



# **ESCUELA POLITÉCNICA NACIONAL**

## **FACULTAD DE CIENCIAS**

**ANÁLISIS SIMPLIFICADO DE PROCESOS DE ALTA ENERGÍA CON  
DATOS ABIERTOS DEL COMPACT MUON SOLENOID (CMS)**

**ESTUDIO DE LA PRODUCCIÓN DE CUATRO QUARKS  
TOP EN COLISIONES PROTÓN-PROTÓN EN PROCESOS  
CON UN ESTADO FINAL COMPUESTO POR UN MUÓN Y  
JETS, UTILIZANDO DATOS ABIERTOS DE LA  
COLABORACIÓN CMS.**

**TRABAJO DE INTEGRACIÓN CURRICULAR PRESENTADO COMO  
REQUISITO PARA LA OBTENCIÓN DEL TÍTULO DE FÍSICA**

**PAMELA CAROLINA LLERENA DELGADO**

[pamela.llerena@epn.edu.ec](mailto:pamela.llerena@epn.edu.ec)

**DIRECTOR: EDY RODRIGO AYALA AMAYA**

[edy.ayala@epn.edu.ec](mailto:edy.ayala@epn.edu.ec)

**DMQ, FEBRERO 2024**



## **CERTIFICACIONES**

Yo, PAMELA CAROLINA LLERENA DELGADO, declaro que el trabajo de integración curricular aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

---

Pamela Carolina Llerena Delgado

Certifico que el presente trabajo de integración curricular fue desarrollado por Pamela Carolina Llerena Delgado, bajo mi supervisión.

---

Edy Rodrigo Ayala Amaya  
**DIRECTOR**



## **DECLARACIÓN DE AUTORÍA**

A través de la presente declaración, afirmamos que el trabajo de integración curricular aquí descrito, así como el(los) producto(s) resultante(s) del mismo, es(son) público(s) y estará(n) a disposición de la comunidad a través del repositorio institucional de la Escuela Politécnica Nacional; sin embargo, la titularidad de los derechos patrimoniales nos corresponde a los autores que hemos contribuido en el desarrollo del presente trabajo; observando para el efecto las disposiciones establecidas por el órgano competente en propiedad intelectual, la normativa interna y demás normas.

Pamela Carolina Llerena Delgado

Edy Rodrigo Ayala Amaya



## **AGRADECIMIENTOS**

Le agradezco a mi familia, a mis padres y abuelos, por su gran apoyo durante todos estos años para poder alcanzar esta meta. En especial a las mujeres de mi vida que tanto me han inspirado: mi Nona, mi tía Mónica, y mi mamá Lorena. Gracias por su incondicional apoyo, por todo su amor y su confianza en mí. A mi papá, por sus consejos y palabras de aliento.

A mis amigos que me han acompañado durante todos estos años, haciendo mi etapa universitaria una de las más bonitas: Eugenia, Alejito, Cristian, Jack, Xime, Joyce, Tati, Alejo, Bry, Germán, Antonio, y David.

A Edy Ayala y Edgar Carrera, por su apoyo y constante guía, los cuáles fueron indispensables para la realización de este proyecto. Les agradezco por compartirme su sabiduría y abrirme las puertas hacia nuevas oportunidades.

A Alejandro Gómez, por su colaboración en el desarrollo de este proyecto.

A mis profesores de la carrera que me han inspirado y enseñado tanto: Ramón, Esteban, y Álvaro. Les agradezco por formar parte de mi camino universitario y por impartirme los conocimientos que me permiten estar aquí hoy.

A Tommy, por haberme acompañado en este largo camino y por alentarme a seguir adelante en los momentos más difíciles.

## RESUMEN

Este trabajo presenta el estudio de la búsqueda de la producción de cuatro quarks top, con el objetivo de reproducir de manera simplificada el proceso de análisis realizado por la colaboración CMS. El estudio se centra en la producción de la señal de cuatro tops, en el canal leptónico con un estado final con un muon y jets, a una energía de centro de masa de 13 TeV y una luminosidad de  $2.26 \text{ fb}^{-1}$ . Para ello se utilizaron los datos abiertos del CERN, que incluyen datos de las colisiones del LHC y simulaciones Monte Carlo. Los datos se procesaron utilizando códigos en Python y el software CMSSW, obteniendo datasets en formato NanoAOD. Se aplicaron los filtros necesarios para la selección de eventos de la señal mientras se suprimen los eventos de fondo. Los histogramas de las variables físicas estudiadas mostraron concordancia entre los datos de las colisiones y la simulación. La discriminación entre señal y procesos de fondo se llevó a cabo mediante un análisis multivariante utilizando la técnica Boosted Decision Trees (BDT). La evaluación del algoritmo a través de la curva ROC reveló un desempeño sólido en la tarea de discriminación. En los datos reales, no se encontró evidencia de eventos con la señal de cuatro quarks top, corroborando los hallazgos de la publicación original. Este estudio contribuye a comprender el proceso que se realiza en análisis dentro de la Física de Partículas.

**Palabras clave:** Cuatro quarks top, canal leptónico, CERN, análisis multivariante, Boosted Decision Trees (BDT), Física de Partículas.



## ABSTRACT

This work presents a study on the search for the production of four top quarks, with the aim of reproducing in a simplified manner the analysis process carried out by the CMS collaboration. The study focuses on the production of the four tops signal in the single lepton plus jets channel with a muon in the final state, at a center-of-mass energy of 13 TeV and a luminosity of  $2.26 \text{ fb}^{-1}$ . CERN Open data, including collision data and Monte Carlo simulations, were used for this purpose. The data were processed using Python codes and the CMSSW software, generating NanoAOD format datasets with the expected number of events. Necessary filters were applied for signal event selection while suppressing background events. Histograms of the studied physical variables exhibited agreement between collision data and simulation. Discrimination between signal and background processes was carried out through a multivariate analysis using the Boosted Decision Trees (BDT) technique. The algorithm's performance, assessed through the ROC curve, revealed good discrimination capabilities. No evidence of events with the four top quarks signal in the collision data was found, consistent with the findings reported in the original publication. This study contributes to understanding the processes involved in Particle Physics analyses.

**Keywords:** Four top quarks, leptonic channel, CERN, multivariate analysis, Boosted Decision Trees (BDT), Particle Physics.

---

# Índice general

---

<b>1. Descripción del componente desarrollado</b>	<b>1</b>
1.1. Objetivo general . . . . .	2
1.2. Objetivos específicos . . . . .	2
1.3. Alcance . . . . .	3
1.4. Marco Teórico . . . . .	4
1.4.1. El Modelo Estándar . . . . .	4
1.4.2. Formalismo del SM . . . . .	8
1.4.3. Limitaciones del Modelo Estándar . . . . .	10
1.4.4. El Quark Top . . . . .	11
1.4.5. Producción de cuatro quarks top . . . . .	12
<b>2. Metodología</b>	<b>15</b>
2.1. El Gran Colisionador de Hadrones . . . . .	15
2.1.1. Medición de la Colisiones . . . . .	16
2.2. El Detector CMS . . . . .	18
2.2.1. Subdetectores . . . . .	19
2.2.2. Trigger . . . . .	19
2.3. Simulación y Reconstrucción de objetos . . . . .	20
2.3.1. Reconstrucción de objetos . . . . .	20
2.3.2. CMS Open Data . . . . .	23

2.4. Procesos de señal y background . . . . .	23
2.4.1. Background del proceso $t\bar{t}$ . . . . .	24
2.4.2. Background del proceso Electro-débil . . . . .	24
2.5. Set de Datos . . . . .	24
2.6. Archivos NanoAOD . . . . .	25
2.6.1. CMSSW . . . . .	25
2.6.2. Procesamiento de datos . . . . .	26
2.6.3. Archivo JSON . . . . .	28
2.7. Selección de Eventos . . . . .	29
2.8. Análisis Multivariante . . . . .	30
2.8.1. Método Boosted Decision Trees . . . . .	31
<b>3. Resultados, conclusiones y recomendaciones</b>	<b>33</b>
3.1. Selección de Eventos . . . . .	33
3.2. Variables Discriminatorias . . . . .	35
3.3. Boosted Decision Trees . . . . .	36
3.4. Conclusiones y recomendaciones . . . . .	42
<b>Anexos</b>	<b>44</b>
<b>Bibliografía</b>	<b>46</b>

---

## Índice de figuras

---

1.1. Escala atómica de la naturaleza. . . . .	6
1.2. Modelo Estándar de Física de Partículas. . . . .	7
1.3. Diagramas de Feynman de distintos procesos dentro del Modelo Estándar . . . . .	10
1.4. Diagrama de Feynman del decaimiento de un quark Top a un bosón W y a un quark b. El bosón W puede desintegrarse ya sea leptónica ( $l^+, \nu$ ) o hadrónicamente ( $q, \bar{q}$ ). . . . .	12
1.5. Diagrama de un evento de colisión p-p, en la producción de cuatro quarks top $t\bar{t}t\bar{t}$ . . . . .	13
2.1. Esquema del LHC. . . . .	16
2.2. Sistema de coordenadas del CMS, incluido el LHC. . . . .	17
2.3. Esquema transversal del detector CMS del LHC. . . . .	18
2.4. Diagrama de Feynman de la producción de cuatro quarks top en el canal leptónico. . . . .	23
2.5. Diagrama del procesamiento de datos. . . . .	28
2.6. Esquema de un Decision Tree. . . . .	31
3.1. Distribuciones de $p_T$ para datos y simulación para el canal muón+jets. . . . .	35
3.2. Distribuciones $N_j$ para datos y simulación. . . . .	37
3.3. Distribuciones $H_t^b$ para datos y simulación. . . . .	37

3.4. Distribuciones $N_{bj}$ para datos y simulación. . . . .	38
3.5. Distribución de la señal y background de la simulación Monte Carlo entre las variables $N_{bj}$ y $N_j$ . . . . .	38
3.6. Distribución de la señal y background de la simulación Monte Carlo entre las variables $H_T^{RAT}$ y $H_T^b$ . . . . .	39
3.7. Distribución de la señal y background de la simulación Monte Carlo entre las variables $H_T^{RAT}$ y $N_j$ . . . . .	39
3.8. Curva ROC para los algoritmos XGBoost y Adaptive Booster.	40
3.9. Densidad de eventos para las muestras de Monte Carlo $MC$ y datos de las colisiones <i>Real Data</i> , en la probabilidad de que sean procesos de señal (1) o de fondo (0). . . . .	41

# Capítulo 1

---

## Descripción del componente desarrollado

---

Los experimentos de física de partículas están diseñados para detectar e identificar partículas producidas en colisiones de dos partículas de alta energía. El Gran Colisionador de Hadrones del CERN (LHC) (actualmente el colisionador de partículas más grande jamás construido) colisiona protones a casi la velocidad de la luz. Las partículas resultantes de las colisiones se detectan en puntos específicos, en donde se ubican detectores como el Compact Muon Solenoid (CMS), el cual tiene el potencial de observar cualquier fenómeno físico nuevo que pueda revelar el LHC [1][2].

El quark top es la partícula elemental más pesada conocida dentro del Modelo Estándar. Tiene el acople más grande con el bosón de Higgs, y se espera que esté estrechamente relacionado con posibles nuevas partículas hipotéticas en modelos más allá del Modelo Estándar. En este sentido, los procesos que involucran a los quarks top son particularmente importantes en esta línea de investigación [3].

En las colisiones protón-protón de alta energía en el LHC, la mayoría de los quarks top se producen en pares  $t\bar{t}$  a través de los canales  $q\bar{q} \rightarrow t\bar{t}$  (10%),  $g\bar{g} \rightarrow t\bar{t}$  (90%) con una pequeña contribución de la producción de un solo quark top. Sin embargo, también es posible producir cuatro quarks top  $t\bar{t}t\bar{t}$  dentro del Modelo Estándar en colisiones gluón-gluón descritas por diagramas de orden superior en cromodinámica cuántica. Su producción es un proceso estudiado dentro de los muchos procesos en el LHC [4].

Cada uno de estos quarks top decae en un quark bottom y un bosón W, y cada bosón W puede decaer en un leptón cargado (electrón, muón, tau) y uno neutro (neutrino electrón, neutrino mu, neutrino tau). Como resultado, se tiene un estado final con un chorro de partículas (jets), generados por la hadronización del quark bottom, y leptones provenientes del decaimiento del bosón W.

En este estudio se propone replicar el procedimiento de tratamiento y análisis de datos del estudio realizado por la colaboración CMS en la publicación “*Search for production of four top quarks in final states with same-sign or multiple leptons in proton-proton collisions at  $\sqrt{s} = 13$  TeV*” [5].

## 1.1. Objetivo general

Realizar un análisis de los datos obtenidos en el detector multipropósito CMS del LHC, en busca de procesos que pueden llevar a nueva física más allá del modelo estándar. Este análisis se llevará a cabo mediante la reproducción simplificada del análisis realizado por la colaboración CMS en el proceso de producción de cuatro quarks top en la colisión protón-protón con energía de centro de masa de 13 TeV y luminosidad  $2.26 \text{ fb}^{-1}$ . En el estado final, uno de los quarks top decae, dando lugar a la presencia de un muón [5][6].

## 1.2. Objetivos específicos

1. Realizar el procesamiento de los datos con los cuales se va a trabajar, del formato MiniAOD al formato NanoAOD.
2. Filtrar los archivos de datos generales con la selección específica para el proceso de  $t\bar{t}t\bar{t}$  y estudiar variables adicionales para su posterior discriminación.
3. Comparar los datos reales obtenidos en la colisión entre protones con la simulación de Monte Carlo correspondiente.
4. Realizar el análisis multivariable para diferenciar e incrementar la

significancia de la señal  $t\bar{t}t\bar{t}$  con respecto a los procesos de background o proceso de fondo.

### 1.3. Alcance

Se busca reproducir los resultados del artículo científico [6] mediante el uso de datos abiertos de la colaboración CMS, dentro del plazo de tiempo establecido para el Trabajo de Integración Curricular. El estudio consiste en la búsqueda de la producción de cuatro quarks top ( $t\bar{t}t\bar{t}$ ) mediante el análisis de los datos de colisión entre protones con centros de masa de  $\sqrt{s} = 13$  TeV, y una luminosidad de  $2.26 \text{ fb}^{-1}$ . Para este proyecto, se utilizaron los datos de las colisiones protón-protón en el canal semi-leptónico con un muón+jets, que están disponibles en el portal de datos abiertos de CMS.

El procesamiento de los datos del formato MiniAOD al formato nanoAOD se realiza mediante códigos en Python, utilizando el software CMSSW y las herramientas computacionales respectivas. Para llevar a cabo el análisis, se utilizan únicamente los datos que contienen información sobre muones en el proceso de producción de cuatro quarks top. Estos datos son filtrados y procesados para seleccionar solo aquellos que cumplen con los criterios de selección específicos para el proceso  $t\bar{t}t\bar{t}$ , de acuerdo a la publicación que se emplea como una guía del estudio.

Se utilizan muestras simuladas de Monte Carlo para evaluar la aceptación del proceso  $t\bar{t}t\bar{t}$ , mediante un análisis multivariante entre los datos experimentales y simulados. Esto implica comparar las distribuciones de las variables de interés entre los datos reales y las simulaciones usando técnicas de machine learning. Se utilizan métricas específicas, como la curva ROC (Receiver Operating Characteristic) y el área bajo la curva (AUC), para cuantificar la eficacia del modelo.

El trabajo finalizado se va a compartir públicamente a través del repositorio de código abierto en Github, permitiendo su accesibilidad y uso por parte de quienes quieran emplear el uso de los datos abiertos del CERN.



## 1.4. Marco Teórico

La física de partículas es la ciencia que permite estudiar los componentes más pequeños que conforman la materia. Aquellos componentes son conocidos como *partículas elementales*, y su estudio hace posible la comprensión de la estructura del universo en su nivel más fundamental. El *Modelo Estándar* (SM) es la mejor teoría desarrollada para describir a estas partículas y sus interacciones, puesto que ha demostrado ser compatible con los datos experimentales obtenidos hasta la actualidad. Sin embargo, aún quedan muchas preguntas sin resolver, motivando la creación de modelos más allá del estándar.

La investigación en este campo se realiza con experimentos en aceleradores de partículas, como el Gran Colisionador de Hadrones, ubicado en la Organización Europea para la Investigación Nuclear (CERN). Luego del exitoso descubrimiento del Bosón de Higgs, interrogantes como: *¿Qué es la materia y energía oscura? ¿Cómo incluir a la gravedad en el modelo estándar?*, hacen que la investigación dentro de este campo continúe.

Las propiedades del quark top lo convierten en la partícula ideal para investigar fenómenos más allá del Modelo Estándar en la búsqueda de nueva física, particularmente en la producción de cuatro quarks top. Este fenómeno es de gran interés en la física de partículas. Aunque su producción está descrita dentro del Modelo Estándar, es extremadamente rara. Por estas razones, este proceso analiza en varios experimentos del LHC. La búsqueda de su producción es el objetivo principal del siguiente estudio.

### 1.4.1. El Modelo Estándar

El SM contiene tres tipos de partículas: leptones, quarks y bosones. Los leptones y quarks pertenecen a la familia de los fermiones, caracterizados por tener espín semi entero  $1/2$ . Están organizados en tres generaciones, cada una de las cuales contiene partículas con mayor masa que la anterior. Las propiedades de los leptones y quarks están expuestas en el Cuadro 1.1 [7].

Los leptones cargados, el electrón  $e^-$ , muón  $\mu$  y tau  $\tau$ , tienen masas

considerablemente mayores en comparación con sus leptones no cargados o neutrinos correspondientes: neutrino electrón  $\nu_{e^-}$ , neutrino muón  $\nu_{\mu}$ , y neutrino tau  $\nu_{\tau}$ . Además, tienen carga eléctrica igual a -1, mientras que los neutrinos, por otra parte, tienen carga cero.

La familia de los quarks está conformada por seis partículas, los quarks up  $u$ , down  $d$ , charm  $c$ , strange  $s$ , top  $t$  y bottom  $b$ . Además de la propiedad de carga eléctrica de +2/3 (para  $u$ ,  $c$  y  $t$ ) o -1/3 (para  $d$ ,  $s$  y  $b$ ), poseen una propiedad adicional conocida como carga de color, y viene en tres variedades denominadas 'red', 'blue' y 'green' en inglés.

Los quarks up  $u$  y down  $d$  forman la estructura del protón y neutrón dentro del núcleo atómico. Junto con el leptón electrón  $e^-$  forman los átomos, los cuales, a su vez, constituyen las moléculas que componen toda la materia existente, como se puede observar en la Figura 1.1.

El modelo también predice a sus antipartículas, es decir, seis anti-leptones y seis antiquarks, todos con cargas opuestas a sus partículas correspondientes. Los valores de masa de mantienen iguales.

Las partículas fundamentales forman la materia observada debido a la manera en que interactúan. Lo hacen a través de las cuatro fuerzas fundamentales de la naturaleza: la interacción electromagnética, la interacción nuclear débil, interacción nuclear fuerte y la interacción gravitacional.

Tanto los leptones como los quarks interactúan débilmente, la interacción débil cargada cambia el sabor de los quarks (el tipo de quarks, así por ejemplo  $d \rightarrow u$ ). Los leptones cargados y los quarks interactúan electro magnéticamente puesto que poseen carga de eléctrica. Los quarks interactúan fuertemente a través de su carga de color, esta interacción tiene la propiedad de "confinamiento" que impide aislar un quark o anti-quark y solo se observará sistemas de quarks, principalmente  $qqq$  y  $q\bar{q}$ . Esta interacción es la que mantiene unido a los protones y neutrones en el núcleo atómico. El cuadro 1.3 muestra las interacciones que sufren cada una de las partículas.

La fuerza gravitacional es la responsable de la atracción entre objetos con masa. Sin embargo, su magnitud es muy débil en la interacción entre partículas, por lo cual su influencia se considera despreciable. Esta interacción no es parte del Modelo Estándar.

Fermiones	Leptones			Quarks		
Generación	Partícula	Masa ( $MeV$ )	Carga	Partícula	Masa ( $MeV$ )	Carga
I	$e^-$	0.501	-1	$u$	$2.2^{+0.6}_{-0.4}$	+2/3
	$\nu_{e^-}$	$< 2.05 \cdot 10^{-6}$	0	$d$	$4.7^{+0.5}_{-0.4}$	-1/3
II	$\mu$	105.65	-1	$c$	$1.27 \pm 0,03 \cdot 10^3$	+2/3
	$\nu_{\mu}$	$< 0.19$	0	$s$	$96^{+8}_{-4}$	-1/3
III	$\tau$	$1776.6 \pm 0.12$	-1	$t$	$173.21 \pm 0.51 \cdot 10^3$	+2/3
	$\nu_{\tau}$	$< 18.2$	0	$b$	$4.18^{+0.04}_{-0.03}$	-1/3

Cuadro 1.1: Familia de los Fermiones: Leptones y Quarks [8].

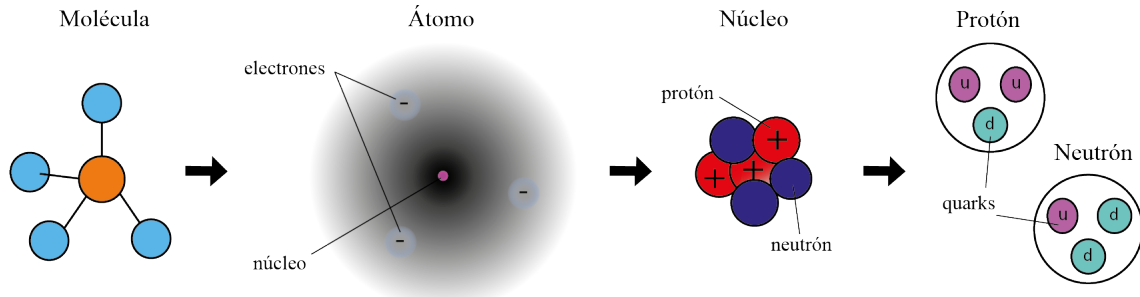


Figura 1.1: Escala atómica de la naturaleza.

Cuando las partículas interactúan entre sí, lo hacen a través del intercambio de partículas mediadoras, los *bosones de gauge*. Una interacción se interpreta como un intercambio de partículas virtuales. Los bosones tienen espín 1, y a cada fuerza de interacción le corresponde cierto tipo de bosón, como se expone en el Cuadro 1.2.

Para la interacción electromagnética, la partícula de intercambio es el fotón  $\gamma$ . En el caso de la interacción fuerte, las partículas portadoras de fuerza son los gluones  $g$ , y como los fotones no son masivas. La interacción débil se da mediante el intercambio de los bosones  $W^+, W^-$  y  $Z$ . Estos bosones, a diferencia de los anteriores, si tienen masa cuyos valores están entre los 80 y 90  $GeV$ .

Partícula medidora (boson)	Interacción	Masa ( $GeV$ )	Espín
fotón $\gamma$	Electromagnética	0	1
$W^\pm$	Nuclear débil	$30.385 \pm 0.015$	1
$Z$	Nuclear débil	$91.1876 \pm 0.0021$	1
gluón $g$	Nuclear fuerte	0	1

Cuadro 1.2: Las tres fuerzas fundamentales y sus bosones correspondientes.

Interacción	Electromagnética	Nuclear débil	Nuclear fuerte
<b>Fermiones</b>			
$e^-$ , $\mu$ , $\tau$	✓	✓	
$\nu_{e^-}$ , $\nu_{\mu}$ , $\nu_{\tau}$		✓	
Quarks	✓	✓	✓

Cuadro 1.3: Interacciones que experimentan las partículas del Modelo Estándar.

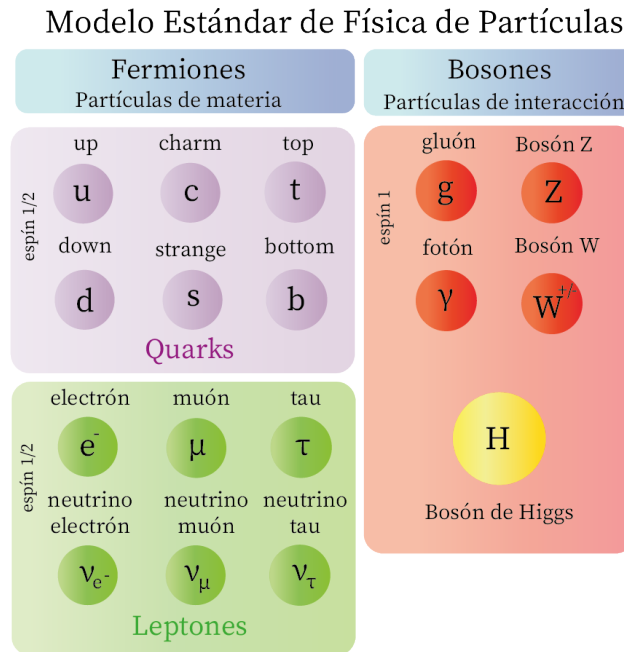


Figura 1.2: Modelo Estándar de Física de Partículas.

Finalmente, dentro del SM está el bosón de Higgs. Es el único bosón escalar o con espín cero. Es una de las partículas más importantes dentro del SM, ya que la masa de cada partícula fundamental está directamente relacionada con su interacción con el campo de Higgs. A mayor interacción, mayor masa. Mientras que gluones y fotones no interactúan con este campo y carecen de masa, leptones, quarks y bosones  $W^+$ ,  $W^-$  y  $Z$  sí lo hacen, lo que contribuye a la diversidad de masas observadas. El esquema completo del Modelo Estándar se lo puede simplificar al mostrado en la Figura 1.2.

El bosón de Higgs fue descubierto en el año 2012, en el Consejo Europeo para la Investigación Nuclear. En este centro se encuentra El Gran Colisionador de Hadrones, el acelerador de partículas más potente del mundo, el cuál fue fundamental en el descubrimiento de la partícula de Higgs, y hoy en día es fundamental en la investigación experimental de la

física más allá del MS, y en particular en el proyecto que se desarrolla en este trabajo [9].

### 1.4.2. Formalismo del SM

Hasta ahora, se ha explorado de una manera descriptiva cómo El Modelo Estándar involucra tres de las cuatro fuerzas fundamentales: la nuclear débil, la nuclear fuerte y la electromagnética. El Modelo Estándar se basa en la unión de teorías gauge, la Cromodinámica Cuántica (QCD)  $SU(3)$  de color y la Teoría Electro débil (EW)  $SU(2) \otimes SU(1)$ , las cuales son teorías cuánticas de campo.

La primera pieza en la construcción del modelo fue la electrodinámica cuántica, la cual describe las interacciones de los todos los leptones cargados, incluyendo muones, con el campo electromagnético. Posteriormente, esta teoría se generalizó a una teoría de las interacciones débiles con vértices vectorial y pseudo vectorial (V-A). Luego se desarrolló la teoría de gauge débil  $SU(2)$  y se la unificó con la electrodinámica cuántica  $U(1)$  para formar la teoría Electro débil, que describe las interacciones electromagnéticas y nucleares débiles. Los bosones  $W^{+-}$  y  $Z$  median la interacción débil, mientras que el fotón media la interacción electromagnética [10].

El modelo asocia a cada partícula e interacción campos cuánticos, que deben ser representaciones del grupo propio de Lorentz para que sean consistentes con la relatividad especial. A las partículas se les asocia campos de Dirac,  $\psi$ ; a las interacciones se les asocia campos vectoriales  $B^\mu$ , para cada uno de los cuales se define  $F^{\mu\nu} = \partial^\mu B^\nu - \partial^\nu B^\mu$ . Además se introduce un campo escalar  $\phi$ , encargado de asignar masa a las partículas [11].

El lagrangiano del modelo estándar se lo puede escribir de una manera reducida como:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}D\psi + h.c. + \bar{\psi}_i y_{ij} \psi_j \phi + h.c. + |D_\mu \phi|^2 - V(\phi)$$

donde el primer término del lado derecho representa todos los campos vectoriales que describen los bosones de interacción: fotón, bosones  $W^+$ ,

$Z$  y 8 Gluones. El segundo término contiene todos los términos de las partículas libre y los términos de interacción (con las tres interacciones); las siglas *h.c.* involucran los términos hermitiano-conjugado. El cuarto y quinto término involucran los términos que asignan masa a las partículas. El sexto término involucra la asignación de masa a los bosones pesados  $W^{+-}$ ,  $Z$  por el mecanismo de Higgs. El último término que considera la interacción del bosón de Higgs consigo mismo. Cabe mencionar que del lagrangiano desarrollado se puede inferir las reglas de Feynman y con ello describir cuantitativamente, hasta cierto grado de aproximación, cualquier proceso de interacción entre partículas.

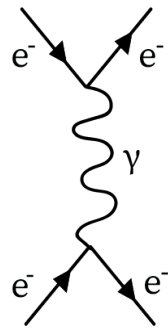
Un ejemplo de la interacción electromagnética, es la interacción coulombica entre dos electrones. En este proceso la interacción coulombica como el intercambio de *fotones virtuales*. Una partícula virtual es aquella que aparece temporalmente (en tiempos menores a los permitidos por el principio de indeterminación) en todo proceso de interacción, pero que nunca se los pondrá observar directamente en experimentos.

En la Figura 1.3a, se expone una gráfica de la interacción electromagnética entre dos electrones. Esta representación visual de la interacción entre partículas es lo que se conoce como Diagrama de Feynman, utilizados de gran manera dentro de la física de partículas. La flecha del tiempo se ha escogido de izquierda a derecha. Cualquier proceso de interacción puede describirse de una manera pictórica a través de un diagrama de Feynman (que no solo da la posibilidad de una representación gráfica del proceso; sino que además, es una herramienta poderosa para el cálculo de la sección eficaz del proceso) [12].

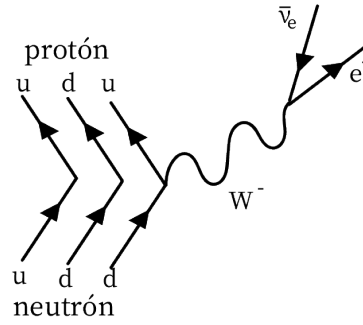
Esto es lo que se conoce como un proceso electromagnético. En general, un *proceso* es una interacción o cambio que se da entre partículas. Para el trabajo presente, se utilizará la convención de que las partículas en el estado inicial se muestran a la izquierda y las partículas en el estado final se muestran a la derecha.

El decaimiento de partículas un tipo particular de reacción que involucra la transformación de partículas en otra más ligeras. En la Figura 1.3b se observa el decaimiento beta de un neutrón. Un quark down del neutrón emite un bosón  $W^-$ , transformándose en un quark up. El bosón  $W^-$  decae posteriormente en un electrón  $e^-$  y antineutrino tipo electrón

$\bar{\nu}_{e^-}$ .



(a) Interacción entre electrones mediante el intercambio de un fotón.



(b) Decaimiento beta de un neutrón.

Figura 1.3: Diagramas de Feynman de distintos tipos interacción o procesos dentro del Modelo Estándar.

### 1.4.3. Limitaciones del Modelo Estándar

El SM es una teoría exitosa que puede predecir y describir procesos físicos que son observados experimentalmente. Sin embargo, no es una teoría completa, puesto que tiene sus limitaciones y aún existen interrogantes por responder, algunas de las cuáles se detallan a continuación. Estas interrogantes indican que el SM puede ser una aproximación de una teoría subyacente más fundamental [13][14].

- El estudio de la cosmología permitió exponer una de las primeras interrogantes. Los datos del satélite Planck muestran que la materia “ordinaria”, que comprende átomos y moléculas, representa sólo el 4.9% del contenido energético del Universo. Otro 26.8% es lo que se conoce como *materia oscura*, la cuál no ha sido detectada a través de interacciones fuertes o electromagnéticas. La combinación de materia ordinaria y materia oscura, solo constituyen el 31.7% del contenido total. El porcentaje restante de 68.3% se atribuye a la *energía oscura*, la cuál se cree que influye en la expansión acelerada del Universo. El SM no ofrece ninguna partícula candidata para la materia oscura y tampoco proporciona una explicación para la energía oscura.

- El SM no considera la gravedad. Eso limita la validación del modelo a escalas de energía donde la gravedad es insignificante en comparación con las otras interacciones. Los efectos de la gravedad cuántica se vuelven importantes a energías muy altas. Esta escala se cuantifica en primera aproximación mediante la escala de Planck de  $\approx 10^{19}$  eV, más allá de la cual se espera que el SM ya no sea válido.
- El SM no proporciona una explicación fundamental sobre las correcciones radiativas de lazo fermiónico en la determinación teórica de la masa de la partícula de Higgs, requiriéndose una sintonización fina (fine tuning) de dichas correcciones para que se eliminen esas correcciones idénticamente.

Estas y otras limitaciones del SM impulsan la búsqueda de modelos más allá del Modelo Estándar, lo cual se conoce como *Nueva Física*. Las propiedades del quark top lo convierten en una herramienta de gran utilidad para investigar dentro de este contexto.

#### 1.4.4. El Quark Top

El quark top  $t$  es el más pesado de todas las partículas fundamentales. Tiene una masa de  $173.2 \pm 0.5$  GeV, que está cerca de la escala de la ruptura de simetría electro-débil. Su gran masa indica por lo tanto un acoplamiento grande con el bosón de Higgs. El quark top decae principalmente<sup>1</sup> en un quark bottom  $b$  y un bosón  $W$ , con el bosón  $W$  desintegrándose posteriormente en otras partículas, como se puede observar en la Figura 1.4. El quark top tiene un tiempo de vida corto, de alrededor  $5 \times 10^{-25}$  segundos, lo cual permite que decaiga antes de su hadronización. Lo que quiere decir que no existen hadrones formados por quarks top y hay que estudiar sus propiedades a través de sus productos de decaimiento [15].

El estudio del quark top se lleva a cabo principalmente en colisiones de hadrones de alta energía. Estas colisiones de alta energía implican el choque de partículas aceleradas a velocidades cercanas a la velocidad de la luz, como ocurre en experimentos realizados en el LHC. Los quarks top

---

<sup>1</sup>probabilidad de ocurrencia del 98%



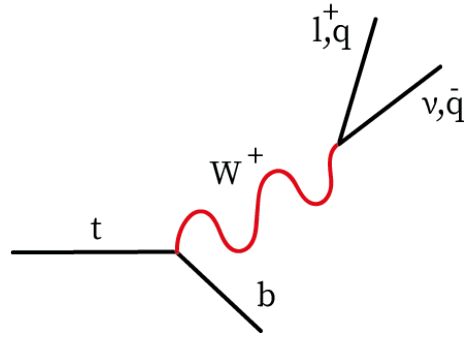


Figura 1.4: Diagrama de Feynman del decaimiento de un quark Top a un bosón W y a un quark b. El bosón W puede desintegrarse ya sea leptónica ( $l^+, \nu$ ) o hadrónicamente ( $q, \bar{q}$ ).

son generados principalmente en *pares* a través de interacciones fuertes (colisiones entre gluones y colisiones quark anti-quark  $q\bar{q}$ , o individualmente mediante interacciones electro-débiles). De acuerdo al modo de producción del quark top, se puede investigar las distintas fuerzas involucradas en estos procesos.

La producción de quarks top juega un papel importante en la búsqueda de nueva física. Varios modelos predicen la existencia de nuevas partículas que se desintegran al quark top. De manera semejante, la producción de *cuatro quarks top* (four top quarks en inglés), está presente dentro de modelos más allá del Modelo Estándar. De hecho, se ha propuesto el estudio de *eventos* con producción de cuatro quarks top como una herramienta para investigar la naturaleza de la ruptura de la simetría electro-débil. Su producción también aparece en modelos específicos que proporcionan un candidato para la materia oscura. Evidentemente, la búsqueda de eventos que contengan cuatro quarks y una posible diferencia en la tasa de generación de este tipo de eventos con relación a los predicho por el SM podría revelar nuevos aspectos en el ámbito de la física.

### 1.4.5. Producción de cuatro quarks top

En el LHC se realizan colisiones de protones a altas energías. Los protones contienen dos quarks up y un quark down, además de gluones que median la fuerza fuerte. En las colisiones protón-protón, la mayoría de los quarks top se producen en pares, principalmente en el canal  $g\bar{g} \rightarrow t\bar{t}$ . Esto

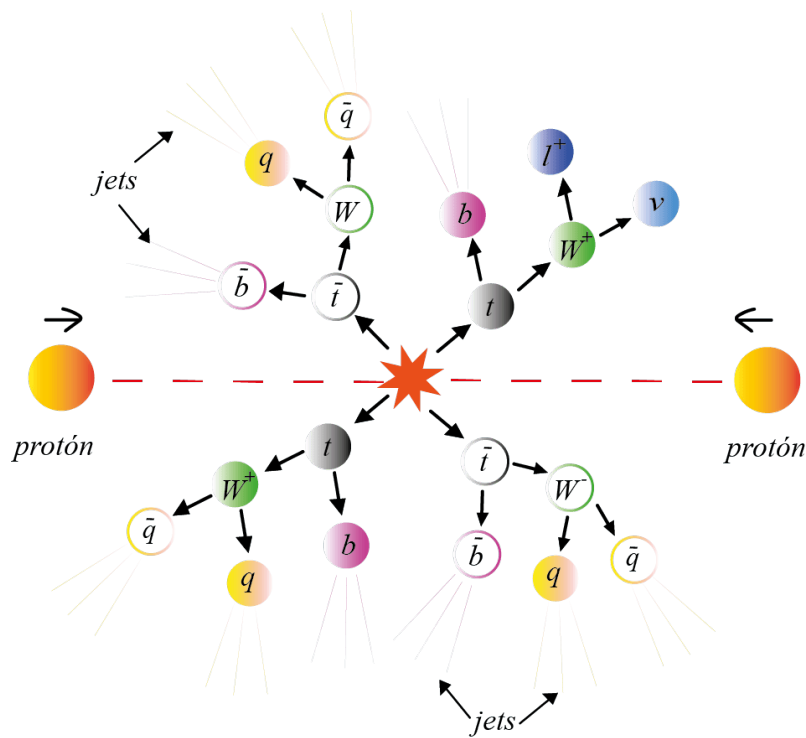


Figura 1.5: Diagrama de un evento de colisión p-p, en la producción de cuatro quarks top  $t\bar{t}t\bar{t}$ .

es, el quark top  $t$  con su antipartícula,  $\bar{t}$ . Sin embargo, también es posible producir cuatro quarks top  $t\bar{t}t\bar{t}$  dentro del SM en colisiones gluón-gluón descritas por diagramas de orden superior en cromodinámica cuántica. La interacción de los gluones viene dada por la colisión entre los protones.

Hay tres modos de decaimientos posibles dependiendo de cómo se desintegra cada par de quark top:

- Decaimiento hadrónico: ambos bosones W ( $W^+$ ,  $W^-$ ) se desintegran en un quark y antiquark  $q\bar{q}$ .
- Decaimiento semi-leptónico: un bosón W se desintegra en  $q\bar{q}$  y el otro bosón W se desintegra en un leptón cargado  $l^{+-}$  y un neutrino  $\nu$ .
- Decaimiento dileptónico: ambos bosones W se desintegran en un leptón y un neutrino cada uno.

Cabe recalcar que cada  $l^{+-}$  podría ser un electrón, un muón o tau. Mientras que cada  $\nu$  podría ser un neutrino electrón, neutrino muón o neutrino tau. A cada uno de estos tipos de decaimientos se los llama canales de desintegración.

En el caso de la producción de cuatro quarks top, los eventos se vuelven más complejos debido a la gran cantidad de partículas involucradas. Los canales de desintegración de los cuatro quarks top podrían involucrar combinaciones de los canales leptónicos y hadrónicos. Por ejemplo, algunos de los quarks top podrían decaer de manera leptónica, mientras que otros decaen de manera hadrónica, como se observa en el diagrama de la figura 1.5. Como resultado de la colisión, se tiene un estado final con un chorro de partículas o *jets*. Este producto del decaimiento se denomina *estado final*.

Una medida utilizada dentro de la física de partículas es el *branching ratio* BR (tasa de desintegración), el cuál cuantifica la probabilidad de que una partícula decaiga a través de un determinado canal de desintegración en lugar de otro. Se expresa como la proporción de la tasa de desintegración en un canal específico con respecto a la tasa total de desintegración.

La producción de cuatro quarks top en colisiones protón-protón es uno de los procesos físicos más raros dentro del modelo estándar, y actualmente uno de los más buscados. En el siguiente capítulo, se expone generalidades sobre el colisionador de Hadrones (LHC) y el experimento CMS que es de relevancia para el siguiente proyecto.

# Capítulo 2

---

## Metodología

---

### 2.1. El Gran Colisionador de Hadrones

El LHC del CERN consiste de un anillo con imanes superconductores de 27 km de largo. Colisiona protones en cuatro puntos a lo largo del anillo en donde se encuentran detectores. Estos detectores son: ATLAS, ALICE, LHCb, y CMS. Los protones se obtienen al aplicar un campo eléctrico a un gas de hidrógeno que lo ioniza. Posteriormente, son acelerados mediante el uso de una serie de aceleradores lineales y circulares interconectados donde el LHC es el acelerador de última etapa, como se puede observar en la Figura 2.1.

En el LHC, los protones se inyectan en *bunches*<sup>1</sup> en sentido horario y antihorario, para luego colisionar a altas energías. En la corrida o Run 2 del LHC (2015-2018), los haces de protones colisionaron a energías de centro de masa,  $\sqrt{s}$ , de 13 TeV. Un *run* es un periodo en donde se registran datos de las colisiones entre grandes periodos de cierre. En la corrida 3 del LHC (2022-actualidad) se alcanzan energías de alrededor  $\sqrt{s} = 13.6$  TeV, lo que presenta una nueva oportunidad en el desarrollo de la física de partículas [16][17].

---

<sup>1</sup>grupos o paquetes de partículas

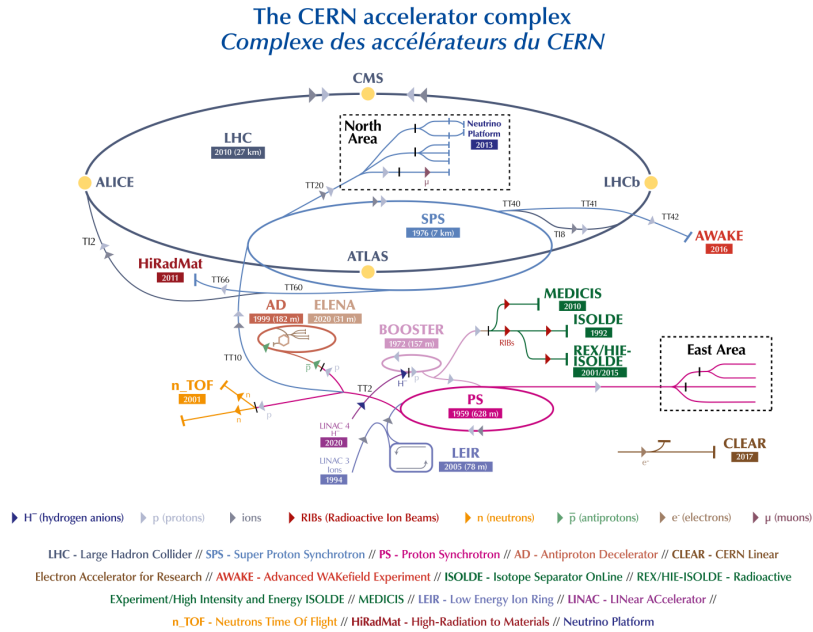


Figura 2.1: Esquema del LHC. Los protones son acelerados desde LINAC 2 al sincrotrón BOOSTER. Luego son acelerados en dos sincrotrones: PS y SPS. Finalmente dos haces en direcciones opuestas son inyectados en el LHC. Los haces se cruzan en los cuatro experimentos: ALICE, ATLAS, LHCb y CMS [18].

### 2.1.1. Medición de la Colisiones

El LHC está diseñado para generar una tasa de cruce de haces de alrededor de 40 MHz, lo que equivale a aproximadamente 600 millones de colisiones por segundo. Un *evento* se refiere, esencialmente, a una colisión energética, por lo que los detectores recopilan una gran cantidad de eventos. Por esta razón, es importante conocer las tasas de diversos eventos que pueden surgir durante una colisión, representando la producción de diferentes estados finales. Se introduce para ello el término *sección eficaz*  $\sigma$ , el cual es una medida de la probabilidad de que un proceso en particular ocurra. La sección eficaz se mide en términos de submúltiplos de barns, la cual es una unidad de área. Otro término relevante dentro de las colisiones es la *luminosidad* integrada  $L_{int}$ . Esta última es, de manera práctica, una medida del número total de eventos. La unidad de la luminosidad es inversa al área  $fb^{-1}$ . Para un proceso de colisiones de partículas, el número de eventos  $N_{eventos}$  se puede calcular a partir de la sección eficaz, con la expresión 2.1.

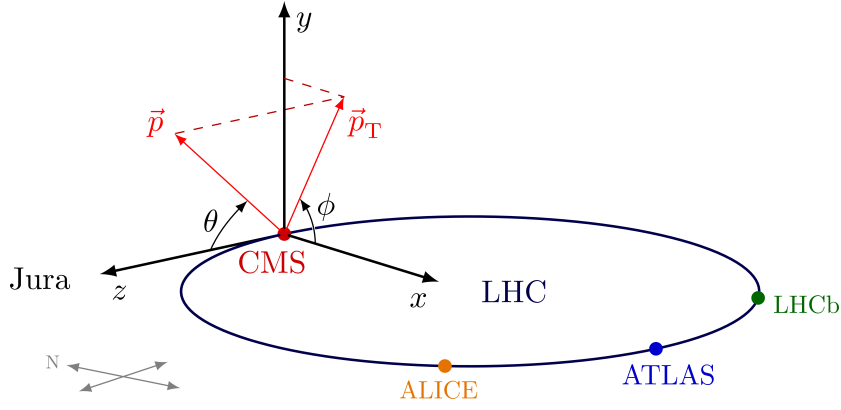


Figura 2.2: Sistema de coordenadas del CMS, incluido el LHC [19].

$$N_{eventos} = L_{int} \times \sigma \quad (2.1)$$

La Figura 2.2 muestra el sistema de coordenadas de uno de los detectores del LHC, el CMS. En esta figura se observa tanto el ángulo azimutal  $\phi$  como el ángulo polar  $\theta$ . La no invariancia de Lorentz del ángulo polar hace que se defina una nueva coordenada que es la pseudorapidez. Los valores de pseudorapidez  $\eta$  indican a qué distancia del centro del detector se produce una partícula. Los valores de pseudorapidez varían de menos a más infinito dentro de la geometría del detector, y su valor puede ser obtenido midiendo el ángulo polar  $\theta$ .

$$\eta = -\ln \tan \theta/2$$

La separación angular  $\Delta R$  se define en base a la pseudorapidez y el ángulo azimutal:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$$

En el contexto de colisiones de protones en el LHC, se denomina *señal* aquello que se busca o se espera medir del experimento. En otras palabras, representa la evidencia de la existencia de un fenómeno físico que pueda indicar la presencia de señales de nueva física. Por otro lado, el proceso de fondo o *background* es cualquier otra cosa que pueda producir el mismo estado final que caracteriza a la señal.

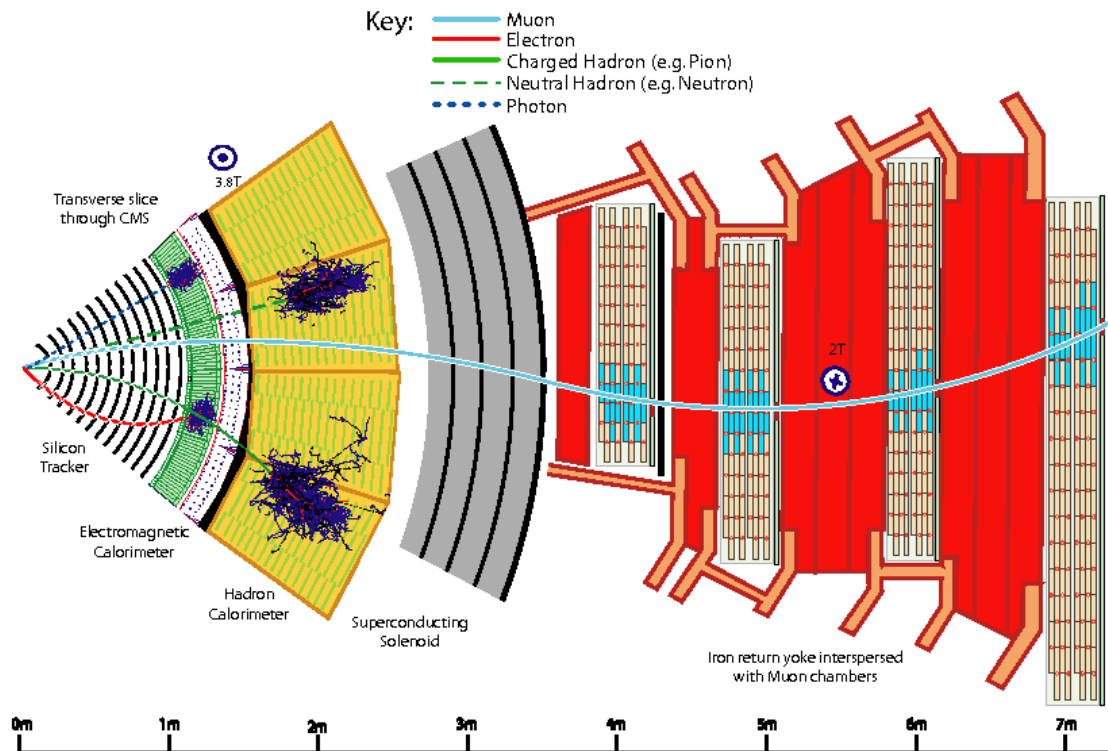


Figura 2.3: Esquema transversal del detector CMS del LHC [20].

## 2.2. El Detector CMS

Uno de los detectores del LHC es el Compact Muon Solenoid (CMS). El CMS es un detector en forma de cebolla cilíndrica, con diferentes tipos de subdetectores en distintas capas. Está compuesto de un solenoide magnético que hace que las partículas cargadas sigan una trayectoria curva a medida que atraviesan el detector, como se observa en la Figura 2.3. Estos subdetectores son: el Silicon Tracker, los Calorímetros, y Cámaras de Muones, recopilan información sobre las partículas generadas en las colisiones [2].

Las partículas de interés, generadas por la colisión, no son detectadas directamente puesto que decaen mucho antes de llegar a la primera capa sensible del CMS. Sin embargo, podemos inferir su presencia a través de sus productos de desintegración.

Es esencial conocer la ubicación del punto de colisión de las partículas. En este contexto, el punto de colisión se denomina comúnmente como punto de vértice primario *PV*. Hay también un vértice secundario

SV que se forma debido a la descomposición de partículas primarias, las cuales viajan unos pocos milímetros después de ser generadas. Posteriormente se desintegran, creando un vértice en el punto de descomposición [21].

### **2.2.1. Subdetectores**

El Silicon Tracker es el encargado de recopilar información que sirve para reconstruir las trayectorias de las partículas cargadas procedentes del vértice de colisión. La reconstrucción de los vértices secundarios también es particularmente importante para distinguir estados finales que involucran quarks top.

El calorímetro electromagnético (ECAL), es el encargado de detectar electrones y fotones. Tanto los electrones como los fotones que impactan con los cristales de tungstanato de plomo, de los que está hecho el ECAL, generan una cascada electromagnética. Por lo tanto, la energía y la posición de estas partículas pueden medirse en los calorímetros.

El calorímetro de hadrones (HCAL), por otro lado, detecta partículas hadrónicas. Mide la posición, energía y tiempo de llegada de partículas más pesadas, como bariones y mesones, mediante capas alternas de absorbentes de latón y acero, junto con materiales centelleadores de plástico fluorescente. Estos materiales generan pulsos de luz cuando una partícula los atraviesa, y la energía luminosa total en una región determinada sirve como medida de la energía de la partícula.

Finalmente, los muones son detectados en lo que se conoce como cámaras de muones. Los muones al ser partículas cargadas como los electrones pero 200 veces más pesadas, son relativamente fáciles de detectar. Es por esto que la capa más externa está diseñada para capturar todos los muones restantes.

### **2.2.2. Trigger**

Como fue mencionado previamente, después de que ocurre una colisión de partículas en un detector, se generan enormes cantidades de datos. La cantidad de información que se puede almacenar es mucho



menor a la cantidad producida. Antes de almacenar datos de un evento, se implementa un sistema de *trigger* para decidir cuáles eventos deben ser registrados y cuáles pueden descartarse en tiempo real antes de almacenar los datos de manera permanente. En el detector CMS, esto ocurre en dos etapas principales: Level-1 trigger (L1) y High-Level trigger (HLT) [22].

## 2.3. Simulación y Reconstrucción de objetos

Para explorar nueva física en el LHC, es esencial modelar los procesos de interés dentro del Modelo Estándar. Los métodos de Monte Carlo se emplean en la física de partículas para simular estos eventos y eventos similares que pueden dar lugar a la misma señal. En el análisis en busca de una señal, los datos obtenidos de la simulación Monte Carlo se comparan con los datos reales provenientes del detector. Desviaciones significativas con el modelo pueden abrir la posibilidad de nueva física. En este análisis, la producción de  $t\bar{t}\bar{t}$  dentro del SM se considera como proceso de señal.

La reconstrucción de objetos es un proceso crucial para identificar y medir las partículas que fueron producidas en las colisiones reales. Esto implica ajustar los parámetros de las simulaciones y los algoritmos de reconstrucción para que coincidan de la mejor manera posible con los datos observados.

### 2.3.1. Reconstrucción de objetos

Las partículas estables productos del decaimiento son las que dejan señales en los subdetectores. Estas señales se utilizan para reconstruir los productos de desintegración; a estos los llamamos *objetos físicos*. La reconstrucción de objetos físicos se realiza mediante el algoritmo *Particle-flow* (PF). Este algoritmo usa la información de todos los subdetectores con el objetivo de reconstruir e identificar todas las partículas estables del evento con el fin de determinar su momento, trayectoria, y energía. Los objetos físicos incluyen muones, electrones, jets, y fotones [23]. Estos datos son almacenados en archivos con formato ROOT. El formato

Requerimientos	Tight Muons	Loose Muons
Es GlobalMuon y TrackerMuon	✓	
Es GlobalMuon o TrackerMuon		✓
$p_T$ (GeV)	$> 26$	$> 10$
$ \eta $	$< 2.1$	$< 2.5$
Rellso	$< 0.15$	$< 0.25$
Impactos en el Tracker	$> 5$	n.a
Impactos en el MuonTracker	$> 0$	n.a

Cuadro 2.1: Filtros para la reconstrucción del objeto físico Muón. La Isolación Relativa o Rellso es definida como una variable dentro del algoritmo PF.

ROOT es un entorno y sistema de análisis de datos desarrollado por el laboratorio CERN y ampliamente utilizado en la física de partículas. Al almacenar los objetos en archivos ROOT, se facilita la navegación directa y el análisis de los datos utilizando herramientas y funciones proporcionadas por este entorno [24]. El proceso de reconstrucción de objetos (*object selection*) se realizó en el código del análisis, *Muon Analyzer*, del cual se detallará posteriormente.

## Reconstrucción de Muones

Es importante poder identificar muones que se originan del proceso de señal de los que aparecen por otros procesos, como por desintegraciones de jets. Para ello se aplican ciertos *criterios de identificación*, asegurando una alta pureza de estos muones. Se consideran dos tipos de muones: *Tight* y *Loose*. Los muones Tight tienen criterios más rigurosos, especialmente relacionados con el momento en el plano transversal del detector  $p_T$ . Los muones Loose se emplean para vetar objetos adicionales que podrían ser muones pero que podrían estar mal identificados, como un pión. Por otro lado, un muón global o *GlobalMuon* es un muón que ha sido detectado tanto en el subdetector Tracker como en la cámara de muones, mientras que un *TrackerMuon* solo se ha identificado a partir de los impactos en el tracker. Los muones Tight y Loose deben cumplir los requerimientos de acuerdo al Cuadro 2.1.

Requerimientos	Tight Electrosn	Veto Electrons
$p_T$ (GeV)	$> 30$	$> 15$
$ \eta $	$< 2.1$	$< 2.5$

Cuadro 2.2: Filtros para la reconstrucción del objeto físico Electrón.

### Reconstrucción de Electrones

De manera similar a los muones, se consideraron dos tipos de electrones: Tight y Veto. Para la reconstrucción de electrones se debe tomar en cuenta la radiación de frenado *Bremsstrahlung*, para medir de manera correcta el momento transversal. Los electrones Tight y Veto deben cumplir los requerimientos de acuerdo al Cuadro 2.2.

### Reconstrucción de Jets

Los *jets* son colecciones de partículas agrupadas espacialmente que se producen cuando un quark o gluón se hadroniza. Las propiedades cinemáticas de los jets se parecen a las de las partículas iniciales que los produjeron, por lo que de ellos se infieren las energías y direcciones de los quarks y gluones de origen. Se utilizan varios criterios para poder contener a jets “falsos” que pueden surgir de la radiación de frenado, y considerar sólo a los jets de interés. Los criterios son los siguientes:

- $p_T > 10$  (GeV)
- $|\eta| < 2.5$
- La distancia angular entre un muón y jets  $\Delta R > 0.4$

### b-tagging

La presencia de los jets indica que de la colisión se obtuvieron partículas como quarks y gluones. En la búsqueda de estados finales que contengan cuatro quarks top, es necesario poder identificar quarks bottom *b* que se originan del decaimiento de quarks top. Esto es especialmente útil para poder hacer la discriminación entre señal y background. El algoritmo Combined Secondary Vertex (CSV) es usado para etiquetar jets que se

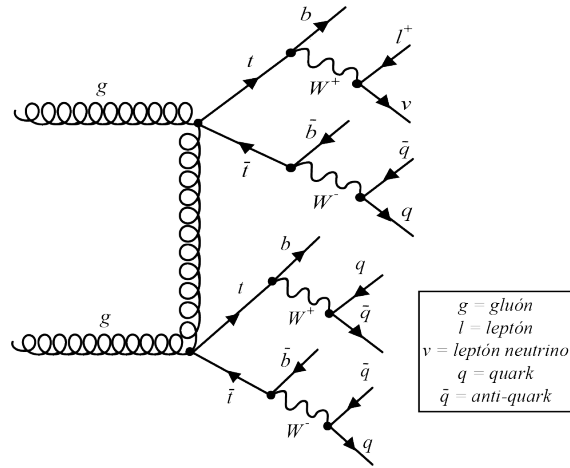


Figura 2.4: Diagrama de Feynman de la producción de cuatro quarks top en el canal leptónico.

originan de quarks b, también llamados *b-jets*. En el presente trabajo se consideró una mejora del algoritmo, CSVv2.

### 2.3.2. CMS Open Data

Desde 2014, la colaboración CMS ha hecho accesibles al público tanto los datos de colisiones reales recopiladas por el detector (data) como las simulaciones de Monte Carlo. El uso de estos *datos abiertos* aporta beneficios significativos, incluyendo transparencia, colaboración y reproducibilidad en la investigación. En este estudio en particular, se emplearon los datos abiertos para la replicación y validación de los resultados presentados en la investigación original [25][26].

## 2.4. Procesos de señal y background

Entre los canales de decaimientos de la producción de cuatro quarks top, el conformado por un único leptón o *single lepton*, es el que tiene mayor BR con respecto a los demás (39%). Esto lo convierte en un canal oportuno para trabajar. Un diagrama de Feynman de la producción de esta señal se puede observar en la Figura 2.4. Este proyecto se centrará en el canal single lepton donde se consideran los estados finales con un solo muón.

### 2.4.1. Background del proceso $t\bar{t}$

Como fue mencionado, la señal a buscar son los cuatro quarks top. El principal background de esta señal es la producción de un par de quarks top,  $t\bar{t}$ . El estado final del proceso  $t\bar{t}$  depende de si cada bosón W decae leptónica o hadrónicamente. De hecho, el canal semi-leptónico de  $t\bar{t}$  puede generar estados finales que se parecen mucho a un estado final de  $t\bar{t}t\bar{t}$ . Es decir, en ambos casos se puede observar un leptón y jets, lo que puede complicar la distinción entre el proceso de señal y el background. A este proceso se lo denominará **ttbar**.

### 2.4.2. Background del proceso Electro-débil

Otro contribuyente al background es el generado por procesos electro-débiles (electro weak) **EW**. Entre los cuales se producen bosones W y Z junto con jets, en donde los bosones W y Z decaen leptónicamente, proporcionando uno o dos leptones en el estado final, respectivamente.

Existen otros procesos de background que pueden contribuir, pero con un valor menor de sección eficaz, por lo cual no serán considerados en este proyecto.

## 2.5. Set de Datos

Para este proyecto se utilizaron los datos de las colisiones del Run 2 del LHC, registrados por el detector CMS en 2015. Los set de datos o datasets de los datos reales y de Monte Carlo MC utilizados en el proyecto son expuestos a los Cuadros 1 y 2 en Anexos. Cada uno de ellos fueron obtenidos del CERN Open Data portal [27].

Nomenclatura a considerar:

- $TTTT$ : se refiere al dataset correspondiente a cuatro quarks top, de la simulación MC.
- $TT$ : se refiere al dataset correspondiente al proceso de background ttbar de la simulación MC.

- *WJets*: se refiere al dataset correspondiente a procesos de producción de pares de bosones W en asociación con jets, de la simulación MC.
- *DYJets*: se refiere al dataset correspondiente a procesos de producción de bosones Z en asociación con jets, de la simulación MC. Este proceso es conocido como Drell-Yan.

Cabe considerar que *WJets+DYJets* conforman el background de MC generado por procesos electro-débiles, EW.

Los datos fueron obtenidos en el formato MiniAOD, específicamente de la versión 76X MC *campaign Fall15MiniAODv1*. El MiniAOD es una forma de organizar y almacenar datos para los análisis físicos. Está diseñado para ser eficiente al ocupar poco espacio (aproximadamente 30-50 kilobytes por evento), al mismo tiempo que proporciona la información esencial necesaria para la mayoría de los análisis. Sin embargo, el formato NanoAOD es un formato que es aproximadamente 20 veces más compacto que el formato MiniAOD. En el trabajo presente se han procesado los datos y se han convertido al formato NanoAOD. Esta transformación ha permitido reducir significativamente el tamaño de los eventos en comparación con el formato MiniAOD, manteniendo al mismo tiempo la información esencial necesaria para el análisis físico [28][29].

## 2.6. Archivos NanoAOD

Todos los datos de la colisión (data) y de montecarlo se procesaron utilizando la versión 76X del software CMS, CMSSW. Para ello se utilizó el servicio de clúster LXPLUS<sup>2</sup> del CERN. Se detalla a continuación información relevante sobre el proceso.

### 2.6.1. CMSSW

El software CMSSW consiste en un conjunto de bibliotecas de software escritas en C++ empleado por el experimento CMS para la producción,

---

<sup>2</sup>Servicio de inicio de sesión interactivo en Linux para todos los usuarios del CERN.

procesamiento y análisis de sus datos. Su configuración se controla mediante el lenguaje Python, lo que facilita su adaptabilidad y hace que sea más asequible para la comunidad científica. El modelo de procesamiento de eventos CMSSW consta de un ejecutable, llamado `cmsRun`, y muchos módulos complementarios administrados por un Framework. Todo el código necesario en el procesamiento de eventos (calibración, algoritmos de reconstrucción, etc.) está contenido en los módulos. Se utiliza el mismo ejecutable tanto para procesar los datos del detector como para generar los de Monte Carlo [30][31].

El Modelo de Datos de Eventos (EDM) en CMS se centra en el concepto de *evento*. En este contexto, un evento representa todos los datos asociados con una colisión específica en el detector. Se utiliza un contenedor de objetos C++ para representar estos datos. Los módulos interactúan y comparten información a través de la estructura del evento. Todos los objetos contenidos pueden almacenarse individual o colectivamente en archivos ROOT.

## 2.6.2. Procesamiento de datos

Los códigos utilizados en el presente trabajo para el procesamiento de datos y el análisis se los puede encontrar en el siguiente repositorio de Github [32].

El procesamiento de datos se realizó mediante el uso de varios programas desarrollados en el lenguaje de programación *Python* y scripts de *shell*. Estos scripts están hechos para ejecutarse con *HTCondor batch system*. HTCondor es un sistema de gestión de recursos distribuidos utilizado para la ejecución de trabajos en clústeres de computadoras. Todos los usuarios que tengan acceso al servicio Linux del CERN pueden enviar trabajos a HTCondor, especialmente cuando se necesitan grandes cantidades de recursos computacionales [33].

Los datasets fueron descargados y organizados en el directorio *data*. Los datasets en formato tipo index contienen los archivos `.root` con formato MiniAOD. Para la ejecución de los códigos, se define la variable *process* que indican los diferentes datos a procesar dentro de directorio. Los diferentes valores que toma son: Run2015D (datos), TT (ttbar), TTTT,

WJets, y DYJets. Se detalla a continuación sobre los códigos utilizados.

Cabe mencionar que los archivos resultantes del procesamiento de datos, archivos `.root` en formato NanoAOD, fueron almacenados en EOS. Esta es una tecnología de almacenamiento utilizada en CERN. La Figura 2.5 muestra un diagrama del proceso llevado a cabo.

- El script `job.sh` contiene la configuración o set up de CMSSW. Prepara el entorno de trabajo y establece las configuraciones necesarias para que el software pueda funcionar correctamente. Toma de argumento la variable `process`, y devuelve el valor de `True` si son datos (Run2015D) o `False` (cualquier otro proceso, datos MC).
- El programa `create_job.py` toma como argumento el proceso y lo busca en el directorio `data`. Va leyendo cada línea de los index archivos correspondientes y escribe en un archivo llamado `argumentos.txt`: el número de proceso, el proceso en sí y la línea (archivo `root`). En el directorio del proceso crea los subdirectorios `out`, `log` y `err`. Escribe el archivo `job.dl`, que contiene información que lee HTCondor sobre el trabajo o job a realizar<sup>3</sup>. Finalmente, abre el archivo `job.sh` y lo copia a uno nuevo con el nombre del proceso que será el ejecutado. Ej: `TT.sh`
- El programa principal es `nanoAOD.py`. Este es el responsable del procesamiento de los datos y de obtener los archivos `.root` como archivo de salida en formato nanoAOD.
- El archivo `job.jdl` ejecuta el archivo `.sh`, (por ejemplo, `TT.sh`). Este crea el output directorio `ntuples` en EOS. Configura el sistema para CMSSW. Corre el archivo `nanoAOD.py` con el argumento `True` (`data`) o `False` (`MC`). Copia el archivo de salida a EOS.
- Todo se ejecuta con `submit_jobs`. Este ejecutable define el path<sup>4</sup> en donde se crea la carpeta `JobLaunch`. Define el proceso y para el mismo crea un directorio y llama al archivo `create_job.py`. Luego,

---

<sup>3</sup>El comando `condor_submit` toma un archivo de descripción del trabajo (`jdl`) como entrada y envía el trabajo a HTCondor. Para más información véase: <https://htcondor.readthedocs.io/en/latest/users-manual/submitting-a-job.html>

<sup>4</sup>Ruta de un archivo o directorio en un sistema de archivos.



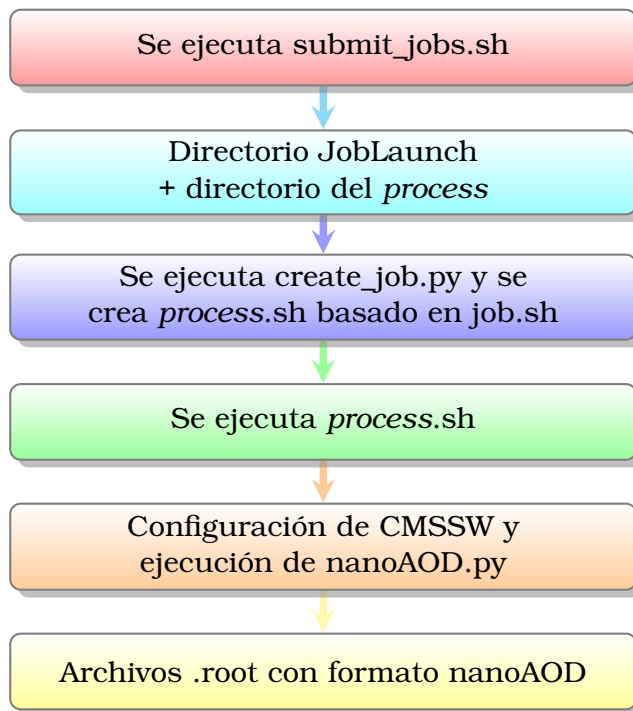


Figura 2.5: Diagrama del procesamiento de datos.

dentro del directorio del proceso en cuestión ejecuta `job.jdl`. Se ejecuta este código con Condor.

Es necesario mencionar que, los archivos `out`, `log` y `err` creados para cada uno de los datos procesados, son de gran utilidad. Estos dan información sobre la ejecución de los códigos, los errores encontrados e información en caso de fallas.

### 2.6.3. Archivo JSON

Los archivos `.root` obtenidos son unidos en un solo archivo en formato JSON. Un *archivo JSON* es un archivo que almacena objetos y estructuras de datos simples en formato de notación de objetos JavaScript (JSON). Al ser livianos, son bastantes prácticos para realizar el análisis. Para ello se utilizó el script `make_jsonfile.py`. Este archivo contiene información del proceso en cuestión y el número de eventos. Para ello se utilizó la librería de `uproot` de Python, que lee y escribe archivos ROOT. A priori se utilizó el script `check_files.py` para comprobar que todos los archivos `.root` obtenidos para cada dataset no estuvieran dañados o

vacíos.

## 2.7. Selección de Eventos

La estrategia para seleccionar eventos  $t\bar{t}\bar{t}$  mientras se suprimen procesos de background, es hacer que el canal con el que se trabaje cumpla con ciertos requerimientos o cortes. Esto es lo que se conoce como *Baseline Event Selection*. Para el canal leptónico muon+jets, se debe cumplir lo siguiente.

- Exactamente un solo muón.
- Exactamente cero muones sueltos adicionales (loose muons).
- Exactamente cero electrones de veto (veto electrons).
- Al menos 6 jets.
- Al menos 2 jets cumplen el criterio CSVv2M de 0.8.

Estos filtros fueron tomados en cuenta de acuerdo a los objetos físicos construidos conforme al *Analysis note* de la publicación original. Los eventos en el set de datos se recopilaron mediante triggers considerando la presencia de un único muón. El trigger utilizado para este canal fue *HLT\_IsoMu18*.

Se realizó un código en Python para realizar el análisis, *Muon Analyzer*. Este código principal lee el archivo JSON y toma en cuenta la selección de eventos y el trigger. Para un mayor control de los filtros empleados, se obtuvo el número de eventos de cada proceso después de la aplicación de cada filtro.

El resultado esperado consta de histogramas, de distintas variables físicas, que representan la selección del eventos. Estos histogramas permitieron definir las variables a utilizar para el análisis multivariante.

## 2.8. Análisis Multivariante

El análisis multivariante (MVA) ha demostrado ser exitoso en distintos análisis dentro de la Física de Partículas. Particularmente, es útil en la discriminación de señal y background (clasificación). En general se utiliza para encontrar patrones y correlaciones entre un conjunto de datos analizando dos o más variables a la vez. Este análisis se lleva a cabo mediante técnicas de *machine learning*, que utiliza algoritmos avanzados para aprender patrones complejos a partir de los datos [34][35].

Un algoritmo común de machine learning recibe como entrada los observables, que son las características o variables de interés de un conjunto de datos específico. El algoritmo utiliza estas características para realizar una tarea específica, ya sea regresión (predicción de valores continuos) o clasificación (asignación a categorías discretas).

La variables de entrada (input) pueden ser, por ejemplo, la energía y los momentos de las partículas que se obtienen de una colisión, donde cada punto de datos representa un evento único detectado en el experimento [36].

En el caso de clasificación, el modelo de machine learning aprende a asignar un valor o una etiqueta a los observables. Esta etiqueta suelen ser valores discretos. En el contexto de discriminación entre señal y background, los algoritmos se entrenan en eventos simulados etiquetados, donde tradicionalmente los procesos de señal se etiquetan con 1 y los procesos de background con 0.

Después de la fase de aprendizaje, el algoritmo puede hacer predicciones sobre datos desconocidos. El modelo es puesto a prueba con datos que no “ha visto” antes y que no están etiquetados. Estos pueden ser eventos o datos reales de colisiones.

Dentro del ámbito del análisis multivariable, se encuentran diversos modelos. Entre ellos destaca el método de Boosted Decision Trees (BDT), debido a su poder de clasificación. Este enfoque se vuelve especialmente valioso en situaciones donde las variables de un conjunto de datos no presentan una relación clara entre sí.

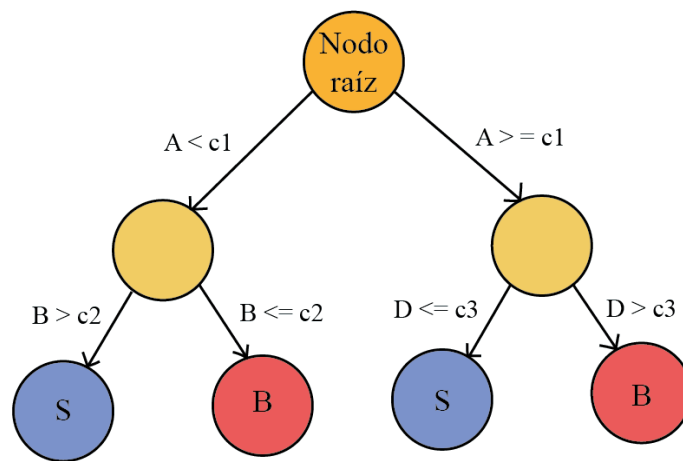


Figura 2.6: Esquema de un Decision Tree. Para una muestra de eventos de señal (S) y de fondo (B), es descrito por el conjunto de variables, A, B y D. Cada vez que una de estas variables cumplan con alguna condición de acuerdo a  $c1$ ,  $c2$  y  $c3$ , el nodo se divide en dos: en una rama que cumple con la condición y en otra que no. Esto se realiza de forma recursiva hasta clasificar en señal o background.

### 2.8.1. Método Boosted Decision Trees

Matemáticamente, los *decision trees* son árboles binarios (se consideran árboles con dos clases: señal y background). En la Figura 1.9 se muestra un ejemplo de su algoritmo. El árbol comienza desde un nodo inicial, el nodo raíz. Cada nodo se puede dividir de forma recursiva en dos ramas, hasta que se alcance alguna condición en particular [37].

Un Boosted Decision Tree o BDT es una mejora de este árbol mediante un proceso llamado *boosting*. El boosting implica la combinación de múltiples decision trees débiles para formar un modelo más robusto y preciso. Cada árbol se construye de manera secuencial, y se da más peso a los datos mal clasificados en cada iteración. Esto permite que el modelo se ajuste y mejore continuamente, reduciendo así los errores de predicción.

La evaluación más común de un clasificador binario se realiza a través de su curva ROC (Receiver Operating Characteristic). Muestra la eficiencia del clasificador en la señal (true positive rate) frente a la ineficiencia de eliminar el background (false positive rate). Cada punto de esta curva corresponde a un umbral de valor de corte. La calidad del clasificador se cuantifica comúnmente mediante el área bajo la curva (AUC), de manera

que un buen clasificador tendrá un AUC cercano a 1.

Se utilizó la técnica de Boosted Decision Trees para la discriminación entre la señal  $t\bar{t}\bar{t}$  y background. Para ello se utilizó el tutorial Análisis Multivariante de Python en [38]. Para el entrenamiento del modelo, se utilizó la técnica *Cross Validation*, en la que se dividen los datos en múltiples subconjuntos para entrenamiento y prueba. Se utilizaron los datos de la simulación MC para el entrenamiento ( $t\bar{t}\bar{t}$  como señal, demás procesos como background), mientras que los datos reales se utilizaron para la prueba del modelo. Cabe recalcar que se utilizaron dos tipos de algoritmos: *Train XGBoost Classifier* y *Adaptive Booster*.

# Capítulo 3

---

## Resultados, conclusiones y recomendaciones

---

### 3.1. Selección de Eventos

Se presentan a continuación los resultados de la selección de eventos realizada en el *Muon Analysis*. A través de esto, se busca analizar la correspondencia entre los datos de la simulación de Monte Carlo y los datos experimentales para cada una de las variables de interés.

En el Cuadro 3.1, se expone el número de eventos para cada uno de los procesos, Monte Carlo y datos, luego de la aplicación de cada filtro. Es evidente que  $t\bar{t}$  tiene un mayor aporte en la contribución final del background, de acuerdo a lo mencionado en la Sección 2.4.1.

En el proceso de análisis, se implementó la normalización de eventos y la asignación de pesos para asegurar la coherencia y validez de los resultados. Para las muestras de Monte Carlo, la normalización se calculó mediante la fórmula 2.1, dividiendo la expresión para el número de eventos ( $N_{eventos}$ ). En un estudio previo, que involucraba datos reales del canal leptónico electrones + jets junto con las mismas muestras de MC, se identificaron inconsistencias en la normalización de estos datos [39]. Sin embargo, a través del reprocesamiento de estos datos, se logró obtener una normalización congruente.

Como resultado de la selección de eventos, se obtuvieron también histogramas de distintas variables físicas. Cada histograma representa el

Filtros	Data	$T\bar{T}$ bar	$t\bar{t}\bar{t}$	Wjets	Dyjets
Inicial	74022265	97994442	250000	47161328	8842520
Trigger	48920060	14950386	66292	7972985	2180157
Un sólo muón	18521253	11241049	45597	5404174	881431
Cero leptones adicionales	17741556	10758863	40796	5375539	522561
Jets $\geq 6$	42412	1823952	39722	2566	643
Jets b-tagging $\geq 2$	14091	819169	30769	96	36

Cuadro 3.1: Número de eventos luego de cada corte para cada proceso.

número de eventos que se obtiene para los distintos valores que toma la variable física correspondiente. Esto se realizó para cada uno de los procesos de background ( $t\bar{t}$ bar,  $W$ Jets,  $DY$ Jets), la señal de Monte Carlo ( $t\bar{t}\bar{t}$ ), y los datos reales ( $Data$ ). Cabe destacar que la señal  $t\bar{t}\bar{t}$  es pequeña en comparación con las fuentes de background, por lo que fue necesario multiplicarla por un factor de 20 para una mejor visualización de su distribución en cada uno de los histogramas.

Por otro lado, la sub-gráfica inferior  $DATA/MC$  representa la razón entre los datos experimentales y las simulaciones. Su funcionalidad es de comparar los datos de las colisiones reales con la simulación MC. A medida que esta razón se acerca a 1, los datos experimentales se ajustan mejor con las simulaciones. El patrón de trama  $///$  es una representación visual de cómo se verían las incertidumbres. Este fuera el caso si se utilizaran un mayor número de datos de MC, lo cuál queda fuera del alcance del trabajo. Este patrón se aplica a cada barra de error calculada en el gráfico, creando líneas diagonales que muestran la incertidumbre asociada con estos datos.

La Figura 3.1 muestra el histograma para la variable  $p_T$  del muón. Se puede observar que, a pesar de no haber considerado todos los posibles eventos de background, no hay una desviación evidente entre datos y procesos de fondo. Un estudio estadístico más detallado, como el BDT, se utiliza para buscar si existe probabilidad de la presencia de señal.

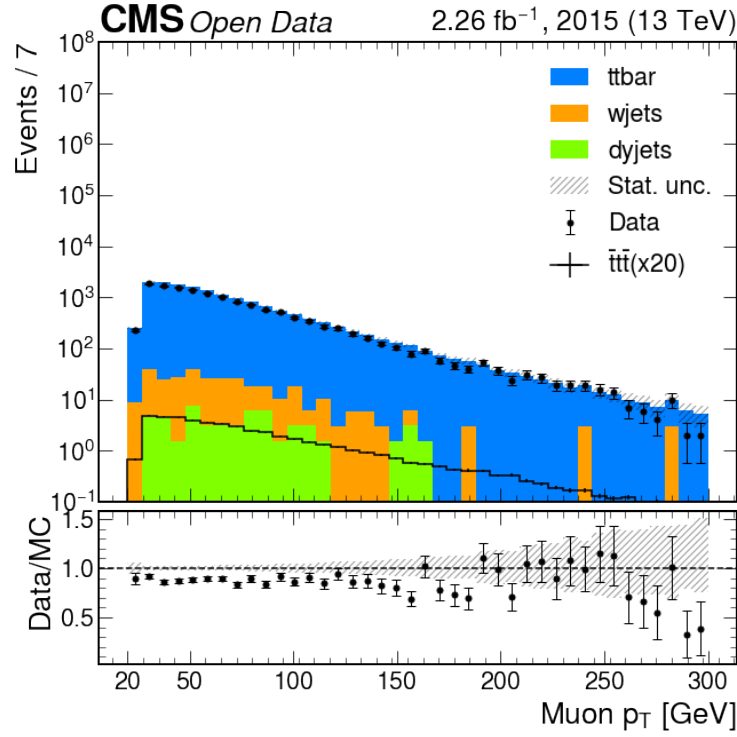


Figura 3.1: Distribuciones de  $p_T$  para datos y simulación para el canal muón+jets.

### 3.2. Variables Discriminatorias

Respecto al BDT (Boosted Decision Tree), el objetivo es incorporar variables que puedan ayudar a distinguir entre el background y la señal. Como el background dominante es  $t\bar{t}$ , se busca variables que permitan discriminar entre este background y  $t\bar{t}t\bar{t}$ .

En los eventos  $t\bar{t}t\bar{t}$ , la desintegración puede tener hasta diez jets, mientras que los eventos  $t\bar{t}$  pueden tener hasta cuatro jets, como se observa en la Figura 3.2. En consecuencia, la variable que representa el número de jets reconstruidos  $N_j$ , puede ayudar a discriminar entre la señal y background.

Por otro lado, los b-jets que se originan a partir de la desintegración de los quarks top tienden a tener valores más altos de momento transversal faltante,  $p_T$ . Como resultado, es útil utilizar una variable para cuantificar la suma de los momentos transversales de los b-jets que tenderán a tener valores mayores para los eventos de señal en comparación con los eventos de background,  $H_T^b$ . La Figura 3.3 muestra las distribuciones de esta



variable para datos y simulación.

Además, la tasa de desintegración BR del quark top a un quark b es de prácticamente el 100 %, por lo que los eventos que contengan cuatro quarks top contendrán cuatro quarks b. El background  $t\bar{t}$  normalmente produce dos quarks b, mientras que  $t\bar{t}\bar{t}\bar{t}$  produce cuatro. Por lo tanto, la presencia de más de dos b-jets es una fuente potencial de discriminación. Ergo, la variable número de b-jets  $N_{bj}$  fue escogida para el análisis. La Figura 3.4 muestra las distribuciones de esta variable para datos y simulación.

Los jets con los valores de momento transversal más altos se denominan jets principales y sus fuentes pueden variar según el tipo de evento. En eventos de background, los cuatro jets principales se originan principalmente a partir de la desintegración de los top quarks, mientras que los otros jets surgen de otros procesos. En consecuencia, se espera que el  $H_T$  (suma escalar de  $p_T$ ) de los cuatro jets principales sea relativamente mayor en comparación con el  $H_T$  de los jets restantes en el evento. Esta diferencia en la distribución de los valores de  $H_T$  se puede aprovechar utilizando la relación entre la suma escalar de  $p_T$  de los jets principales y el  $H_T$  de los jets restantes en el evento,  $H_T^{RAT}$ .

En resumen, se eligieron las variables físicas  $N_j$ ,  $N_{bj}$ ,  $H_t^b$  y  $H_T^{RAT}$  para realizar el análisis multivariante.

### 3.3. Boosted Decision Trees

Para determinar la relación entre dichas variables, se graficó la distribución de la señal (puntos azules) y background (puntos rojos) entre pares de variables. Esto se puede observar en las Figura 3.5-7.

Luego de realizar el entrenamiento, se obtuvieron las curvas ROC para cada una de los clasificadores, como se muestra en la Figura 3.8. El algoritmo XGBoost mostró tener un mayor poder de clasificación. Fue este mismo, por lo tanto, el que se usó para probar el modelo entrenado con los datos reales.

Finalmente, en la Figura 3.9 se presenta la clasificación de eventos realizada por el modelo seleccionado, utilizando tanto los datos de Monte

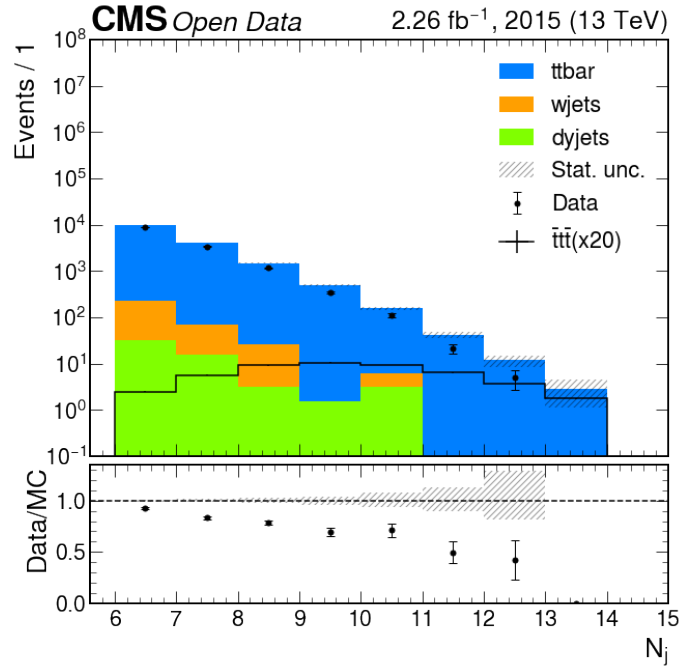


Figura 3.2: Distribuciones  $N_j$  para datos y simulación.

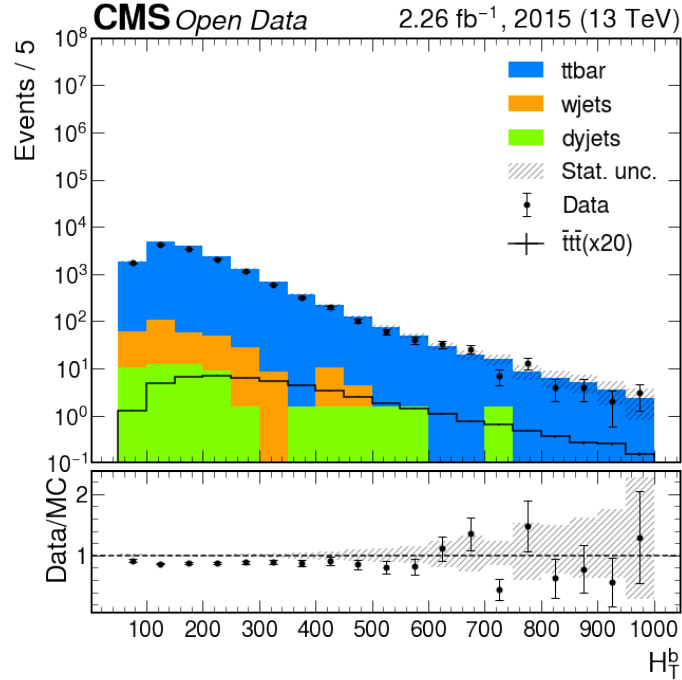


Figura 3.3: Distribuciones  $H_t^b$  para datos y simulación.

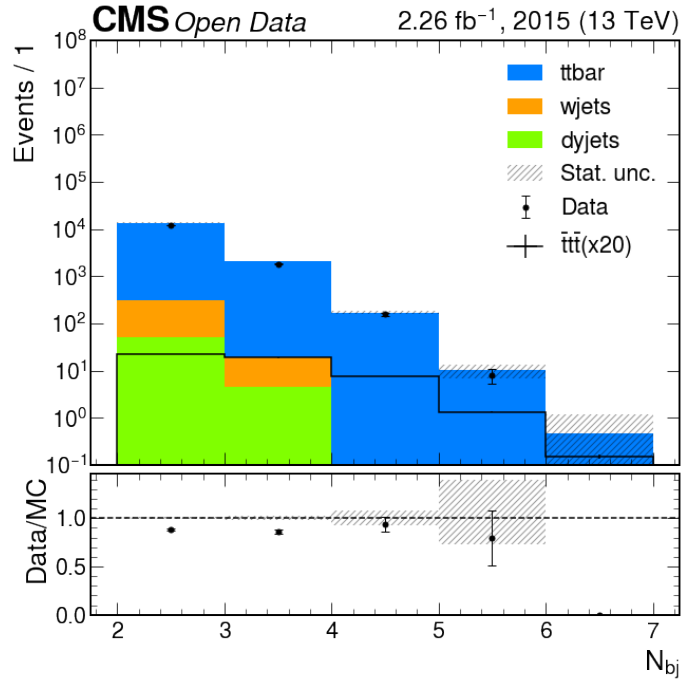


Figura 3.4: Distribuciones  $N_{bj}$  para datos y simulación.

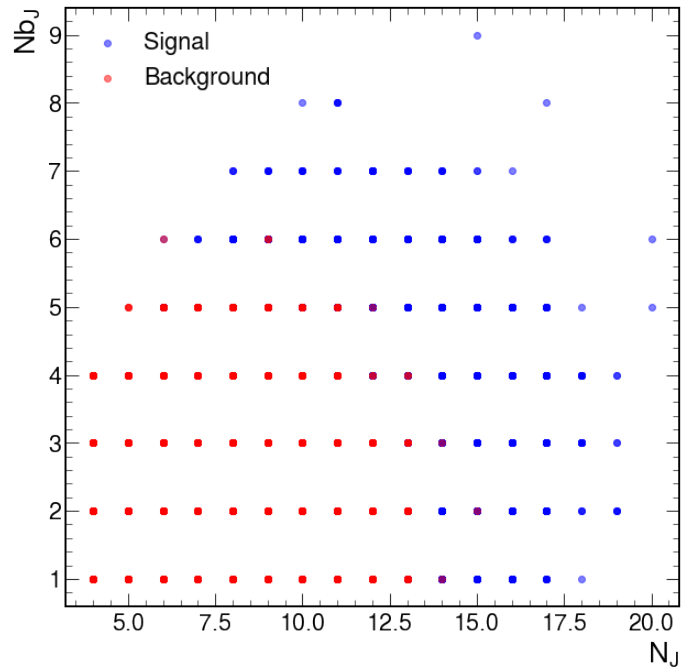


Figura 3.5: Distribución de la señal y background de la simulación Monte Carlo entre las variables  $N_{bj}$  y  $N_j$ .

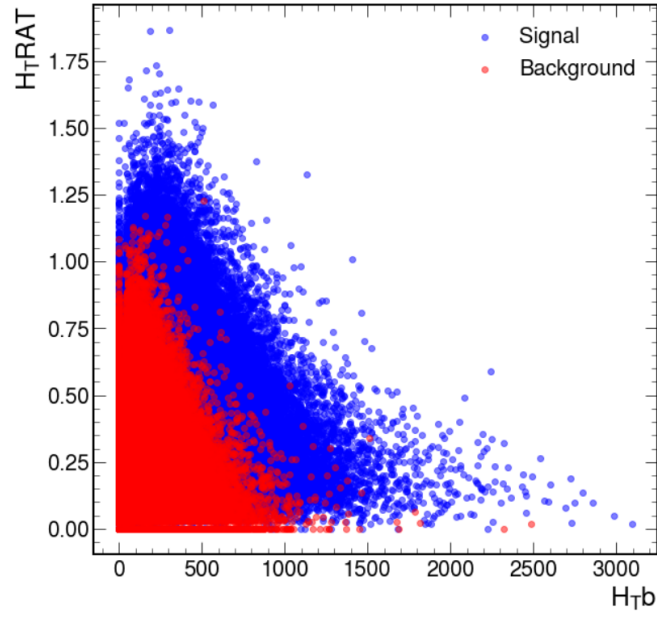


Figura 3.6: Distribución de la señal y background de la simulación Monte Carlo entre las variables  $H_T^{RAT}$  y  $H_T^b$ .

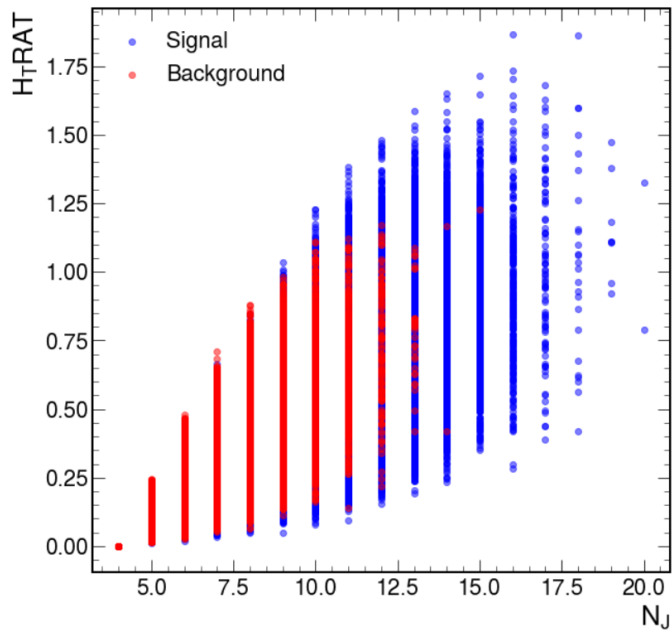


Figura 3.7: Distribución de la señal y background de la simulación Monte Carlo entre las variables  $H_T^{RAT}$  y  $N_j$ .

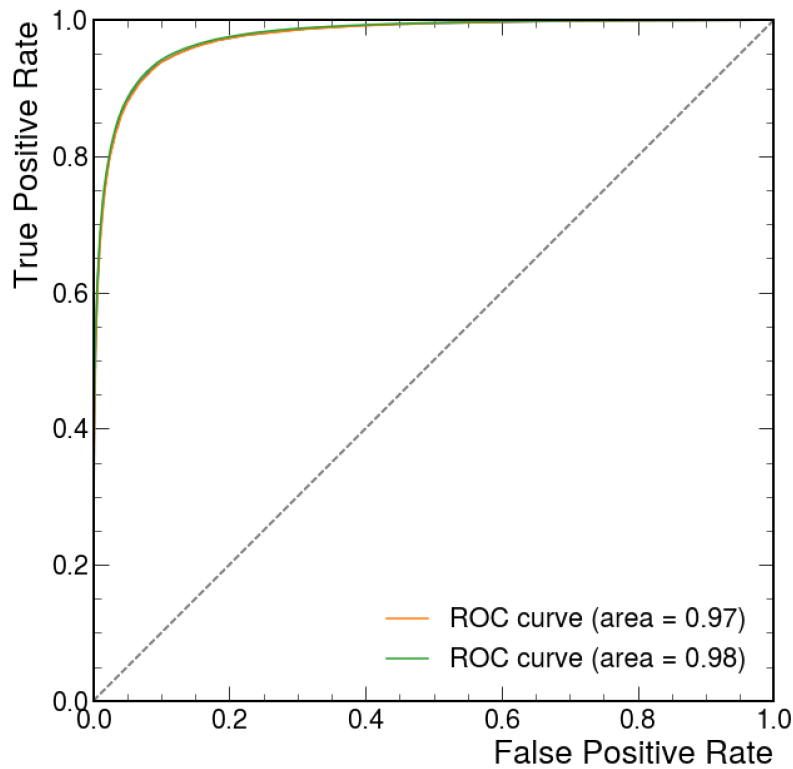


Figura 3.8: Curva ROC para los algoritmos XGBoost (en verde) y Adaptive Booster (en naranja).

Carlo como los datos reales de las colisiones. En este gráfico, se representa la probabilidad de que un evento sea clasificado como señal (tendiendo a 1) o como background (tendiendo a 0).

Se observa que, para la simulación Monte Carlo, hay algunos eventos clasificados como señal, los cuales coinciden con la señal utilizada en la simulación. Sin embargo, al analizar los datos reales de las colisiones, que son de nuestro interés, se aprecia que la probabilidad de tener eventos con la señal de cuatro quarks top  $t\bar{t}t\bar{t}$  tiende a cero. En otras palabras, no se observan eventos con la señal dentro de los datos reales, lo cual coincide con el resultado reportado en el análisis original.

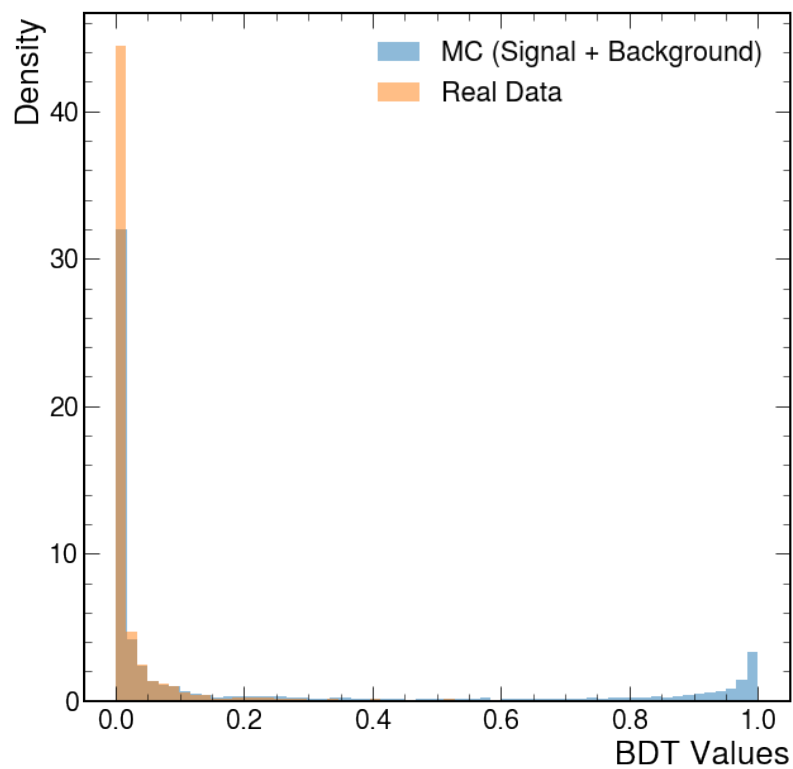


Figura 3.9: Densidad de eventos para las muestras de Monte Carlo *MC* y datos de las colisiones *Real Data*, en la probabilidad de que sean procesos de señal (1) o de fondo (0).

### 3.4. Conclusiones y recomendaciones

Se realizó el análisis simplificado en la búsqueda de cuatro quarks top, usando los datos abiertos del CERN. Se utilizaron los datos correspondiente al Run 2 del 2015, para una luminosidad de  $2.26 \text{ fb}^{-1}$  con centros de masa de  $\sqrt{s} = 13 \text{ TeV}$ . El estudio se centró en el canal leptónico muon+jets.

La primera parte del trabajo constó del procesamiento de los datos correspondientes a las colisiones reales y la simulación Monte Carlo en formato MiniAOD. Para ello se utilizó el software CMSSW y códigos computacionales, en donde se consideraron los distintos procesos de fondo o background y los datos experimentales. Se obtuvieron de manera satisfactoria los archivos root correspondientes en el formato NanoAOD, los cuales fueron almacenados en el archivo json con el número de eventos esperado para cada proceso.

La reconstrucción de los objetos físicos y la selección de eventos se llevó a cabo mediante un código principal, incorporando los filtros propuestos en la publicación original. Los histogramas obtenidos de las distintas variables físicas, mostraron concordancia entre los datos reales con la simulación Monte Carlo. Es relevante mencionar que no se ha realizado un análisis de posibles errores sistemáticos. Se recomienda su evaluación en futuros estudios para mejorar la confiabilidad del análisis.

Posteriormente, se realizó un análisis multivariante utilizando el método BDT para la discriminación entre señal y fondo. La evaluación del rendimiento de dos algoritmos a través de la curva ROC reveló un desempeño sólido en la tarea de discriminación. En los resultados obtenidos del BDT, no se observaron eventos con la señal de cuatro quarks top dentro de los datos reales, corroborando así los hallazgos de la publicación original.

En resumen, este estudio ha cumplido de manera integral con el objetivo de adquirir las habilidades y el conocimiento necesarios para llevar a cabo análisis similares dentro del campo de la física de partículas. La selección de eventos, el procesamiento de los datos y la aplicación de técnicas avanzadas de análisis multivariante no solo han arrojado resulta-

dos coherentes, sino que también han consolidado una base sólida para la participación en futuras investigaciones dentro de este dominio.

CMS reportó recientemente el descubrimiento de la producción de cuatro quarks top. En este estudio se utilizaron muchos más datos y se afinó el entendimiento de los componentes sistemáticos. Se propone un incremento sustancial de los datos y un estudio de los errores sistemáticos para determinar si es posible encontrar dicha señal. Este enfoque adicional permitiría una comprensión más completa de los errores involucrados en el análisis de la producción de cuatro quarks top y, en general, en otros eventos de estudio que se realizan dentro de la física de partículas.



---

## Anexos

---

Set de Datos	$L_{int}$ ( $pb^{-1}$ )
/SingleMuon/Run2015D-16Dec2015-v1	2256.38

Cuadro 3.2: Dataset de las colisiones para el canal muon + jets. Las muestras se encuentran en el formato MiniAOD y pueden ser encontradas en el CERN Open Data Portal.

Set de Datos	Sección eficaz ( <i>pb</i> )	Proceso
/TTTTTuneCUETP8M113TeV-amcatnlo-pythia8/ RunIIFall15MiniAODv2-PU25nsData2015v176X mcRun2asymptoticv12ext1-v1	0.009	Señal tttt
/TTTuneCUETP8M113TeV-powheg-pythia8/ RunIIFall15MiniAODv2-PU25nsData2015v176X mcRun2asymptoticv12ext3-v1/	831	ttbar
/WJetsToLNuTuneCUETP8M113TeV-madgraphMLM-pythia8/ RunIIFall15MiniAODv2-PU25nsData2015v176X mcRun2asymptoticv12-v1	61526	EW
/DYJetsToLLM-50TuneCUETP8M113TeV-madgraphMLM-pythia8/ RunIIFall15MiniAODv2-PU25nsData2015v176X mcRun2asymptoticv12-v1	6025	EW

Cuadro 3.3: Datasets de los datos MC utilizados para el canal muon + jets, junto con la sección eficaz y el nombre del proceso. Las muestras se encuentran en el formato MiniAOD y pueden ser encontradas en el CERN Open Data Portal.

---

## Referencias bibliográficas

---

- [1] Mark Thomson. *Modern Particle Physics*. University of Cambridge, 2013.
- [2] The CMS Collaboration et al. 3(08):S08004, aug 2008.
- [3] Georges Aad et al. Evidence for  $t\bar{t}\bar{t}$  production in the multilepton final state in proton–proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector. *Eur. Phys. J. C*, 80(11):1085, 2020.
- [4] Search for production of four top quarks in final states with same-sign or multiple leptons in proton–proton collisions at  $\sqrt{s} = 13$ tev. *The European Physical Journal C*, 80(2), 2020.
- [5] Lana Beck. The Search for the Standard Model Production of Four Top Quarks, 2016.
- [6] Search for standard model production of four top quarks in proton–proton collisions at  $s=13$  tev. *Physics Letters B*, 772:336–358, 2017.
- [7] David Griffiths. *Introduction to Elementary Particles*. Second edition. Wiley-VCH, 2009.
- [8] C. Patrignani. Review of particle physics. *Chinese Physics C*, 40(10):100001, 2016.
- [9] Cern. The higgs boson. <https://home.cern/science/physics/higgs-boson>, 2023. Accessed: (29-12-2023).
- [10] Tom W. B. Kibble. The standard model of particle physics, 2014.

- [11] Julia Woithe, Gerfried J Wiener, and Frederik F Van der Veken. Let's have a coffee with the standard model of particle physics! *Physics Education*, 52:034001, 2017.
- [12] Brian Martin. *Nuclear and particle physics*. 3rd Edition. Wiley, 2009.
- [13] T. S. Virdee. Beyond the standard model of particle physics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2075)., 2016.
- [14] Daniela Paredes Hernández. *Search for New Physics in events with 4 top quarks in the ATLAS detector at the LHC*. Theses, Université Blaise Pascal - Clermont-Ferrand II, September 2013. N° DU : 2375, N°EDSF : 758.
- [15] Michele Gallinaro. Top quark physics: A tool for discoveries. *Journal of Physics: Conference Series*, 447(1):012012, jul 2013.
- [16] Deepak Kar. *Experimental Particle Physics*. 2053-2563. IOP Publishing, 2019.
- [17] Cern. Lhc run 2. <https://home.cern/tags/lhc-run-2>, 2023. Accessed: (29-12-2023).
- [18] Cern. Facts and figures about the lhc. <https://home.cern/resources/faqs/facts-and-figures-about-lhc>, 2023. Accessed: (29-12-2023).
- [19] Izaak Neutelings. Cms coordinate system. [https://tikz.net/axis3d\\_cms/](https://tikz.net/axis3d_cms/), 2023. Accessed: (29-12-2023).
- [20] Woodbank Communications. Cms particle detector. <https://www.mpoweruk.com/CMS-Detector.htm>, 2023. Accessed: (29-12-2023).
- [21] Kazunori Hanagaki, Junichi Tanaka, Makoto Tomoto, and Yuji Yamazaki. *Particle Identification*, pages 69–114. Springer Japan, Tokyo, 2022.
- [22] Somnath Choudhury. Performance of the high-level trigger system at cms in lhc run-2. *IEEE Transactions on Nuclear Science*, 2021.

- [23] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. Technical report, CERN, Geneva, 2009.
- [24] ROOT Team 2023. Root: analyzing petabytes of data, scientifically. <https://root.cern/>. Accessed: (29-12-2023).
- [25] CMS Collaboration. 2020 cms data preservation, re-use and open access policy. <https://opendata.cern.ch/docs/about-cms>, 2023. Accessed: (29-12-2023).
- [26] CMS Collaboration. 2020 cms data preservation, re-use and open access policy. <https://opendata.cern.ch/record/415>, 2023. Accessed: (29-12-2023).
- [27] Cern open data portal. <http://opendata.cern.ch/docs/about>. Accessed: (29-12-2023).
- [28] Twiki CERN. Miniaod analysis documentation. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookMiniAOD2015>. Accessed: (29-12-2023).
- [29] Marco Peruzzi, Giovanni Petrucciani, Andrea Rizzi, and for the CMS Collaboration. The nanoaod event data format in cms. *Journal of Physics: Conference Series*, 1525(1):012038, apr 2020.
- [30] Twiki CERN. Cmssw application framework. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSSWFramework>. Accessed: (29-12-2023).
- [31] The CMS Collaboration. Intro to cmssw. <https://cms-opendata-workshop.github.io/workshop2022-lesson-cmssw/01-introduction/index.html>. Accessed: (29-12-2023).
- [32] Pamela Llerena, Alejandro Gomez, and Edgar Carrera. Repository for four top analysis using cms open data. <https://github.com/pamelallerena/FourTopsCMSOpenData>, 2024.
- [33] ABP Computing at CERN. Htcondor batch system. [https://abpcomputing.web.cern.ch/computing\\_resources/cernbatch/](https://abpcomputing.web.cern.ch/computing_resources/cernbatch/). Accessed: (29-12-2023).

- [34] Pushpa Bhat. Multivariate analysis methods in *Annual Review of Nuclear and Particle Science*, 61:281–309, 11 2011.
- [35] Alan S. Cornell, Wesley Doorsamy, Benjamin Fuks, Gerhard Harmen, and Lara Mason. Boosted decision trees in the era of new physics: a smuon analysis case study. *Journal of High Energy Physics*, 2022(4), April 2022.
- [36] Tommaso Isolabella. A machine learning approach to particle physics data analysis: the process  $j/\psi \rightarrow g p pbar$ . [https://fse.studenttheses.ub.rug.nl/20112/1/mPHYS\\_2019\\_IsolabellaT.pdf](https://fse.studenttheses.ub.rug.nl/20112/1/mPHYS_2019_IsolabellaT.pdf). [Online; accessed 16-August-2020].
- [37] Coadou, Yann. Boosted decision trees and applications. *EPJ Web of Conferences*, 55:02004, 2013.
- [38] HSF. Advanced python tutorial 3: Multivariate analysis. <https://hsf-training.github.io/analysis-essentials/advanced-python/30Classification.html>. Accessed: (29-12-2023).
- [39] Chicaiza Luis. Análisis simplificado de procesos de alta energía con datos abiertos del compact muon solenoid (cms) : estudio de la producción de cuatro quarks top a través del canal electrón+jets, en colisiones protón-protón, utilizando datos abiertos del experimento cms correspondientes a  $2.3 fb^{-1}$  de luminosidad integrada y energía de centro de masa  $\sqrt{s} = 13$  tev. 2023.