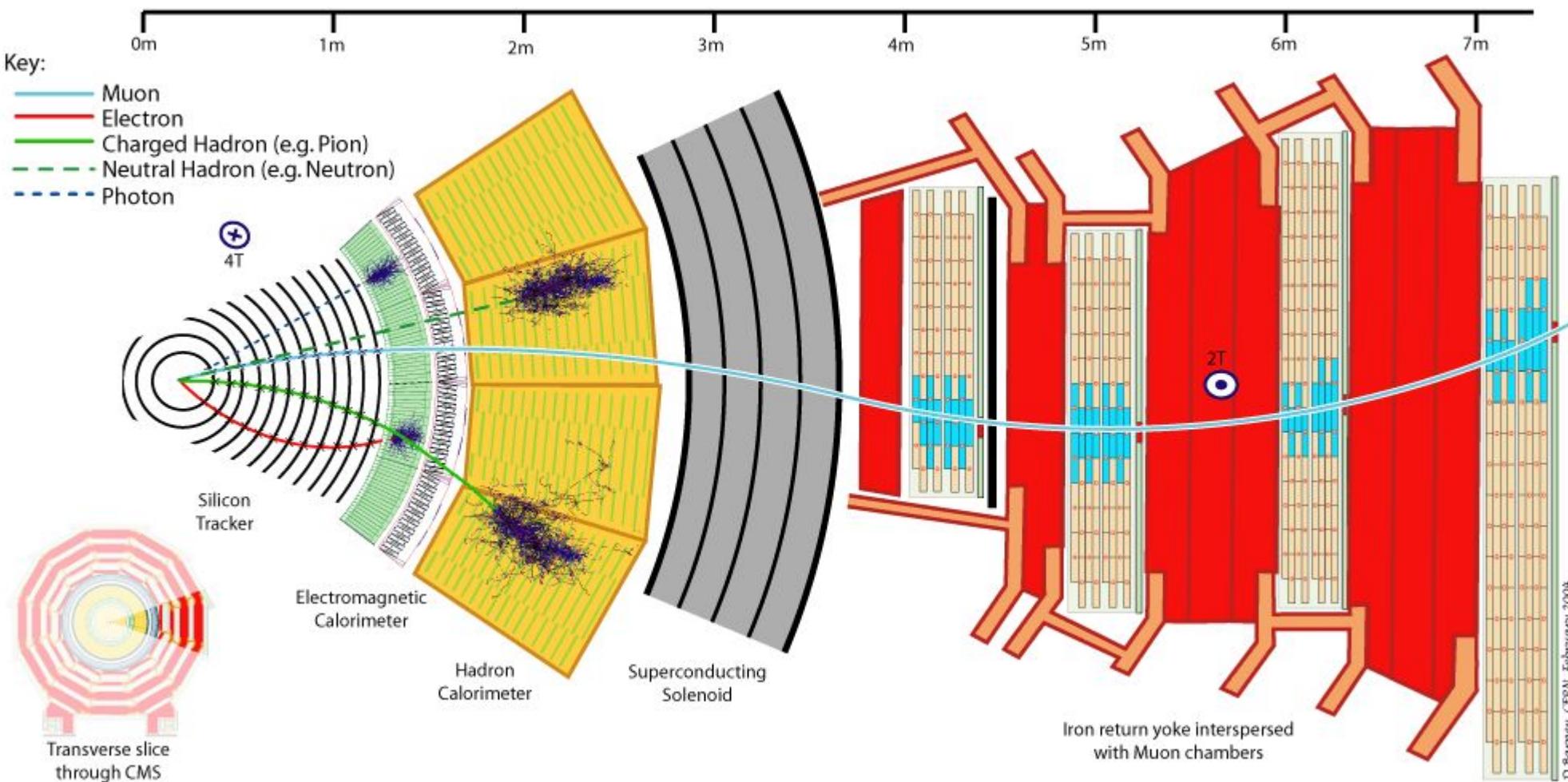


<https://www.youtube.com/watch?v=S99d9BQmGB0>

# EXPERIMENTO CMS

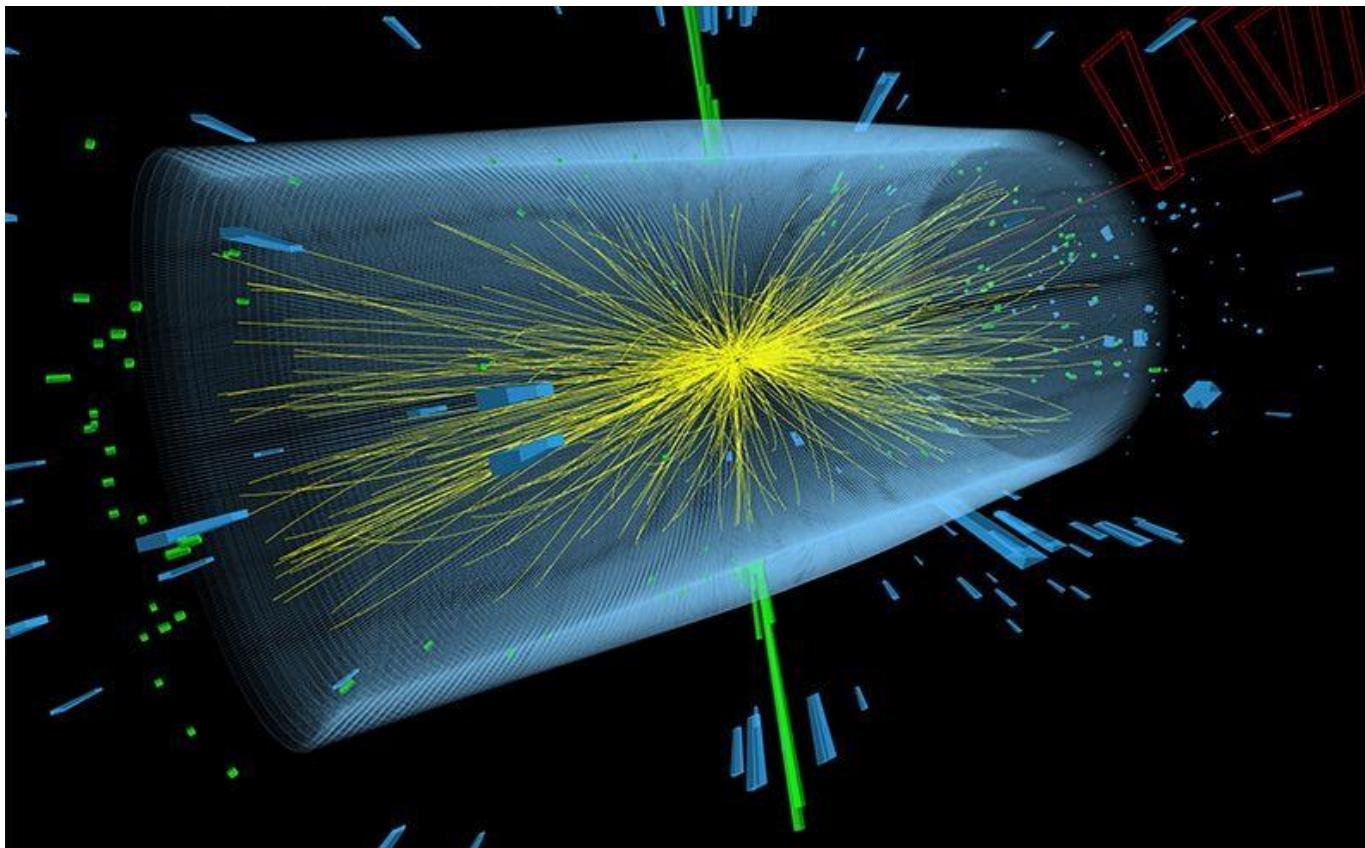


# EXPERIMENTO CMS



# PROCESO ESTADÍSTICO

- Millones de colisiones producidas cada segundo
- Los partículas se generan con una probabilidad (sección eficaz) diferente

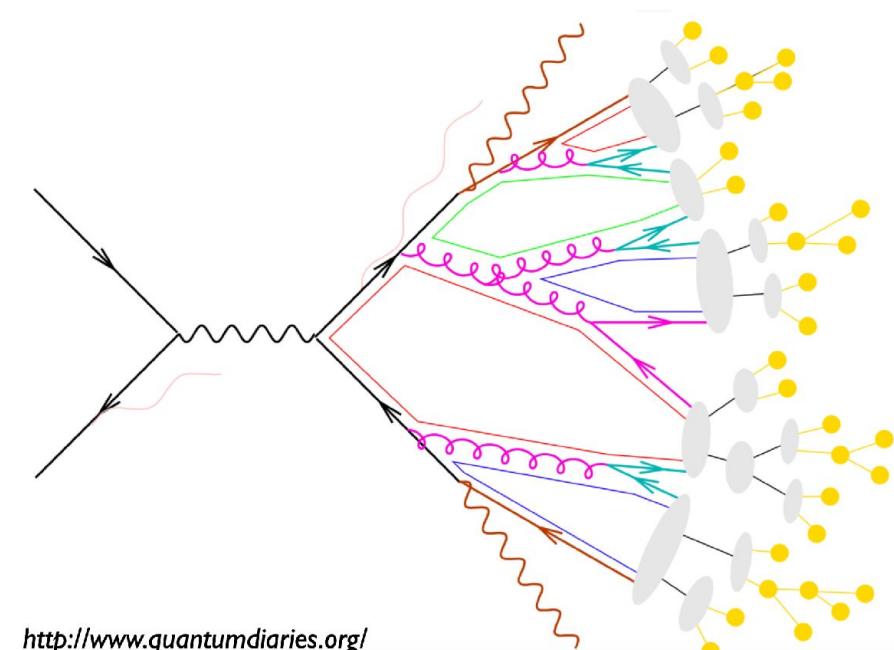


# ¿CÓMO ENCONTRAR UN QUARK TOP?

- PRODUCIÉNDOLO: En un acelerador como el LHC
- DETECTÁNDOLO: Con un detector (como CMS) que reconstruya la energía y la posición de cada partícula.
- RECONSTRUYÉNDOLO: Con un algoritmo de reconstrucción de jets que combine las partículas en un objeto de alto nivel

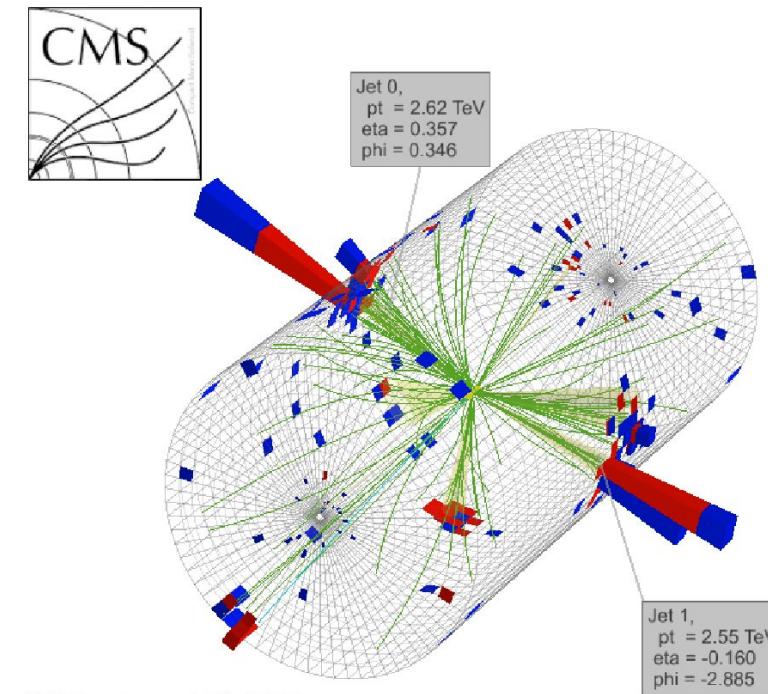
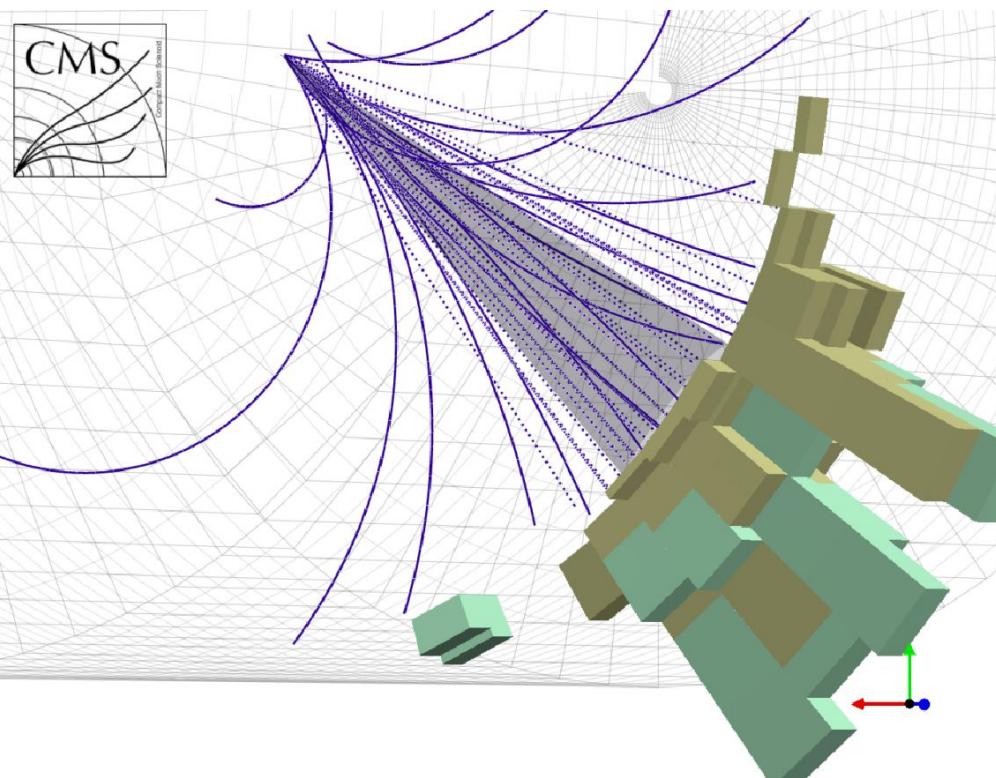
# ¿POR QUÉ ES COMPLICADO EL TOP TAGGING?

- Debido a la naturaleza de la interacción fuerte los quarks no viajan libres
- Son forzados a estar confinados dentro de hadrones
- Los quarks no se detectan como una partícula aislada, pero como un jet de partículas
- Los algoritmos de reconstrucción de jets clusterizan las partículas que vienen de un quark
- Se diseñan de manera que el momento del jet clusterizado es proporcional a la energía inicial del quark



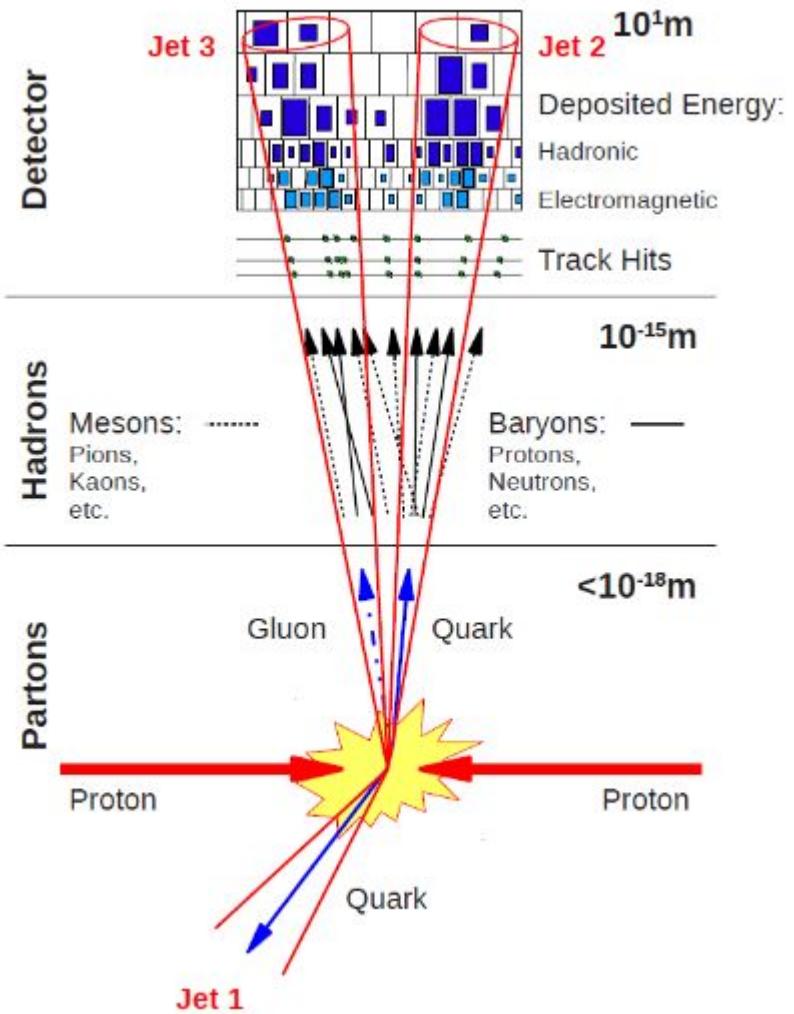
<http://www.quantumdiaries.org/>

# ALGORITMO RECONSTRUCCIÓN DE JETS



CMS Experiment at LHC, CERN  
Data recorded: Sun Jul 12 01:52:51 2015 CDT  
Run/Event: 251562 / 310157776  
Lumi section: 347  
Dijet Mass : 5.4 TeV

# DE QUARKS A JETS



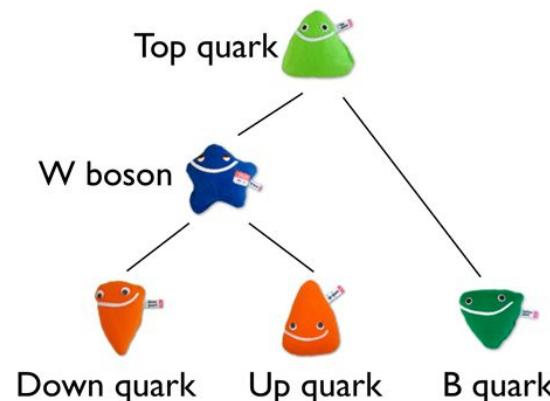
# ¿CÓMO ENCONTRAR UN QUARK TOP?

- PRODUCIÉNDOLO: En un acelerador como el LHC
- DETECTÁNDOLO: Con un detector (como CMS) que reconstruya la energía y la posición de cada partícula. Los quarks se detectarán como jets de partículas.
- RECONSTRUYÉNDOLO: Con un algoritmo de reconstrucción de jets que combine las partículas en un objeto de alto nivel
- DISTINGUIÉNDOLO → Utilizar los conocimientos de física para entender las diferencias entre señal y fondo

# ¿POR QUÉ ES COMPLICADO EL TOP TAGGING?

- Producir quarks top es “difícil”
- La producción de quarks top es relativamente rara (sección eficaz pequeña)
- Otros procesos iniciados por la interacción fuerte (QCD) ocurren mucho más a menudo.
- Producen quarks más ligeros (up, down, strange...)
- Tienen un aspecto similar a los quarks top pero pasan mucho más a menudo
- ¡Nuestro reto es luchar contra este fondo!

SEÑAL (RARO)



FONDO (COMÚN)

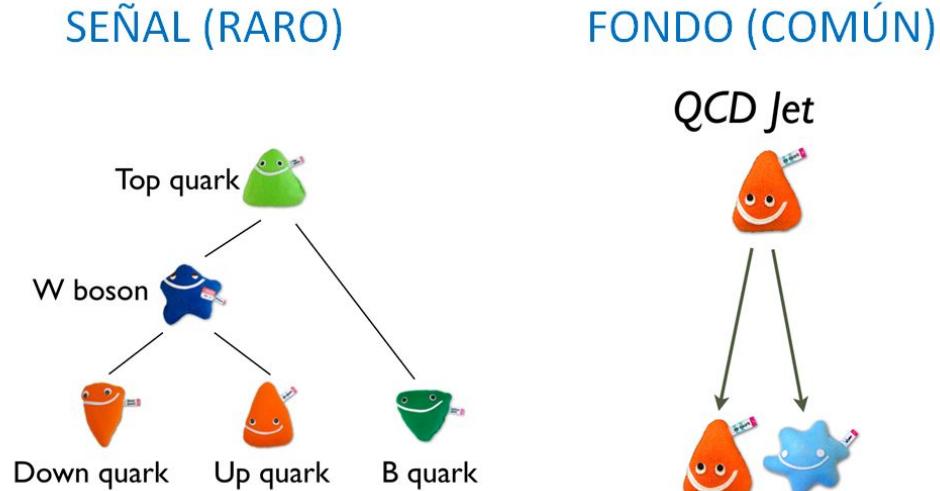
*QCD Jet*



# INTUICIÓN FÍSICA

Idea intuitiva:

- Los quarks top producen 3 quarks
- Los procesos mediados por interacción fuerte (normalmente) 1 quark
- *n-subjettiness*: distingue cuantos sub-jets hay en un jet
- Top → 3-pronged jet
- QCD → 1-pronged jet
- La masa invariante del jet puede ser otro buen discriminante
- Estas propiedades pueden ser aprendidas con ML y DL



# OBJETIVO DEL DATALAB: TOP TAGGING CON ML

**Problema propuesto:** Abordar el problema de top tagging utilizando técnicas de Machine Learning:

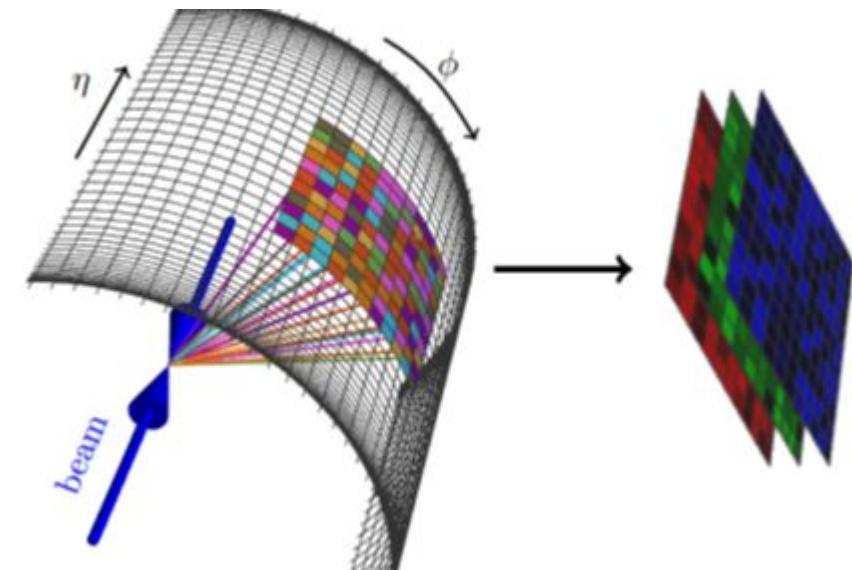
- Este algoritmo tendrá como objetivo discriminar jets que vengan de un quark top (señal) de jets generados por otros procesos físicos como QCD (fondo).
- Hay 2 puntos de partida:
  - **Top tagging 1:** Utilizar los valores numéricos de energía y momento de las partículas que constituyen el jet.
  - **Top tagging 2:** Utilizar imágenes construidas con los valores de energía de esas partículas.

## TOP TAGGING 1

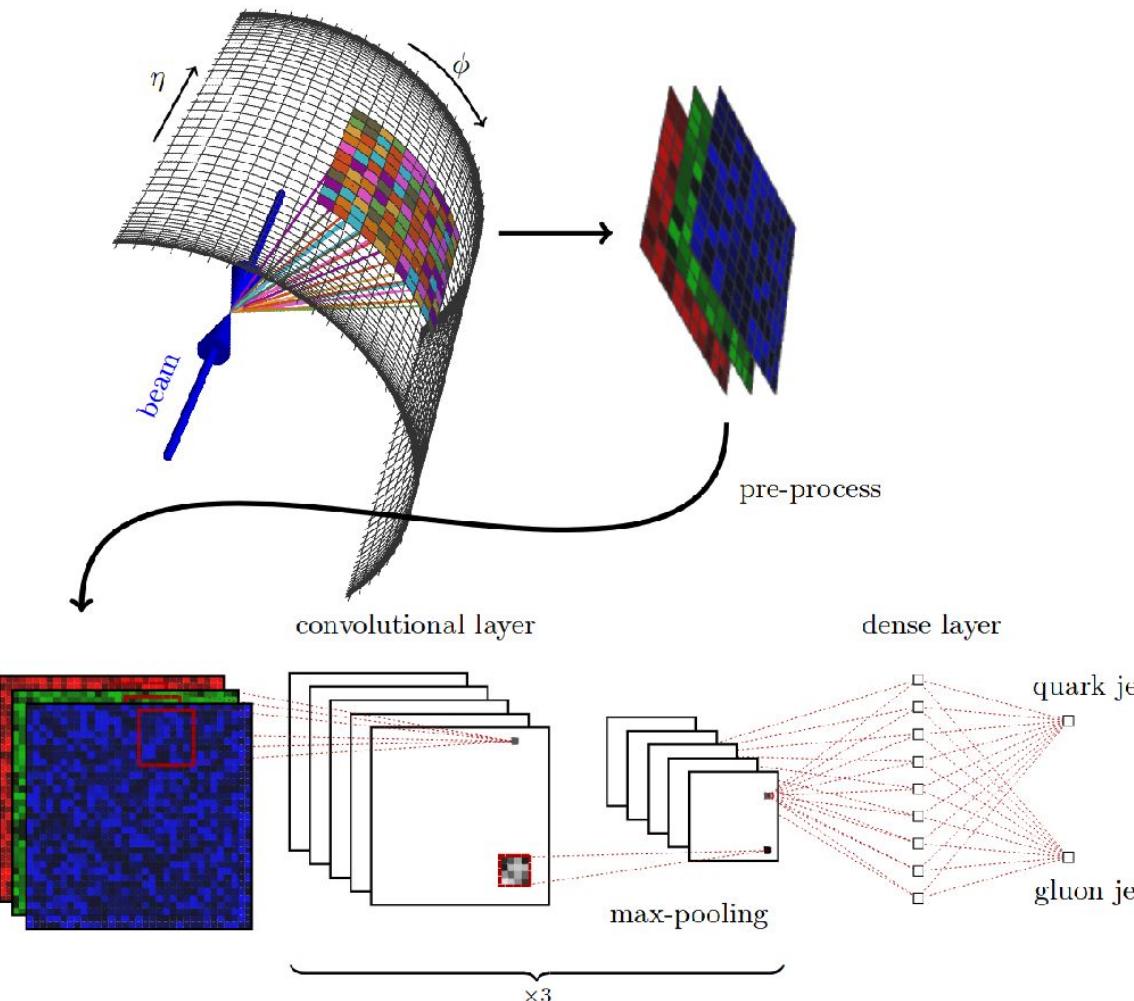
- Utilizaréis el cuadrimomento de las partículas que han sido clusterizadas dentro de un jet como **valores de input de la red**.
- **E, px, py, pz de 200 constituyentes** de jets almacenados en un dataframe de pandas.
- Los constituyentes están ordenados por su momento transverso (el primer constituyente es el más energético).
- Se le asigna el **label 1** a los jets que vienen del top y **0** para los que vienen fondo
- Empezaréis con una red *fully connected*
  - *El notebook os guiará para entender y visualizar las imágenes de los jets y para evaluar el resultado (ROC curve)*
  - *También encontraréis alguna pista para mejorar los resultados*

## TOP TAGGING 2

- Forma del detector CMS → Un cilindro.
- La superficie cilíndrica puede ser “desenrollada” dando lugar a una superficie rectangular que puede ser dividida en píxeles”
- Los depósitos de energía de las partículas pueden convertirse en intensidades de color dentro de cada pixel
- Cuanto más energía tiene una partícula, más intensidad de color en un pixel particular
- Trabajaremos en blanco y negro

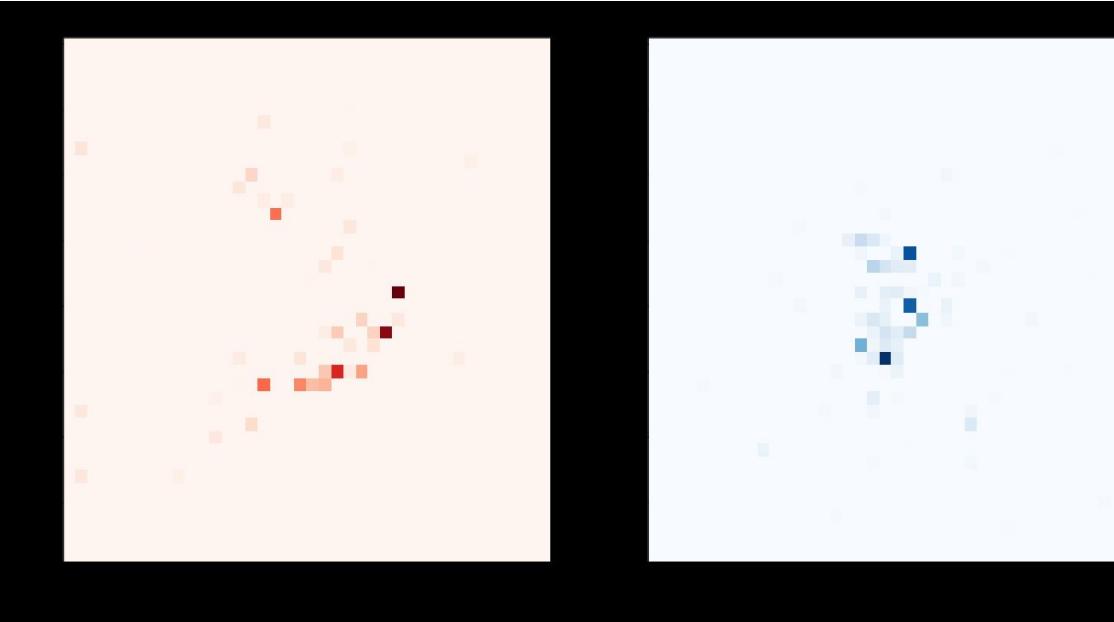


## TOP TAGGING 2



- Los depósitos de energía de los constituyentes de los jets se transforman en “intensidades” de pixel de una imagen 2D en blanco y negro
- Así podemos aplicar algoritmos de reconocimiento de imágenes a problemas de altas energías!

## TOP TAGGING 2



- ¿Qué pinta tiene la imagen de un jet?
- ¿Puedes decir cual es señal y cual es fondo?
- ¡No es fácil!

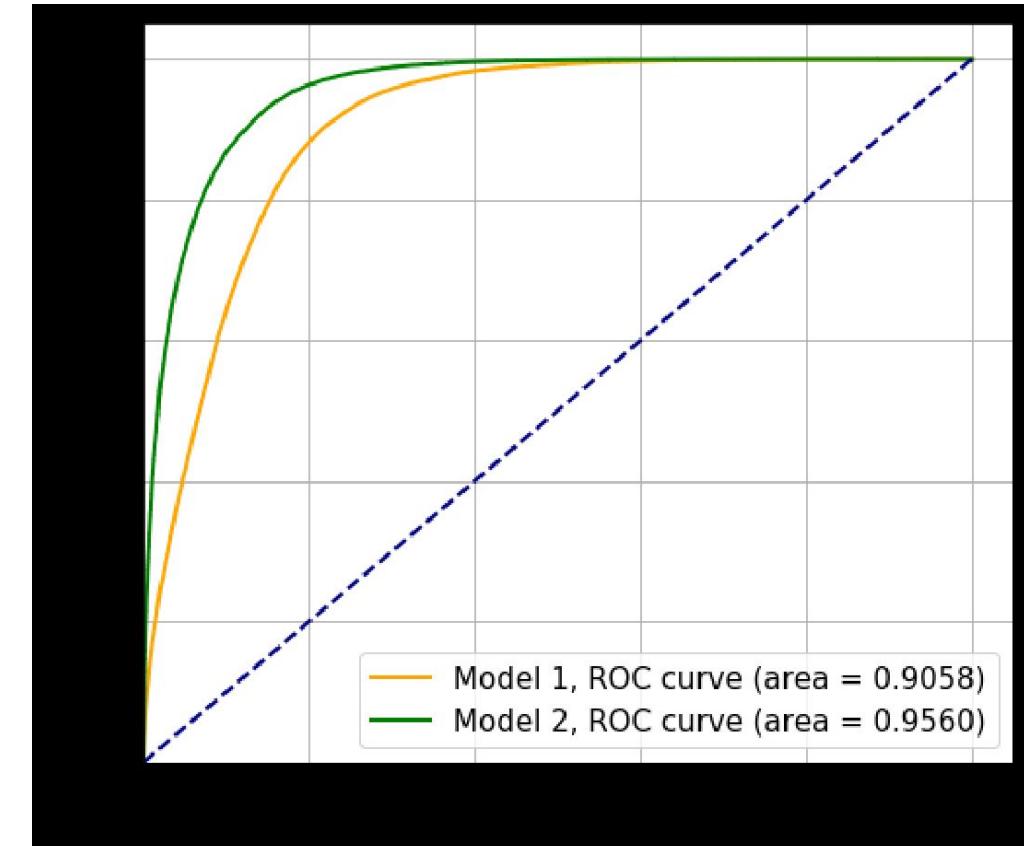
## TOP TAGGING 2

- El cuadrimiento de las partículas que han sido clusterizadas dentro de un jet son transformadas en **imágenes de 40x40 pixels**.
- El contenido de estos 16000 pixels se guarda como columnas en un pandas frame.
- Se le asigna el **label 1 a los jets que vienen del top y 0 para los que vienen fondo**
- Tendréis que utilizar **redes convolucionales** y conceptos más avanzados de Deep Learning
  - *El notebook os guiará para entender y visualizar las imágenes de los jets y para evaluar el resultado (ROC curve)*
  - *También encontraréis alguna pista para mejorar los resultados*

# EVALUANDO EL RESULTADO

Medida del rendimiento para clasificación binaria: Receiver Operating Characteristic curve (**ROC curve**):

- Compara cuantas veces la red predice señal cuando el resultado es señal (TPR) vs cuantas veces la red predice señal cuando el resultado es fondo (FPR)
- Cuanto mayor sea el área bajo la curva, mejor será el resultado



# DATASET PÚBLICO Y RESULTADO

- Los datos utilizados en este ejercicio se encuentran disponibles publicamente aqui:

<https://goo.gl/XGYju3>

- Se están utilizando para comparar distintos algoritmos de “top tagging” □
- Estaréis trabajando en un problema real de Machine Learning
- Si obtenéis un AUC mayor de 0.98...se puede publicar!

	AUC	Acc	1/ $\epsilon_B$ ( $\epsilon_S = 0.3$ )			#Param
			single	mean	median	
CNN [16]	0.981	0.930	914±14	995±15	975±18	610k
ResNeXt [30]	0.984	0.936	1122±47	1270±28	1286±31	1.46M
TopoDNN [18]	0.972	0.916	295±5	382± 5	378 ± 8	59k
Multi-body $N$ -subjettiness 6 [24]	0.979	0.922	792±18	798±12	808±13	57k
Multi-body $N$ -subjettiness 8 [24]	0.981	0.929	867±15	918±20	926±18	58k
TreeNiN [43]	0.982	0.933	1025±11	1202±23	1188±24	34k
P-CNN	0.980	0.930	732±24	845±13	834±14	348k
ParticleNet [47]	0.985	0.938	1298±46	1412±45	1393±41	498k
LBN [19]	0.981	0.931	836±17	859±67	966±20	705k
LoLa [22]	0.980	0.929	722±17	768±11	765±11	127k
Energy Flow Polynomials [21]	0.980	0.932	384			1k
Energy Flow Network [23]	0.979	0.927	633±31	729±13	726±11	82k
Particle Flow Network [23]	0.982	0.932	891±18	1063±21	1052±29	82k
GoaT	0.985	0.939	1368±140		1549±208	35k

<https://arxiv.org/pdf/1902.09914.pdf>

# DATATHON

- Cada estudiante participará de **forma individual**.
- Tendréis que presentar vuestra propuesta para **resolver el problema de top tagging**:
  - Podéis presentar hasta **3 propuestas como máximo**.
  - Podéis elegir tomar el punto de partida que prefiráis, top tagging 1 (números) o 2 (imágenes)
- El ganador es el que obtenga el **mejor AUC** sobre la muestra de test.
- En los notebooks encontrareis unas líneas (al final) que preparan un zip con el output en el formato que lo tenéis que presentar
- Poned nombres significativos a los archivos zip. **El mejor de los tres** será vuestro resultado final.

# DATATHON

- El **plazo máximo** para entregar las propuestas será el **Jueves 30**.
- Hasta ese tiempo podréis escribirnos para preguntarnos las dudas que tengáis.
- Si lo necesitáis, podemos organizar **tutorías** en horario de clase vía GoToMeeting.

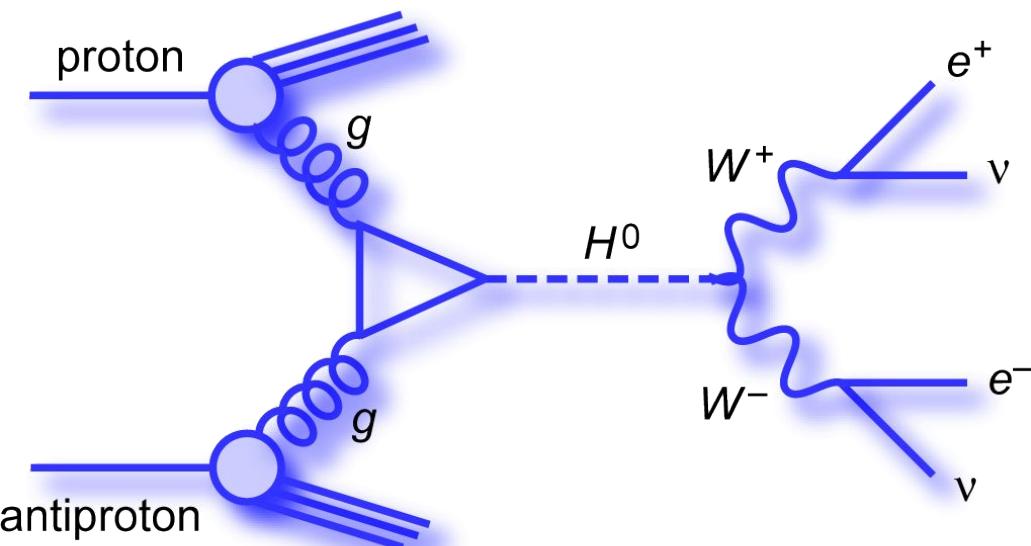
# A POR ELLO!

Los dos notebooks con los dos métodos, Top Tagging 1 y Top Tagging 2, pueden encontrarse aquí:

[https://github.com/CeliaFernandez/TopTagging\\_2019-2020](https://github.com/CeliaFernandez/TopTagging_2019-2020)

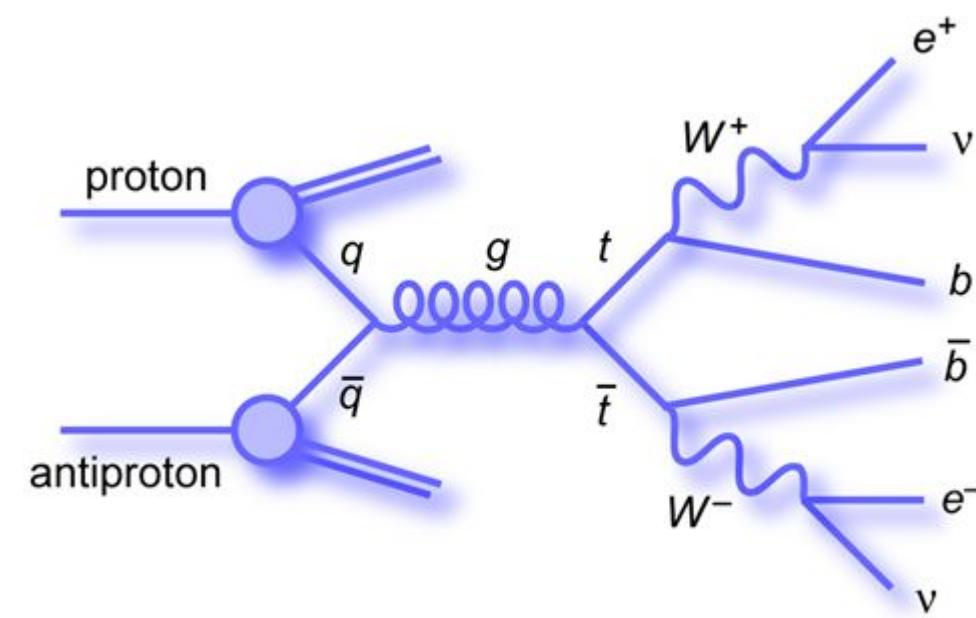
Recomendamos utilizar DataScienceHub  
(Todos los paquetes están instalados y los datos  
descargados en /home/jovyan/share/TopTaggingData/)

## HIGGS



• Estado final:  $e^+$ ,  $e^-$ ,  $2\nu$ , jets

## TOP-ANTITOP



• Estado final:  $e^+$ ,  $e^-$ ,  $2\nu$ , jets