



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 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 $\sqrt{s} = 13 \text{ TeV}$

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

LUIS ANDRÉS CHICAIZA CHUQUITARCO

andres.chicaiza@cern.ch

DIRECTOR: EDY RODRIGO AYALA AMAYA

edy.ayala@epn.edu.ec

DMQ, AGOSTO 2023

CERTIFICACIONES

Yo, LUIS ANDRÉS CHICAIZA CHUQUITARCO, 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.

Luis Andrés Chicaiza Chuquitarco

Certifico que el presente trabajo de integración curricular fue desarrollado por Luis Andrés Chicaiza Chuquitarco, 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.

Luis Andrés Chicaiza Chuquitarco

Edy Rodrigo Ayala Amaya

AGRADECIMIENTOS

Agradezco infinitamente a mi madre quien ha sido luz, amor y verdad en mi vida. Gracias a su apoyo invaluable he podido continuar con mi vida académica y personal.

A Gerardo, Naty, Lili, Magui y Willy, por su apoyo fundamental durante toda mi vida, su amor y compañía.

A Edy Ayala, por el apoyo y seguimiento en el desarrollo de este trabajo.

A Edgar Carrera por su gran colaboración, supervisión, guía y confianza, las cuales me permitieron salir de mi zona de confort y conocer nuevos enfoques para hacer ciencia.

A Alejandro Gomez, por su ayuda y colaboración en el desarrollo de este trabajo.

A Jack, Pamela, Xime, Nanda y Daniela, por su cariño y compañía, gracias a ustedes mi experiencia universitaria fue placentera y divertida.

A mi mejor amigo, Aarón, por escucharme y apoyarme en las buenas y en las malas a pesar de la distancia.

A Henry, María José, Eliana y Ramón, docentes de la Facultad de Ciencias, que siempre supieron ayudarme académicamente y además hacer de la ciencia algo divertido.

Al CERN y en particular a la colaboración CMS, por brindarme la oportunidad de continuar mi camino en la ciencia.

DEDICATORIA

“No one can go back and start a new beginning but anyone can start today and make a new ending”.

-Maria Robinson

Para mi amada Mechita, por haber sido un modelo a seguir y haberme enseñado a valorar la belleza en la naturaleza a través de la música

Con amor, para Natita, Gerardo y Lili

RESUMEN

Un análisis simplificado de la búsqueda de la producción de cuatro top quarks ($t\bar{t}t\bar{t}$) en un único canal leptónico, electrón+jets, es presentado. El objetivo es replicar los resultados con la mayor fidelidad posible en comparación con el estudio original mediante el uso del lenguaje de programación Python. Este análisis utiliza datos abiertos producidos por el experimento CMS correspondientes a 2.3 fb^{-1} de luminosidad integrada y energía del centro de masa $\sqrt{s} = 13 \text{ TeV}$. Para distinguir entre señal y background, primero se reprodujeron los cortes de selección utilizados en el análisis original. Por consiguiente, se observó concordancia entre los datos experimentales y las simulaciones Monte Carlo. Luego, para mejorar la sensibilidad con la que se realiza la distinción entre background (i.e. $t\bar{t}$, tW , W +jets) y señal ($t\bar{t}t\bar{t}$), un algoritmo basado en un *Boosted Decision Tree* fue aplicado.

Palabras clave: Señal, background, BDT, quark, jets, electrón, física de partículas, física de colisionadores, CMS, CERN, LHC, datos abiertos.

ABSTRACT

A simplified analysis on the search of the production of four top quarks ($t\bar{t}t\bar{t}$) in a single lepton channel, electron+jets, is presented. The objective is to replicate the results as faithfully as possible compared to the original publication by using the Python programming language. This analysis uses Open Data produced by the CMS experiment which corresponds to 2.3 fb^{-1} of integrated luminosity with $\sqrt{s} = 13 \text{ TeV}$. To distinguish between signal and background, first *the baseline event selection filters* are applied. Then, an agreement between experimental data and Monte Carlo simulations is showed. Next, in order to improve the performance in distinguishing background (i.e. $t\bar{t}$, tW , W +jets) and signal ($t\bar{t}t\bar{t}$), a Boosted Decision Tree algorithm has been applied.

Keywords: Signal, background, BDT, quark, jets, electron, particle physics, accelerator physics, CMS, CERN, LHC, open data

Índice general

1. Descripción del componente desarrollado	1
1.1. Objetivo general	3
1.2. Objetivos específicos	3
1.3. Alcance	3
1.4. El Modelo Estándar (Standard Model SM)	4
1.5. Colisiones protón-protón	6
1.6. Física del Top Quark	7
1.6.1. Producción de pares de top quarks	7
1.6.2. Producción de un único top quark	8
1.6.3. Producción de un bosón W más jets (W+jets)	8
1.6.4. Producción de cuatro top quarks	9
2. Metodología	10
2.1. El Gran Colisionador de Hadrones y el detector Compact Muon Solenoid	10
2.1.1. LHC	10
2.2. El detector CMS	12
2.2.1. Solenoide Magnético	13
2.2.2. Tracker	13
2.2.3. Calorímetros	14

2.2.4. Muon Chambers	14
2.3. Reconstrucción de Eventos	15
2.3.1. Reconstrucción de Trayectoria	15
2.3.2. Vértices Primarios	16
2.3.3. Algoritmo Particle-flow (PF)	16
2.3.4. Electrones	16
2.3.5. Jets	17
2.3.6. b-tagging	17
2.3.7. Momento transversal p_T faltante	18
2.4. Señal y backgrounds	19
2.4.1. Proceso de producción de cuatro top quarks	19
2.4.2. $t\bar{t}$ background	19
2.5. Estrategia de análisis y técnicas	20
2.5.1. Datos y Simulaciones	20
2.5.2. Estrategia para la búsqueda de cuatro top quarks	22
2.5.3. Filtros de evento y de objeto	22
2.6. Análisis Multivariante	24
2.6.1. Boosted Decision Trees	24
2.6.2. Variables BDT	24
3. Resultados, conclusiones y recomendaciones	26
3.1. Selección de eventos	26
3.2. Diferenciación entre Señal y Background	28
3.2.1. Actividad en el evento	28
3.2.2. Cantidad de b-jets	30
3.3. BDT	31
3.3.1. Construcción del BDT	33
3.3.2. Resultados de los clasificadores BDT	35

3.3.3. Rendimiento del clasificador BDT	36
3.4. Conclusiones y recomendaciones	37
A. Conjunto de simulaciones usadas en el análisis	39
Bibliografía	41

Índice de figuras

1.1. Representación de Feynman donde el quark top se descompone en un bosón W y un quark b con el posterior decaimiento del bosón W, ya sea leptónica o hadrónicamente.	7
1.2. Diagramas de Feynman LO para la producción de pares de quarks top. El primer diagrama es el canal de aniquilación de quarks anti-quarks, los siguientes tres diagramas representan el canal de fusión de gluones.	8
1.3. Diagramas representativos de la producción de un solo top quark en el a) canal t, b) canal s y c) canal tW.	8
1.4. Diagramas representativos de la producción de $t\bar{t}\bar{t}$ a primer orden.	9
2.1. Configuración de aceleradores en el CERN	11
2.2. Detector multipropósito CMS	12
2.3. Sistema de coordenadas del detector CMS	13
2.4. Estado final semi-leptónico de $t\bar{t}\bar{t}$	19
2.5. Estado final semi-leptónico de $t\bar{t}$	20
3.1. Las distribuciones del p_T del electrón para datos y simulaciones. En este caso se muestra el canal electrón+jets. Leading Electron hace referencia a un único electrón <i>tight</i> como ya se explicó en la Sección 2.5.2.	27

3.2. Distribuciones del H_T del electrón para datos y simulaciones. En este caso se muestra el canal electrón+jets que concuerda con el resultado original.	27
3.3. Distribuciones del N_j para datos y simulaciones. En este caso se muestra el canal electrón+jets. La cantidad de datos experimentales no es la suficiente para estudiar $N_j > 10$ y por tanto, no se pueden realizar aseveraciones sobre la presencia de eventos de $t\bar{t}\bar{t}$.	28
3.4. Distribuciones del H_T^b para datos y simulaciones. Los errores mostrados en la parte inferior se encuentran dentro del rango esperado comparados con la publicación original.	29
3.5. Distribuciones del H_T^{Rat} para datos y simulaciones. Las barras de error están en el rango esperado.	30
3.6. Distribuciones del N_M^{tags} para datos y simulaciones. En este caso se muestra el canal electrón+jets. Además, se observa que se han detectado hasta 4 b-jets en aproximadamente 10 eventos de $t\bar{t}$. Cuando existe la presencia de 2 b-jets es más probable que los mismos se hayan originado de por la producción de un par de top quarks $t\bar{t}$ como se explicó en la Sección 2.4.2	31
3.7. Distribuciones H_T^{rat} Vs. N_j para simulaciones. Se aprecian dos regiones definidas. Además, $N_j > 14$ corresponde solo a eventos de señal.	32
3.8. Distribuciones N_{tags}^M Vs. N_j para simulaciones. Se puede notar que un número de jets por encima de 14 corresponde a eventos de señal. Además, se corrobora que el numero de jets y b-jets son las variables que pueden ofrecer un mejor rendimiento para distinguir entre señal y background, ya que producen dos regiones claramente separadas.	33
3.9. Distribuciones H_T^b Vs. H_T para simulaciones. Se pueden observar dos regiones, sin embargo, no se aprecia la misma definición o desempeño que N_{tags}^M Vs. N_j ya que existen eventos de background dispersos en la región donde se encuentran eventos de señal. . .	34

3.10	Distribuciones H_T^b Vs. H_T^{ratio} para simulaciones. Se observa un buen desempeño discriminatorio entre señal y background. Es decir, producen dos regiones visiblemente separadas.	34
3.11	Distribuciones de las variables BDT para el canal electrón+jets obtenidas mediante AdaBoost	35
3.12	Distribuciones de las variables BDT para el canal electrón+jets obtenidas mediante XGBoost	36
3.13	Curvas ROC para AdaBoost y XGBoost	37

Capítulo 1

Descripción del componente desarrollado

El quark top es la partícula elemental más pesada conocida del Modelo Estándar (Standard Model SM). Debido a que esta partícula tiene un gran acoplamiento¹ con el bosón de Higgs del SM, su importancia radica en que también podría tener grandes acoplamientos con nuevas partículas hipotéticas en varios modelos más allá del Modelo Estándar (Beyond the Standard model, BSM) [1].

En este proyecto se realizó una reproducción parcial del análisis de datos correspondiente a la búsqueda de la producción de cuatro quarks top en un solo canal leptónico electrón+jets. El análisis utilizó datos abiertos correspondientes a 2.3 fb^{-1} de luminosidad integrada con $\sqrt{s} = 13 \text{ TeV}$ obtenidos por el experimento CMS en 2015.

En este proceso, los quarks y gluones tienen una carga neta de color y no pueden existir libremente debido al confinamiento de color. En consecuencia, no se observan directamente en la naturaleza sino que se unen para formar hadrones de color neutro, un proceso llamado hadronización que conduce a un chorro colimado de hadrones llamado jet. El estudio de jets es importante ya que son las firmas experimentales de quarks y gluones producidos en procesos de alta energía (colisiones protón-protón).

¹En física de partículas, el término acoplamiento se refiere a la fuerza de la interacción entre partículas. Describe el grado en que una partícula interactúa con otra a través de una fuerza o interacción específica. El acoplamiento determina la probabilidad de que ocurra la interacción.

El estudio de los procesos con baja sección eficaz que involucran al quark top son particularmente relevantes. Entre estos, la producción de cuatro quarks top ($t\bar{t}t\bar{t}$) está predicha por el SM y ha sido observada recientemente por la Colaboración CMS con una significancia que supera el famoso umbral de 5 sigma utilizado por los físicos para declarar que se ha realizado un nuevo descubrimiento [2].

En el Modelo Estándar la producción de cuatro quarks top se realiza a través de la fusión de gluones o quarks-antiquarks. Se predice que la sección eficaz de primer orden (LO) para la producción de cuatro quarks top en el LHC es extremadamente pequeña, $\sigma_{t\bar{t}t\bar{t}} \approx 9$ fb a $\sqrt{s} = 13$ TeV y de $\sigma_{t\bar{t}t\bar{t}} \approx 12$ fb en el siguiente orden (NLO) en QCD, incluyendo correcciones electrodébiles [3]. Además, en varias teorías BSM se predice un incremento o mejora en la sección eficaz $\sigma_{t\bar{t}t\bar{t}}$ de la producción de cuatro top quarks; por ejemplo, la producción de pares de gluinos en teorías de supersimetría [4, 5].

Por otro lado, la presencia de eventos $t\bar{t}t\bar{t}$ se encuentra cinco órdenes de magnitud por debajo de la producción de un par de top quarks. Esto se debe a que su tasa de producción es muy pequeña y, en parte, debido a que la topología experimental de este proceso es muy compleja, requiriendo la reconstrucción de hasta 12 objetos físicos pertenecientes a los decaimientos de los cuatro top quarks.

Esta búsqueda también se ve afectada por procesos con topologías similares a $t\bar{t}t\bar{t}$ dentro del SM, también llamados fuentes de background (o procesos de fondo). Es decir, existen procesos del SM que pueden producir eventos con topologías semejantes a $t\bar{t}t\bar{t}$ y, por lo tanto, son considerados como backgrounds. En particular, debido al abrumador background de la producción de $t\bar{t}$ +jets, el cual tiene $\sigma_{t\bar{t}} \approx 831$ pb, la observación directa de los eventos que llevan a una medición de la sección eficaz $\sigma_{t\bar{t}t\bar{t}}$ es poco probable con el set de datos de 2015.

Se sabe que cada quark top decae principalmente en un quark bottom y un bosón W, y cada bosón W se descompone en leptones o quarks. Como resultado, el estado final $t\bar{t}t\bar{t}$ contiene jets, principalmente de la hadronización de quarks ligeros (u, d, s, c) y quarks b (b-jets). También puede contener leptones cargados aislados y momento transversal faltante² que surge de los neutrinos emitidos [6].

²El momento transversal faltante se explica en la Sección 2.3.7.

El canal de leptones+jets es uno de los más sensibles a los eventos $t\bar{t}t\bar{t}$ producidos con propiedades cinéticas estudiadas por el SM. Es por esta razón que este proyecto se centra en la búsqueda de cuatro quarks top en el canal de un único leptón cargado, en este caso, el canal de electrón+jets.

1.1. Objetivo general

Reproducir parcialmente el análisis de la producción de cuatro quarks top, en donde uno de ellos decae hacia un electrón + jets, mediante el uso de datos abiertos de la Colaboración CMS del CERN con energía de centro de masa de 13 TeV.

1.2. Objetivos específicos

- Definir las bases de datos del proceso de producción de cuatro tops.
- Aplicar los filtros necesarios a las diferentes variables para desarrollar software requerido para el análisis.
- Comparar los resultados de las simulaciones Monte Carlo con los datos experimentales.
- Realizar una simplificación de técnicas de análisis multivariable (como Boosted Decision Trees, BDT) para efectos de comparación.

1.3. Alcance

Mediante el uso de datos abiertos de la colaboración CMS se replicó parcialmente el análisis descrito en el artículo [7], consistente en la búsqueda de la producción de cuatro quarks top en el canal de un solo leptón electrón+jets. El análisis utilizó datos equivalentes a 2.3 fb^{-1} de luminosidad integrada con $\sqrt{s} = 13 \text{ TeV}$ registrados por el experimento CMS en el año 2015.

Finalmente, se desarrolló el código informático necesario usando el lenguaje de código abierto Python y librerías como uproot y coffea. De este modo, se proporciona un ejemplo funcional de análisis de datos en física de partículas

para facilitar el uso de CMS Open Data tanto para los colaboradores de CMS como para investigadores externos.

1.4. El Modelo Estándar (Standard Model SM)

El modelo estándar de la física de partículas es la teoría actualmente aceptada para describir los componentes básicos conocidos del universo. Incluye seis quarks y seis leptones (y sus antipartículas) y bosones de gauge. También incorpora el bosón de Higgs, que surge del mecanismo de Brout-Englert-Higgs que describe el origen de la masa de los quarks, leptones y bosones gauge.

Sin embargo, todavía hay muchas preguntas sin respuesta sobre el universo; por ejemplo, ¿cómo incluir la gravedad en el SM?, ¿cuál es la materia oscura en el universo que afecta las curvas de rotación de las galaxias y provoca lentes gravitacionales donde no hay materia bariónica presente?, ¿cuál es el origen de la asimetría materia-antimateria?, etc. El Gran Colisionador de Hadrones, LHC, del CERN, tiene como objetivo responder a algunas de estas preguntas mediante el estudio de posibles evidencias de nueva física que pueden surgir en torno a la escala electrodébil.

En este trabajo se estudia el proceso de la producción de cuatro quarks top. Aunque el SM predice este proceso, es extremadamente raro. Sin embargo, en el presente año (2023), CMS publicó la observación de la producción de cuatro top quarks en un canal di-leptónico y multileptónico del mismo signo. Esta medición es una prueba de la consistencia de la predicción hecha por el SM [8]. La colaboración continuará impulsando la precisión de esta medición en el RUN3 del LHC (2022-actualidad).

Por otro lado, nuevos modelos en la física predicen estados finales que contienen cuatro quarks top, incluida la supersimetría, modelos con dimensiones adicionales, entre otros. La producción de cuatro quarks top en el SM es una referencia importante para estos nuevos modelos. La tasa de producción de cuatro quarks top podría incrementar bajo mecanismos de nuevos modelos físicos.

El Modelo Estándar proporciona actualmente las mejores predicciones y explicaciones para el comportamiento de las partículas elementales. Sus propiedades se miden con gran precisión en muchos experimentos. No se han

observado desviaciones significativas de las predicciones teóricas hasta el momento. Sin embargo, el modelo estándar se limita a una energía en el rango de TeV y tiene varias deficiencias para actuar como una teoría completa (teoría del todo). Algunas limitaciones del SM son: a) la gravedad no está incluida en este modelo, b) solo puede explicar la materia visible, que constituye aproximadamente solo el 5% de la materia en el universo, c) no puede resolver el problema de la jerarquía, entre otros.

Dado que el LHC, en su actual funcionamiento denominado RUN3, tiene acceso a un nuevo dominio de energía (13.6 TeV), se espera que el Modelo Estándar revele sus problemas, es decir, observar nuevas partículas que indiquen la naturaleza de una teoría subyacente en donde el Modelo Estándar aparece como aproximación de baja energía.

El SM [9] es una teoría que describe las partículas elementales y sus interacciones. Cada quark y leptón, y sus respectivas antipartículas, están organizados en tres generaciones. Cada generación contiene partículas más pesadas que la anterior, como se ve en la Tabla 1.1.

Toda la materia ordinaria de la Tierra está formada por partículas de la primera generación, es decir, los quarks up y down que forman protones y neutrones, y combinados con electrones forman átomos.

Generación	Quarks			Leptones		
	Sabor	Carga Eléctrica	Masa (MeV)	Sabor	Carga Eléctrica	Masa (MeV)
I	u	2/3	$2.2^{+0.6}_{-0.4}$	e	-1	0.511
	d	-1/3	$4.7^{+0.5}_{-0.4}$	ν_e	0	$< 2 \times 10^{-6}$
II	c	2/3	$(1.27 \pm 0.03) \times 10^3$	μ	-1	105.66
	s	-1/3	96^{+8}_{-4}	ν_μ	0	< 0.19
III	t	2/3	$(173.21 \pm 0.87) \times 10^3$	τ	-1	1776.86 ± 0.12
	b	-1/3	$4.18^{+0.04}_{-0.3}$	ν_τ	0	< 18.2

Tabla 1.1: Quarks y Leptones del Modelo Estándar [10].

El sector leptónico consta de leptones con carga y sin carga eléctrica que pueden interactuar a través de la fuerza electromagnética y débil. En el caso de los neutrinos, interactúan solo a través de la fuerza débil, no tienen carga y tienen masas muy pequeñas.

Los quarks interactúan a través de las fuerzas electromagnética, débil o

Bosón Gauge	Fuerza	Carga Eléctrica	Masa (GeV)	Spin	Rango (m)
Fotón (γ)	electromagnética	0	0	1	∞
W^\pm	débil	± 1	80.385 ± 0.015	1	10^{-18}
Z	débil	0	91.1876 ± 0.0021	1	10^{-18}
gluón (g)	fuerte	0	0	1	10^{-15}

Tabla 1.2: Bosones Gauge del Modelo Estándar [10].

fuerte. Cada quark tiene carga eléctrica fraccionaria, como se ve en la Tabla 1.1, y además carga de color (i.e. rojo, verde, azul o RGB), donde los tres colores combinados pueden formar un estado singlete de color. Una combinación de quarks con un color y su anticolor también puede formar un estado singlete. El fenómeno del confinamiento de color significa que los quarks solo se pueden encontrar en estados singlete de color, esto es, en bariones o mesones.

Finalmente, los portadores de fuerza consisten en bosones gauge de espín entero, como se ve en la Tabla 1.2. Los fotones y los bosones Z median interacciones electrodébiles neutras, mientras que los bosones W median interacciones débiles cargadas. Los gluones son los mediadores de la interacción fuerte y pueden tener 8 tipos diferentes de carga de color.

1.5. Colisiones protón-protón

En el Gran Colisionador de Hadrones³, las partículas involucradas en las colisiones son protones, que son partículas compuestas complejas que consisten en tres quarks, dos quarks up y un quark down; y gluones que intercambian la fuerza fuerte. El protón también contiene “sea quarks”, que son pares de quarks y antiquarks que aparecen y desaparecen rápida y continuamente dentro del protón debido a la división del campo de color del gluón.

Los protones en el LHC pueden: a) no interactuar en absoluto y continuar acelerándose alrededor del anillo, b) interactuar a través de una dispersión suave donde los productos usualmente viajan principalmente en la dirección del haz de protones, c) participar en una interacción fuerte donde dos partones⁴ dentro de los protones tienen una colisión de alta energía en la que los

³El LHC se aborda con más detalle en la Sección 2.1

⁴Término usado para referirse a los constituyentes fundamentales de hadrones como el protón (i.e. una mezcla de quarks y gluones).

productos viajan transversalmente al haz. En este último caso, los partones restantes que no han participado en la interacción de alta energía se hadronizan y forman lo que se conoce como el evento subyacente (underlying event UE).

1.6. Física del Top Quark

El top quark es el quark más pesado con una masa de $173.21 \pm 0.51 \pm 0.71$ GeV. El quark top es el único quark que decae predominantemente antes de que pueda formar cualquier estado ligado, debido a su corta vida media de 5×10^{-25} segundos [10] y, por lo tanto, es el único quark que puede estudiarse por sus propiedades de polarización y espín.

El principal modo de decaimiento de los quarks top es un quark bottom y un bosón W, como se muestra en la Figura 1.1. Este modo tiene una probabilidad de ocurrencia del $99.8 \pm 3.8(\text{exp.}) \pm 1.6(\text{teo.})\%$ [11].

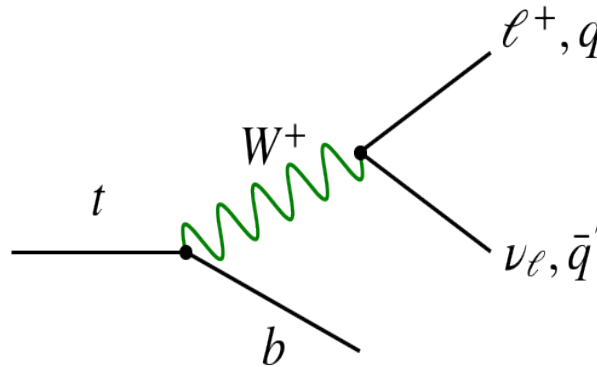


Figura 1.1: Representación de Feynman donde el quark top se descompone en un bosón W y un quark b con el posterior decaimiento del bosón W, ya sea leptónica o hadrónicamente.

1.6.1. Producción de pares de top quarks

Las primeras observaciones de top quarks se realizaron en análisis de la producción de pares de top quarks ($t\bar{t}$), ya que este es el mecanismo que domina en la producción de top quarks en los colisionadores. La Figura 1.2 muestra los mecanismos de producción de primer orden (leading order, LO) a través de la fusión de gluones y la aniquilación de quarks anti-quarks [12].

Hay tres posibles estados finales para este proceso, dependiendo de cómo decaiga cada quark top, como se muestra en la Figura 1.1: hadrónico, donde ambos bosones W del top decaen en un quark y un antiquark; semi-leptónico, donde un bosón W decae en $q\bar{q}'$ y un bosón W se descompone en un leptón y un neutrino; y dileptónico donde ambos bosones W se descomponen en un leptón y un neutrino cada uno.

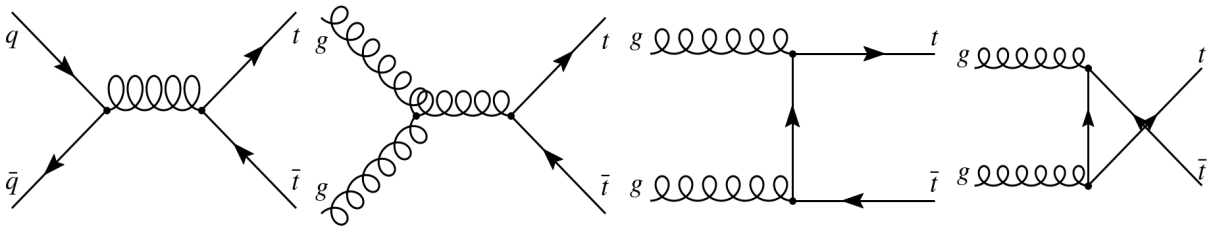


Figura 1.2: Diagramas de Feynman LO para la producción de pares de quarks top. El primer diagrama es el canal de aniquilación de quarks anti-quarks, los siguientes tres diagramas representan el canal de fusión de gluones.

1.6.2. Producción de un único top quark

La producción de un solo quark top es mucho más rara que la producción de $t\bar{t}$ en el SM. Puede ocurrir mediante la aniquilación de $q\bar{q}$, la fusión de un gluon g con un quark q o la fusión de gluones. Otros ejemplos se muestran en la Figura 1.3. En esta figura se muestran el canal t (izquierda), el canal s (centro) y la producción asociada con un bosón W, canal tW (derecha) [13].

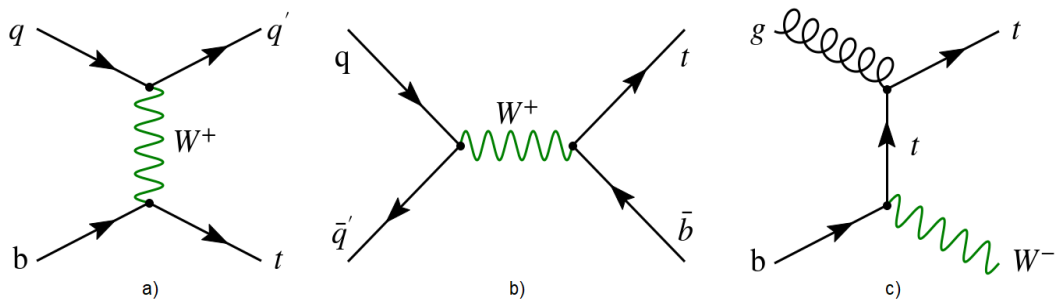


Figura 1.3: Diagramas representativos de la producción de un solo top quark en el a) canal t, b) canal s y c) canal tW.

1.6.3. Producción de un bosón W más jets (W+jets)

La producción W+jets se origina ya sea por quarks o por gluones. La relevancia de este proceso se debe a que su estado final es similar a $t\bar{t}t\bar{t}$ y por esta

razón es considerado como un background.

El bosón W puede decaer leptónicamente, es decir tener un estado final con un electrón o un muón, con su correspondiente neutrino. También puede decaer hadrónicamente generando jets a partir de quarks. El estado final de W +jets puede contener jets y leptones lo cual hace que se asemeje al estado final de $t\bar{t}\bar{t}$.

1.6.4. Producción de cuatro top quarks

Después de haber introducido las principales fuentes de background es necesario introducir el evento de interés, i.e. la producción de cuatro quarks top ($t\bar{t}t\bar{t}$). Este evento ocurre predominantemente a través de la fusión de gluones ($\sim 90\%$), como se ve en la Figura 1.4 [14].

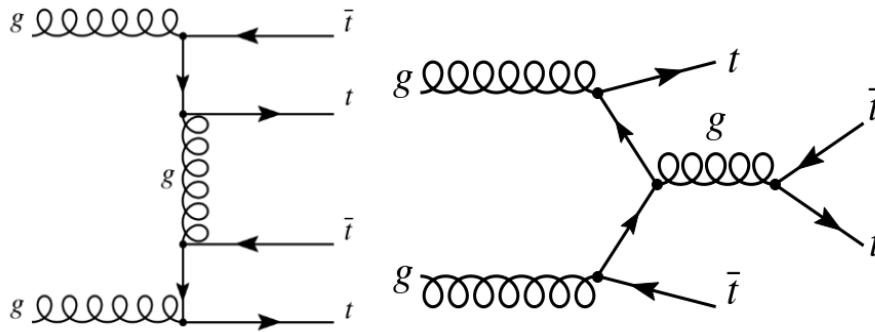


Figura 1.4: Diagramas representativos de la producción de $t\bar{t}t\bar{t}$ a primer orden.

El mecanismo de producción de $t\bar{t}t\bar{t}$ se da predominantemente a través de QCD⁵. El decaimiento para cada quark top producido se repite de la misma manera que para los procesos de background anteriores que se introdujeron, es decir, cada bosón W^\pm puede decaer en \bar{q}, q o l, ν_l como en la Figura 1.1.

⁵Cromodinámica cuántica o QCD, es una teoría cuántica de campos que describe la interacción fuerte.

Capítulo 2

Metodología

2.1. El Gran Colisionador de Hadrones y el detector Compact Muon Solenoid

El LHC se encuentra en la frontera franco-suiza cerca de la ciudad de Ginebra-Suiza, aproximadamente a 100 m bajo tierra en el laboratorio CERN (Conseil Européen pour la Recherche Nucléaire). El LHC es un acelerador de partículas sincrotrón de 26.7 km de perímetro con cuatro puntos de interacción, en donde se ubican cuatro detectores.

El presente trabajo utiliza los datos recopilados por el detector Compact Muon Solenoid (CMS) en el año 2015 descrito en la Sección 2.2. En esta sección se explican brevemente los detalles del funcionamiento y propósito de cada uno de los subdetectores del CMS. Los algoritmos que se utilizan para reconstruir partículas a partir de la información proporcionada por los subdetectores se indican en la Sección 2.3.

En el LHC también existen otros experimentos como ATLAS, LHCb y el detector ALICE, los cuales no son descritos en el presente trabajo.

2.1.1. LHC

El LHC acelera haces de protones que circulan en direcciones opuestas. Los protones son extraídos de una fuente de hidrógeno. Luego, un campo eléctrico

es usado para excitar los electrones del átomo de hidrógeno haciendo que se separen del mismo.

El LHC del CERN, cuenta con aceleradores que funcionan en serie, es decir una sucesión de máquinas que aceleran partículas. Cada acelerador incrementa la energía del haz de partículas, luego el haz es inyectado en la máquina subsecuente como se puede observar en la Figura 2.1.

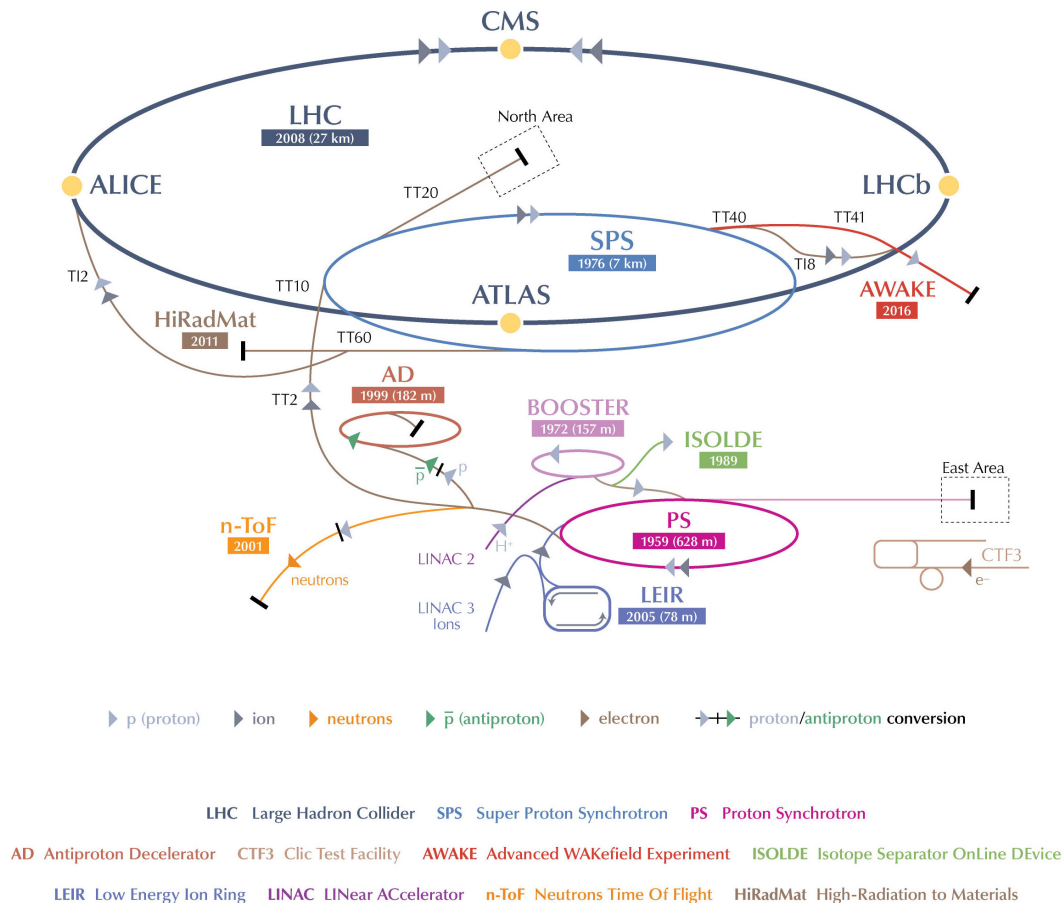


Figura 2.1: Configuración de aceleradores en el CERN [15]

Los protones son acelerados en un acelerador lineal llamado LINAC2, donde alcanzan una energía de 50 MeV. Luego, desde el LINAC2, los protones son acelerados por el Proton Synchrotron Booster (PBS) hasta los 1.4 GeV. Posteriormente, son acelerados en el Proton Synchrotron (PS) alcanzando los 25 GeV. El siguiente paso es acelerarlos en el Super Proton Synchrotron (SPS) hasta los 450 GeV, este es el paso final previo a la inyección de los protones en el anillo del LHC que es donde la colisión final de alta energía se llevará a cabo.

2.2. El detector CMS

CMS es un detector hermético con un gran solenoide magnético que curva la trayectoria de partículas cargadas cada vez que atraviesan el detector. El subdetector más cercano al haz de protones es el Silicon Tracker el cual realiza medidas precisas de la posición de las partículas. A continuación se encuentran dos calorímetros, uno para partículas electromagnéticas y otro para partículas hadrónicas. Mayores detalles pueden ser encontrados en la referencia [16].

Todos estos detectores están contenidos dentro del solenoide magnético formando capas, Figura 2.2. Por otro lado, también están las cámaras de muones (muon chambers) las cuales están fuera del solenoide. Este detector está diseñado para detectar muones ya que son mucho más penetrantes que otras partículas.

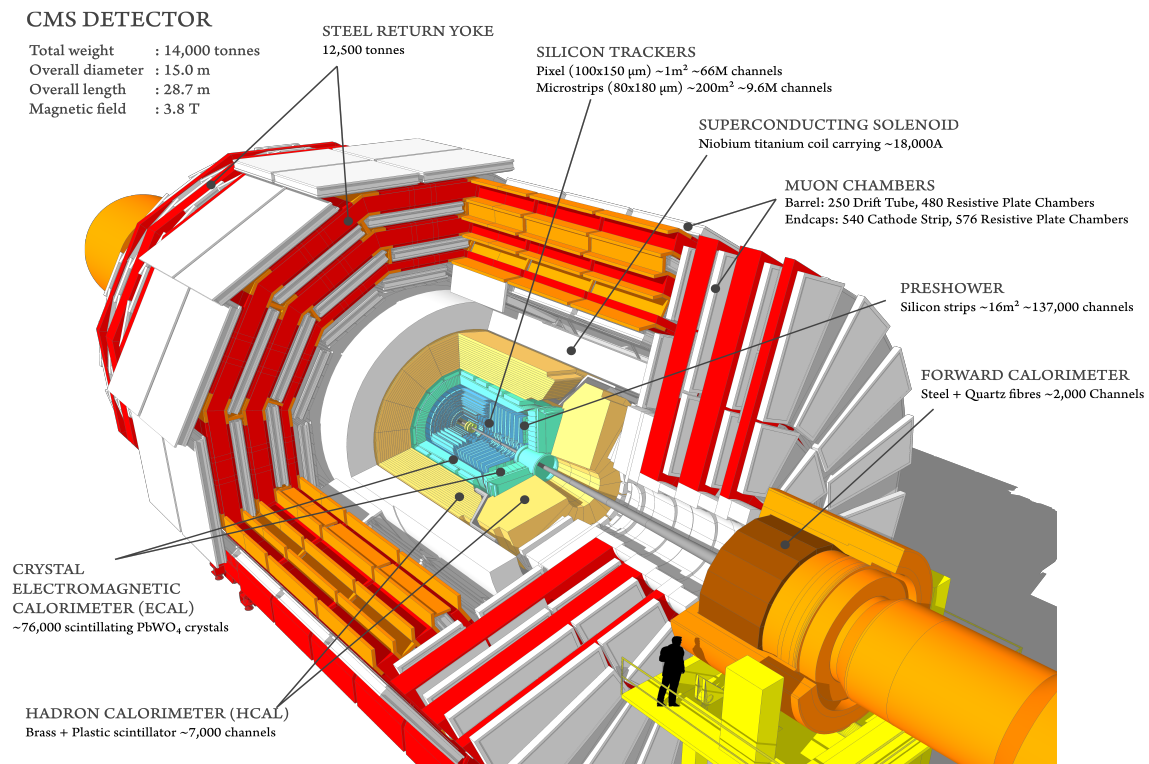


Figura 2.2: Detector multipropósito CMS [17]

El sistema de coordenadas cilíndricas del CMS se define de la siguiente manera: el eje x apunta hacia el centro del anillo LHC, el eje y apunta hacia arriba y el eje z apunta a lo largo de la línea del haz de protones en el sentido contrario a las agujas del reloj. El ángulo azimutal ϕ se mide en el plano (x, y)

en el sentido de las agujas del reloj desde el eje x (Figura 2.3). El ángulo polar θ se mide en el sentido de las agujas del reloj desde el eje z . A menudo, la pseudorapidez (ecuación 2.1), se utiliza en lugar de θ .

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right] \quad (2.1)$$

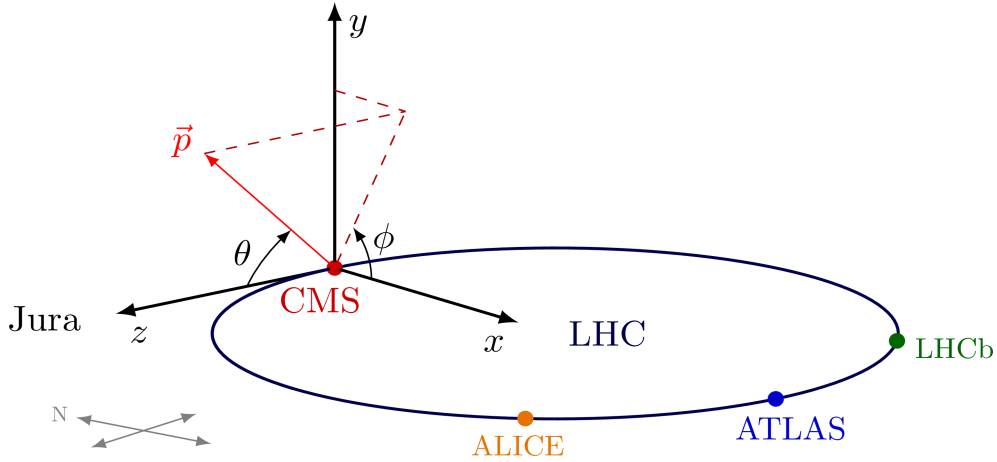


Figura 2.3: Sistema de coordenadas del detector CMS [18]

2.2.1. Solenoide Magnético

El solenoide magnético superconductor se encuentra en el núcleo del CMS y fue diseñado para producir un campo magnético de 4 T, aunque históricamente funciona a 3.8 T.

Este magneto produce un campo magnético homogéneo el cual afecta a las partículas cargadas haciendo que sigan trayectorias curvas transversales a la dirección del haz de protones.

2.2.2. Tracker

El Silicon Tracker o Tracker se encuentra dentro del solenoide superconductor. Los detectores de silicio del tracker son usados para obtener una respuesta rápida requerida para construir de manera precisa y confiable las trayectorias de las partículas generadas en el punto o vértice de colisión¹.

¹A este punto de interacción se lo conoce como vértice primario (primary vertex)

Por otra parte, la reconstrucción de vértices secundarios también es de particular importancia para la identificación de jets que se originan de quarks pesados como los bottom quarks. Esto es de suma importancia para distinguir estados finales que contienen top quarks.

2.2.3. Calorímetros

Para la medición de energía el CMS cuenta con dos calorímetros. El primero es el calorímetro electromagnético (ECAL) que mide la energía de electrones y fotones.

También se encuentra el calorímetro hadrónico HCAL, que recubre al ECAL. El HCAL mide la energía de partículas hadrónicas que sobrepasan el ECAL. Esto se debe a que los hadrones interactúan más fuertemente con la materia por lo que pueden atravesar mayor distancia antes de depositar toda su energía.

Por una parte, el ECAL está hecho de tungstato de plomo que con un poco de oxígeno se convierte en un material centellante. Es decir, cuando electrones y fotones atraviesan este material, se produce luz proporcional a la energía de la partícula.

Por otra parte, el HCAL posee varias capas hechas de materiales centellantes plásticos. De igual manera, cuando una partícula lo atraviesa se emite luz haciendo posible su detección.

2.2.4. Muon Chambers

Estos detectores se encargan particularmente de la detección de muones. Estas partículas son parecidas a los electrones ya que tienen su misma carga eléctrica, sin embargo pesan 200 veces más que un electrón.

Debido a que los muones pueden penetrar muchos más metros de material sin perder tanta energía, a diferencia de otras partículas, no son detenidos por los calorímetros del CMS. Por esta razón, las cámaras para detectar muones son ubicadas en la capa más externa del experimento.

2.3. Reconstrucción de Eventos

En el contexto de análisis de datos en física de partículas, un objeto físico es la reconstrucción de una o varias partículas. Estos objetos se obtienen de las medidas realizadas por CMS. Las mediciones hechas son combinadas y analizadas de manera que se puede reconstruir e identificar partículas.

En la Sección 2.2, cada uno de los subdetectores de CMS ha sido descrito para entender cómo las partículas interactúan con él. El siguiente paso para la reconstrucción de eventos es la combinación de cada lectura hecha por cada subdetector con el objetivo de reconstruir las partículas resultantes de una colisión protón-protón.

La captura de la interacción se denomina *evento*. Un evento también contiene otras interacciones, denominadas Pile-up² que surgen de partículas de otras colisiones simultáneas que no son de interés.

Usualmente una partícula atraviesa más de un subdetector. Por esta razón, se combinan los resultados de cada uno de ellos para reconstruirlas e identificarlas. Esto se consigue mediante el uso del algoritmo *particle-flow* (PF) descrito en la Sección 2.3.3. Los objetos que pueden ser reconstruidos mediante PF, tal como electrones y jets, serán discutidos en las secciones 2.3.4, 2.3.5 respectivamente.

2.3.1. Reconstrucción de Trayectoria

Aproximadamente 1000 partículas cargadas atraviesan el Silicon Tracker de CMS por cada conjunto de protones que colisionan, lo cual produce ≈ 20 interacciones de alta energía. Existen algoritmos diseñados para relacionar los choques de las partículas con el tracker y así reconstruir su trayectoria. De esta manera, también se obtiene información de la carga y momento de la partícula.

La reconstrucción de las trayectorias ayuda a ubicar el vértice de interés también conocido como vértice primario (primary vertex).

²Pile-up hace referencia a la existencia de varias interacciones simultáneas en el detector. Es decir, en adición a una interacción primaria de interés, existen otras colisiones secundarias.

2.3.2. Vértices Primarios

Los vértices primarios son los puntos en los cuales ocurrió la colisión de alta energía. Esto es diferente a los vértices secundarios que se originan del decaimiento de partículas subsecuentes originadas en la colisión.

Las coordenadas de los vértices primarios se obtienen a partir de la suma del momento al cuadrado de todas las trayectorias que se consideran originadas en un vértice de interés. El vértice cuya suma sea la mayor es tomado como el evento más interesante, i.e., vértice primario.

2.3.3. Algoritmo Particle-flow (PF)

Este algoritmo combina la información de todos los subdetectores, descritos en la Sección 2.2, con el objetivo de mejorar la reconstrucción de partículas de estado final tales como electrones, muones, fotones, hadrones neutros y cargados. Usando esta información, objetos más complicados, como jets, también pueden ser reconstruidos.

El algoritmo PF reconstruye objetos empezando por los más fáciles, i.e. muones. Los objetos más difíciles de reconstruir, como hadrones neutros, son los últimos en ser analizados ya que sus propiedades se ven limitadas por los objetos previamente reconstruidos [19].

2.3.4. Electrones

Los electrones normalmente pierden una gran parte de su energía en el Silicon Tracker debido al efecto Bremsstrahlung. Debido a esto, los electrones son uno de los objetos más difíciles de reconstruir ya que la radiación Bremsstrahlung debe tomarse en cuenta para las mediciones de su momento. Además, los fotones de esta radiación producen pares de electrón-positrón creando electrones secundarios que deben distinguirse de la señal del electrón original producido en el proceso de alta energía.

Se pueden definir puntos de funcionamiento (*working points*) para partículas candidatas a electrones. Electrones *tight* son usados cuando se requieren electrones como parte del proceso principal (también llamado señal), es decir, es el criterio de selección más estricto. Por otro lado, también existen crite-

rios más flexibles o menos estrictos (*loose*) con el propósito de vetar electrones adicionales en caso de que se desee realizar una selección más estricta. La selección *loose* contendrá más electrones pero con una pureza menor, es decir, su criterio de selección es menos estricto que *tight* [20].

2.3.5. Jets

Cuando partones como quarks y gluones se hadronizan, forman un conjunto de hadrones neutros y cargados que se mueven aproximadamente en la misma dirección que el partón. Todo este conjunto de partículas puede ser considerado como un solo objeto llamado jet.

En particular, en este trabajo se usa el parámetro $R = 0.4$ denominado radio. En este sentido, en el algoritmo *anti - κ_T* [21], se elige una partícula i inicial del conjunto de hadrones que se quiere analizar, luego se calcula la distancia d_{iB} entre el haz B de hadrones y la partícula i . Si una partícula i no tiene otras partículas cercanas a una distancia R , entonces d_{iB} será menor que d_{ij} para cualquier otra partícula j , luego la partícula i es catalogada como un jet.

Los criterios de identificación de un jet son aplicados para eliminar jets falsos que aparecen de lluvias de electrones debido al efecto Bremsstrahlung. Debido a esto, en este trabajo se requiere $|\eta| < 2.5$, $p_T > 30$ GeV y una separación al electrón *loose* más cercano de $\Delta R > 0.4$.

2.3.6. b-tagging

La presencia de jets en un evento indica que las partículas que emergen de una colisión incluyen quarks y gluones. Ser capaz de identificar o etiquetar qué sabor (flavour) de quark se hadronizó en el detector es de suma utilidad para el análisis de datos.

En particular, para búsquedas de estados finales que contienen cuatro top quarks, la posibilidad de identificar b-quarks originados del decaimiento de top quarks es de alto beneficio, ya que permite discriminar entre la señal de interés ($t\bar{t}t\bar{t}$) y backgrounds.

La lluvia de partículas proveniente de la hadronización de b-quarks contiene mesones B , bariones Λ_B y otros más. El tiempo de vida media de estas

partículas les permite viajar más lejos en el detector que mesones y bariones formados por light flavour quarks (u,d,s).

Por otro lado, para identificar o etiquetar (tag) b-jets (jets originados de b-quarks), se usa el algoritmo CSV (Combinated Secondary Vertex) el cual asigna un valor discriminador entre 0 y 1, donde mayores valores son consistentes con b-quark jets.

Para, este trabajo, donde $\sqrt{s} = 13$ TeV, se usó el algoritmo mejorado CSVv2 (Combinated Secondary Vertex version 2) [22]. Los *working points* o discriminadores para identificación errónea de b-quarks se encuentran en la Tabla 2.1

\sqrt{s} (TeV)	Nombre	Working Point	Eficiencia (%)	Identificación errónea (%)
13	CSVv2L	0.46	82	11.5
	CSVv2M	0.8	67	1.4
	CSVv2	0.935	47	0.15

Tabla 2.1: Working points para el b-tagging y sus eficiencias

2.3.7. Momento transversal p_T faltante

El momento transversal (p_T) hace referencia a la componente transversal del momento que se encuentra perpendicular a la dirección de movimiento de la partícula (Figura 2.3).

Por otro lado, debido a que no es posible detectar neutrinos y partículas potencialmente BSM, ya que su interacción con la materia es extremadamente baja, su existencia puede ser inferida al examinar la suma del momento de partículas ubicadas en el plano transversal al detector. Este plano se define como transversal a la línea del haz de protones.

Asumiendo que el momento total en el plano transversal es cero, un desbalance en la suma del momento p_T de las partículas detectadas, es considerado como momento transversal faltante.

2.4. Señal y backgrounds

2.4.1. Proceso de producción de cuatro top quarks

Es conocido que el estado final de un solo leptón es el más probable ($\approx 39\%$), lo cual hace que el estudio de $t\bar{t}t\bar{t}$ sea más favorable a través de este canal. En la Figura 2.4 se observa el diagrama detallado del estado final de la producción de cuatro top quarks. En adelante, a la producción de cuatro top quarks se la llamará señal.

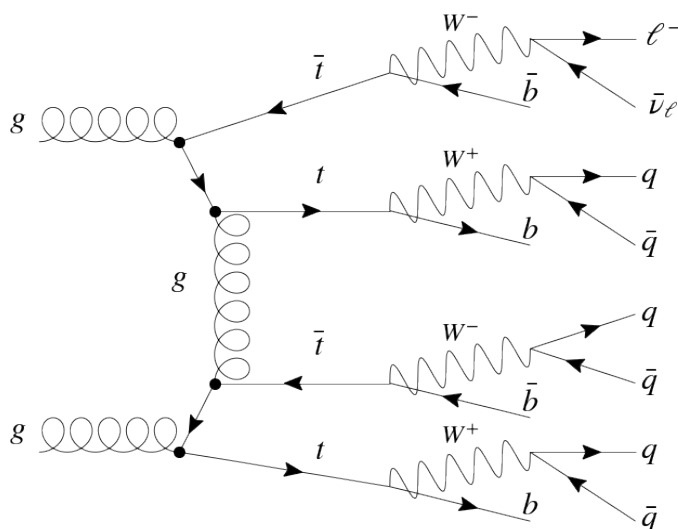


Figura 2.4: Estado final semi-leptónico de $t\bar{t}t\bar{t}$

2.4.2. $t\bar{t}$ background

El principal background³ para $t\bar{t}t\bar{t}$ es la producción de $t\bar{t}$. En este caso, es de particular interés el requerimiento de dos b-jets y un leptón (electrón) en el estado final también llamado semi-leptónico. La Figura 2.5 muestra el estado final semi-leptónico para $t\bar{t}$.

En el caso del canal con un único leptón, el estado final semi-leptónico de $t\bar{t}$ contribuye al background ya que, instrumentalmente, es muy parecido al estado final de $t\bar{t}t\bar{t}$. Existen dos top quarks que producen jets, incluyendo b-jets y un leptón. Además, puede haber jets que caigan fuera del rango de

³W+jets y tW también son fuentes de background pero su contribución es mucho menor comparadas con $t\bar{t}t\bar{t}$.

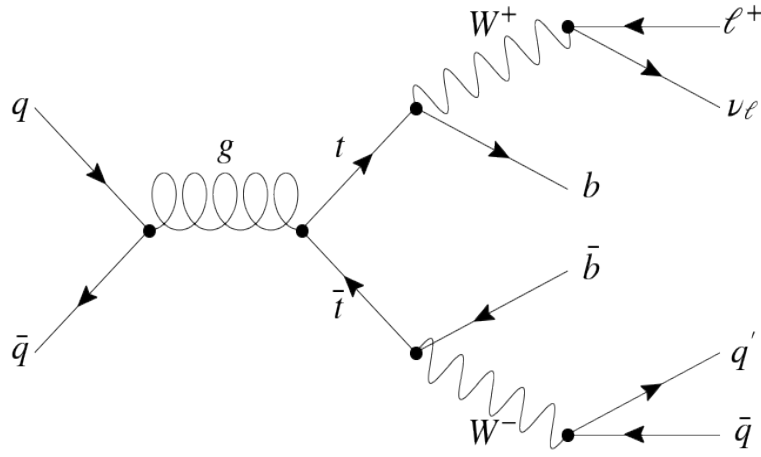


Figura 2.5: Estado final semi-leptónico de $t\bar{t}$

aceptación del detector o jets que pueden unirse, lo cual puede reducir la multiplicidad de jets de $t\bar{t}t\bar{t}$ haciendo que se parezca más a un proceso de $t\bar{t}$.

2.5. Estrategia de análisis y técnicas

El LHC es visto como una fábrica de top quarks debido a la sección eficaz grande para la producción de $t\bar{t}$, a $\sqrt{s} = 13$ TeV, la cual es de 831 pb.

En esta sección, se detalla el proceso de selección de eventos y objetos que luego fue aplicado a través del uso del framework coffea desarrollado en Python.

2.5.1. Datos y Simulaciones

Los datos utilizados en este análisis provienen de los conjuntos de datos *SingleElectron* de 2015 con grupos de protones con un espaciado de 25 ns. Para este análisis, la luminosidad integrada corresponde a ≈ 2256.38 pb $^{-1}$ para el conjunto de datos de electrones, cuyos detalles se pueden encontrar en la Tabla 2.2. Todas las muestras simuladas se produjeron como parte del conjunto denominado *RunI-IFall15MiniAODv2*.

Todos los datos y simulaciones se han procesado utilizando la versión 76X del software CMSSW [23]. En la Tabla A.2 se proporciona una lista completa de las especificaciones de las muestras de simulación. Simulaciones adicionales se muestran en la Tabla A.1.

Además, partiendo de los archivos miniAOD de CMS Open Data, se crearon archivos nanoAOD. El formato nanoAOD es comúnmente usado en CMS. Los pasos para producir archivos nanoAOD pueden encontrarse en el repositorio [24]. También se usaron simulaciones Monte-Carlo procesadas por *Analysis Grand Challenge* [25].

Set de datos	Luminosidad Int.	Global Tag
/SingleElectron/Run2015D-16Dec2015-v1	2256.38 pb ⁻¹	76X dataRun2 v15
Total	2256.38 pb ⁻¹	

Tabla 2.2: Nombre del conjunto de datos y luminosidad integrada para las muestras de datos MiniAODv2 utilizadas para el canal e + jets [26].

De acuerdo con la tabla A.2 (simulaciones Monte Carlo):

- Los conjuntos de datos cuyo nombre comienza como $/TT\dots$, son simulaciones de la producción de pares de top quarks (como se observa en la sección 1.6.1).
- Los conjuntos de datos cuyo nombre comienza como $/W\text{JetsToLNu}\dots$, son simulaciones del proceso específico en el que se produce un bosón W en asociación con un leptón (e.g., electrón, muón) y un neutrino (Sección 1.6.3).

El proceso “WJetsToLN” se estudia comúnmente en experimentos de física de alta energía para comprender la producción y las propiedades de los bosones W y sus interacciones con otras partículas. El término “Jets” se refiere a la emisión de quarks y gluones que acompañan al bosón W en su estado final, formando chorros colimados de partículas conocidas como jets. El término “LNu” indica la presencia de un leptón cargado y su neutrino asociado.

- Los conjuntos de datos cuyo nombre comienza como $/ST_{tW}_{top}\dots$ o $/ST_{tW}_{antitop}\dots$, son simulaciones de la producción de un único top quark del canal tW (como se observa en la Sección 1.6.2).
- Los conjuntos de datos cuyo nombre comienza como $/TTTT\dots$, son simulaciones de la producción de cuatro top quarks (como se observa en la Sección 1.6.4).

Todo el conjunto de simulaciones está codificado en un archivo *.json*. El código desarrollado para este trabajo puede ser encontrado en el siguiente repositorio [27].

2.5.2. Estrategia para la búsqueda de cuatro top quarks

Para el canal de un único electrón+jets, el criterio de selección es el siguiente, y se lo denomina *baseline event selection* :

- Exactamente 1 electrón *tight*
- Exactamente cero electrones *loose* adicionales
- ≥ 6 jets con $p_T > 30$ GeV, $|\eta| < 2.5$ y $\Delta R > 0.4$ desde el electrón *loose* más cercano.
- ≥ 2 CSVv2 b-jets con $p_T > 30$ GeV, b-tag working point de 0.8

En la siguiente Sección se abordan más detalles sobre estos criterios de selección.

2.5.3. Filtros de evento y de objeto

El conjunto de datos que son utilizados para el análisis, utilizan el formato ROOT. Es necesario mencionar que la forma de cada conjunto de datos es la de un *jagged array*. Esto significa que existen sub-arrays dentro de un array principal y cada sub-array contiene cierto número de elementos. De esta manera, en el código computacional, un evento corresponde a un sub-array.

Electrones

En primer lugar, para poder obtener un solo electrón definido como *tight* (Sección 2.3.4), se deben aplicar filtros tal que $p_T > 30$ GeV y $|\eta| < 2.1$.

Al elegir $p_T > 30$ GeV se asegura considerar solo electrones de alta energía, los cuales tienen mayor probabilidad de haberse originado a partir del decaimiento de partículas pesadas como el top quark. En este caso, se usa *tight* para asegurar que el electrón reconstruido tenga altas posibilidades de ser un electrón genuino.

Por otro lado, al usar $|\eta| < 2.1$, significa que solo son tomados en cuenta los electrones producidos con ángulos relativamente alejados del haz de protones (i.e. en la región central del detector). Esto debido a que en la región central, las capacidades de reconstrucción del detector son las mas óptimas.

Ahora, para elegir un solo electrón definido como *loose*, son necesarios $p_T > 15$ GeV y $|\eta| < 2.5$. Loose significa que cualquier evento que no contenga electrones es descartado. Luego, la elección de $p_T > 15$ GeV y $|\eta| < 2.5$, ayuda a suprimir electrones de baja energía. De esta manera se reducen procesos de background.

Jets

El requerimiento de al menos 6 jets tiene que ver con la cantidad total de jets que son producidos cuando 4 top quarks decaen (Figura 2.4). Cada jet es originado a partir del decaimiento de un quark.

De igual manera, $p_T > 30$ GeV y $|\eta| < 2.5$, permiten seleccionar jets con energía suficientemente alta y que además son detectados en la región central del detector. Un p_T alto permite seleccionar jets asociados al decaimiento de top quarks.

A su vez, $\Delta R > 0.4$, permite asegurar que los jets seleccionados están lo suficientemente alejados del electrón *loose* más cercano (Sección 2.3.5). Esto permite reducir el número de jets que se originan en eventos de baja energía y evitar asociarlos a la producción de cuatro quarks top.

b-jets

Como se mencionó en la Sección 2.3.6, el algoritmo CSVv2 asigna un valor numérico a cada jet. Mientras mayor es este valor, mayor es la probabilidad de que el jet seleccionado sea un b-jet, es decir, que fue originado a partir de un quark b. El requerimiento de al menos 2 b-jets incrementa la selección de eventos en los que se produjeron top quarks. Así, se incrementa la posibilidad de capturar las partículas producidas debido al decaimiento de un top quark.

Así mismo, $p_T > 30$ GeV asegura que los b-jets seleccionados provengan de eventos de alta energía debido al decaimiento de partículas pesadas como el top quark.

2.6. Análisis Multivariante

Como $t\bar{t}$ es la mayor fuente de background y además tiene un orden de magnitud mucho mayor que $t\bar{t}t\bar{t}$, en este tipo de análisis el background y la señal pueden ser similares. Debido a esto, el uso de métodos de análisis multivariante (Multivariate analysis, MVA) puede ser de utilidad.

En este trabajo, el principal método MVA usado es el llamado *Boosted Decision Trees* (BDT).

2.6.1. Boosted Decision Trees

Un BDT ayuda a maximizar la separación entre muestras de datos, en este caso, señal ($t\bar{t}t\bar{t}$) y background. El BDT toma ciertas variables como datos de entrada (Sección 2.6.2) y realiza una predicción en cada evento clasificándolo como señal o background. Para esto, el workshop mostrado en [28] fue utilizado como referencia.

2.6.2. Variables BDT

A continuación se detallan las variables que son relevantes para este análisis. Cabe mencionar que estas variables se construyeron luego de aplicar el *baseline event selection*.

- N_j se refiere al número de jets por evento. Esta variable es de interés debido a que el estado final de la producción de cuatro top quarks contiene varios jets como se muestra en la Figura 2.4. También es importante tener en cuenta el número de jets que pueden provenir de la principal fuente de background, i.e. pares de top quarks, Figura 2.5.
- H_T^b , se define como la suma del p_T de todos los jets identificados como b-jets por el algoritmo CSVv2. Esta variable puede ayudar a separar la señal del background. Los eventos generados por la señal pueden tener un valor más alto de H_T^b mientras que los eventos de background tienen menor valor.
- H_T , se define como la suma del p_T de todos los jets en cada evento.

- H_T^{Rat} , es el cociente de la suma de los cuatro valores más altos del p_T de jets, para la suma del p_T de los otros jets en el evento.

Capítulo 3

Resultados, conclusiones y recomendaciones

3.1. Selección de eventos

Como ya se había mencionado anteriormente, la mayor fuente de background para $t\bar{t}\bar{t}$ corresponde a la producción de $t\bar{t}$. Sin embargo, para este análisis también se toma en cuenta el background¹ debido a tW (Sección 1.6.2) y W +jets (Sección 1.6.3).

En esta sección se busca estudiar la concordancia entre las distribuciones de todas las variables relevantes, i.e. datos experimentales y simulaciones². Además, ya que la señal es a menudo muy pequeña comparada con las fuentes de background dominantes, es necesario escalarla por un factor de 20.

Todos los eventos mostrados en las siguientes figuras han pasado la **selección** descrita en la Sección 2.5.2. La Figura 3.1 muestra el p_T del leptón seleccionado, en este caso, el electrón. La Figura 3.2 muestra el p_T total de todos los jets en un evento. En estas figuras se observa una concordancia entre simulaciones y datos³.

Por otro lado, cada figura posee un gráfico secundario en la parte inferior. El objetivo de estos subplots o subgráficos, es mostrar el error relativo que

¹El término background se refiere a todo el conjunto de datos que forman parte del mismo, en este caso $t\bar{t}$, tW y W +jets.

²Tanto background como señal ($t\bar{t}\bar{t}$), son simulaciones Monte Carlo. El término -datos experimentales- se refiere a mediciones o datos reales obtenidos por CMS.

³En todas las figuras, EW es una abreviatura de interacciones Electrodébiles, en este caso W +jets. Drell Yan también forma parte de EW, pero su contribución es relativamente pequeña.

existe entre datos reales (Data) y simulaciones Monte Carlo (MC).

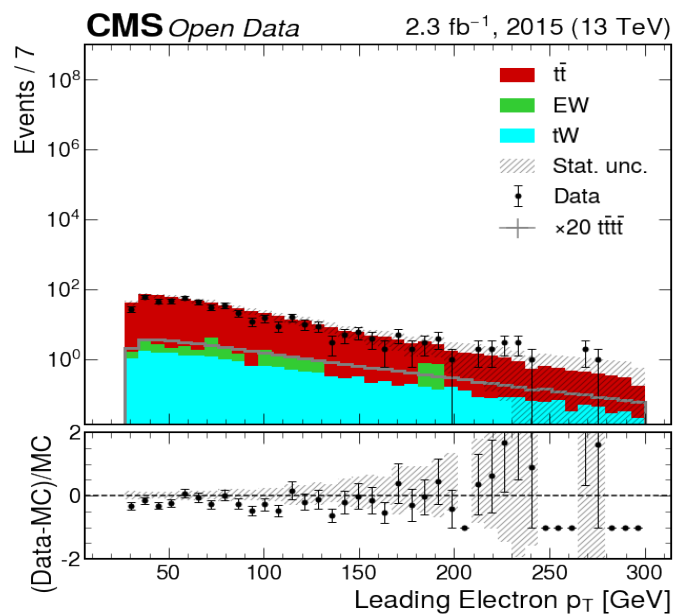


Figura 3.1: Las distribuciones del p_T del electr3n para datos y simulaciones. En este caso se muestra el canal electr3n+jets. Leading Electron hace referencia a un 3nico electr3n *tight* como ya se explic3 en la Secci3n 2.5.2.

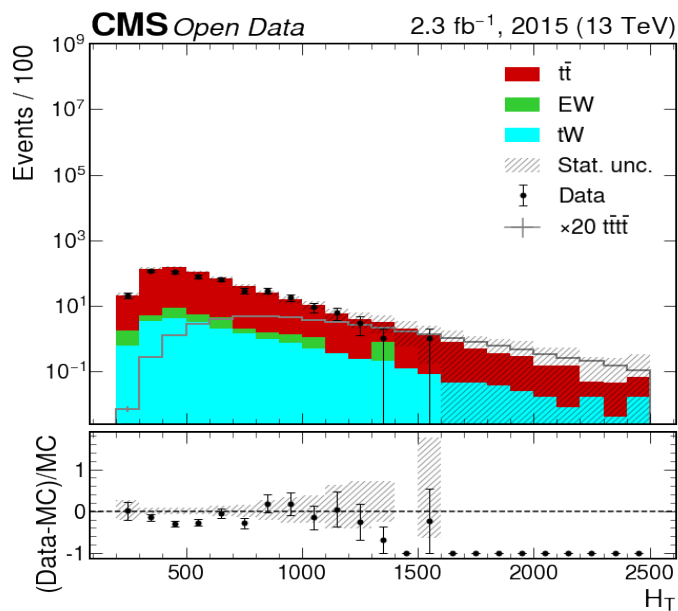


Figura 3.2: Distribuciones del H_T del electr3n para datos y simulaciones. En este caso se muestra el canal electr3n+jets que concuerda con el resultado original.

3.2. Diferenciación entre Señal y Background

Luego de haber aplicado el *baseline event selection*, se observa que $t\bar{t}$ es por mucho el proceso de background dominante. Entonces, la sensibilidad para identificar eventos de $t\bar{t}t\bar{t}$ depende en gran medida de la habilidad para distinguir entre este proceso y $t\bar{t}$.

En esta sección se estudian diferentes variables las cuales muestran potencial para poder distinguir entre señal y background.

3.2.1. Actividad en el evento

Debido a que los eventos $t\bar{t}t\bar{t}$ semileptónicos contendrán hasta 10 jets originados del decaimiento de top quarks, mientras que eventos $t\bar{t}$ contendrán hasta 4 jets, las variables basadas en jets contendrán un alto poder discriminatorio. Sin embargo, también pueden existir jets debido a underlying events/pile up como ya se explicó en la Sección 1.5. Estos jets, por el contrario, pueden originar problemas al momento de diferenciar entre señal y background.

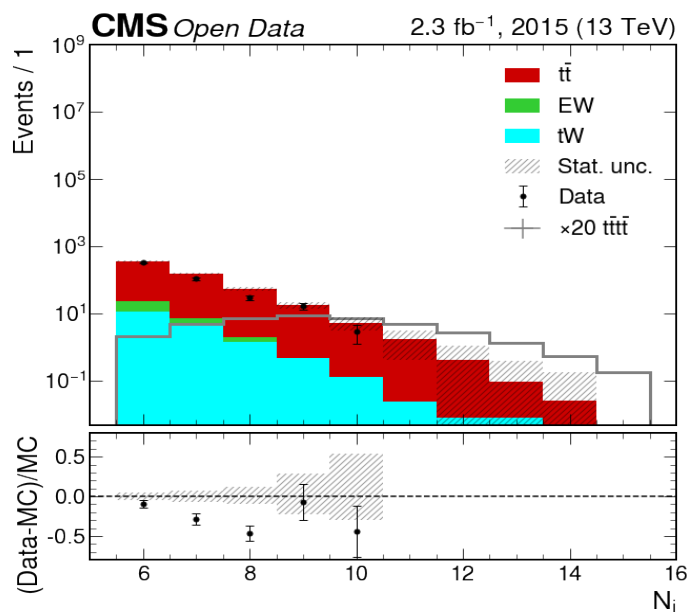


Figura 3.3: Distribuciones del N_j para datos y simulaciones. En este caso se muestra el canal electrón+jets. La cantidad de datos experimentales no es la suficiente para estudiar $N_j > 10$ y por tanto, no se pueden realizar aseveraciones sobre la presencia de eventos de $t\bar{t}t\bar{t}$.

Una de las variables que puede ayudar a discriminar entre $t\bar{t}$ y $t\bar{t}t\bar{t}$ es el número de jets reconstruidos, N_j . En la Figura 3.3, se muestran las distribuciones N_j para datos y simulaciones. Por otra parte, se estima la presencia de hasta 10 jets en aproximadamente 10 eventos de $t\bar{t}$. Además, mientras menos número de jets se tenga, es más probable que hayan sido originados de un evento $t\bar{t}$.

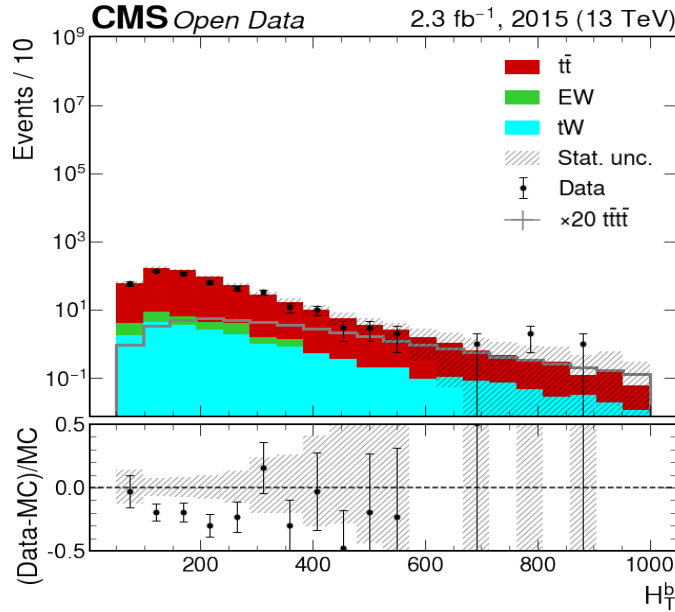


Figura 3.4: Distribuciones del H_T^b para datos y simulaciones. Los errores mostrados en la parte inferior se encuentran dentro del rango esperado comparados con la publicación original.

Otra variable que puede ser usada es H_T^b definida como la suma escalar de los valores p_T de todos los b-jets en el evento. Los eventos pertenecientes a la señal contendrán hasta 4 b-jets originados del decaimiento de top quarks mientras que eventos de background contendrán hasta 2 b-jets y, b-jets adicionales debido a procesos de *gluon splitting*⁴.

Ya que los b-jets originados del decaimiento de top quarks tienen mayores valores de p_T comparados con los que se originan de otros procesos, el H_T de b-jets H_T^b , tenderá a tener mayores valores para la señal que para el background. En la Figura 3.4 se muestran las distribuciones de H_T^b .

Por otro lado, el H_T de los 4 mayores valores del p_T de jets en el evento divi-

⁴El proceso más común de *gluon splitting* tiene como consecuencia el decaimiento del gluón en un par quark-antiquark que luego pueden generar b-jets.

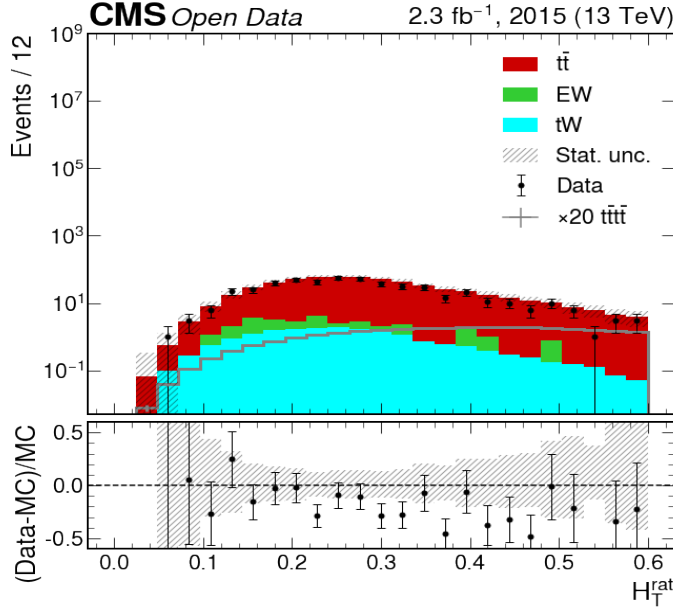


Figura 3.5: Distribuciones del H_T^{Rat} para datos y simulaciones. Las barras de error están en el rango esperado.

dido para el H_T de todos los jets, tenderá a ser alto para eventos de background ya que los cuatro jets principales tienden a ser originados del decaimiento de top quarks. En contraste, para eventos de señal este cociente será más cercano a uno ya que los jets principales en el evento tienden a equilibrarse por otros jets adicionales originados por el decaimiento de otros top quark. Esta variable se denomina H_T^{Rat} y se muestra en la Figura 3.5.

3.2.2. Cantidad de b-jets

Debido a que la probabilidad de que un top quark decaiga en un b quark y un bosón W es $\approx 100\%$, los eventos de señal que contengan cuatro top quarks contendrán cuatro quarks b. El principal proceso de background $t\bar{t}$, típicamente produce dos quarks b mientras que $t\bar{t}t\bar{t}$ generalmente produce cuatro. En consecuencia, la presencia de más de dos jets catalogados como b-jets es una importante fuente con alto poder discriminatorio. Esto se puede visualizar en la Figura 2.4, donde $t\bar{t}t\bar{t}$ contiene 4 b-jets.

En la Figura 3.6 se muestran las distribuciones de b-jets para un working point medio, CSVv2M (Tabla 2.1). Esta variable, denominada N_{tags}^M , ofrece un buen desempeño para discriminar entre señal y background. Sin embargo,

las distribuciones en los histogramas son similares para los eventos de señal y $t\bar{t}$. Además, ya que no se dispone de la cantidad suficiente de datos reales (Data) para un número de b-jets superior a 4, no se puede realizar ninguna aseveración de si se produjeron o no eventos reales de $t\bar{t}\bar{t}\bar{t}$.

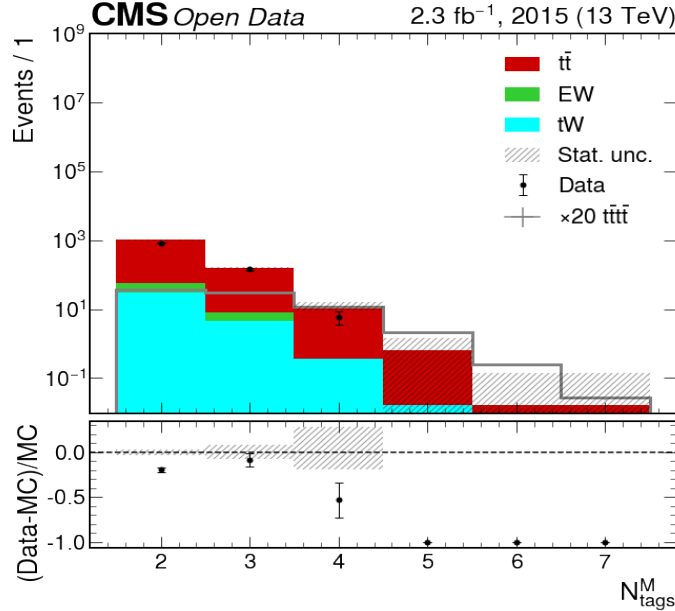


Figura 3.6: Distribuciones del N_M^{tags} para datos y simulaciones. En este caso se muestra el canal electrón+jets. Además, se observa que se han detectado hasta 4 b-jets en aproximadamente 10 eventos de $t\bar{t}$. Cuando existe la presencia de 2 b-jets es más probable que los mismos se hayan originado de por la producción de un par de top quarks $t\bar{t}$ como se explicó en la Sección 2.4.2

Por otra parte, es necesario mencionar que luego de la correcta aplicación de los cortes de selección o filtros, mencionados en 2.5.2, se encontraron inconsistencias en la normalización de las simulaciones con respecto a los datos experimentales, por lo cual, un factor de escala ad hoc fue necesario. Con esta modificación, todas las distribuciones de las variables estudiadas, muestran la concordancia esperada con las simulaciones Monte Carlo.

3.3. BDT

Para obtener una mejor distinción entre señal y background, se usaron dos tipos de algoritmos disponibles en Python, i.e. Adaptive Booster (AdaBoost [29]) y XGBoost classifier el cual incluye un modelo para prevenir el

overfitting. Más detalles se encuentran en [30]. Además, cabe mencionar que estos algoritmos ofrecen una versión simplificada de un clasificador BDT.

Como primer paso, se desea estudiar el desempeño de las variables elegidas. Para este propósito, se tomaron en cuenta las variables descritas en la Sección 2.6.2, ya que presentan un buen desempeño a la hora de discriminar entre señal y background. Esto puede corroborarse en las figuras 3.2-3.5, en particular, el número de jets y b-jets son las variables que ofrecen mejor capacidad de distinción.

Como siguiente paso, se graficaron distintos pares de variables con el propósito de observar su comportamiento. En las figuras 3.7-3.10 realizadas con simulaciones Monte Carlo, se puede observar la presencia de dos regiones (región naranja y región azul).

Cuando un par de variables produce dos regiones claramente separadas, o lo que es lo mismo, no existe una alta dispersión de datos, entonces se dice que poseen buen poder discriminatorio. Esto ratifica que las variables BDT elegidas poseen buen desempeño al momento de distinguir entre señal y background lo cual las hace útiles para el entrenamiento del BDT.

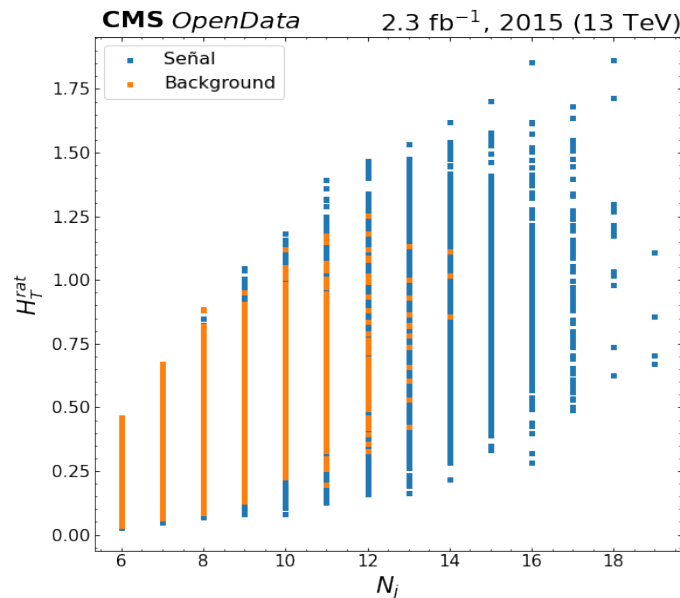


Figura 3.7: Distribuciones H_T^{rat} Vs. N_j para simulaciones. Se aprecian dos regiones definidas. Además, $N_j > 14$ corresponde solo a eventos de señal.

3.3.1. Construcción del BDT

El funcionamiento del algoritmo BDT simplificado es el siguiente. Una vez que se tiene todo el conjunto de datos clasificado entre señal y background, se les asigna una categoría, esto es, 1 y 0 respectivamente. Esto permite entrenar al algoritmo BDT para que pueda diferenciar entre $t\bar{t}$ y background. Estos clasificadores BDT son entrenados de manera similar a cuando se realiza una curva de ajuste (fitting curve).

Una vez que el clasificador ha sido entrenado, se aplican los resultados a los datos experimentales con el objetivo de obtener una variable que indique la probabilidad de que el candidato sea categorizado como 1 (señal) o que se categorizado como 0 (background).

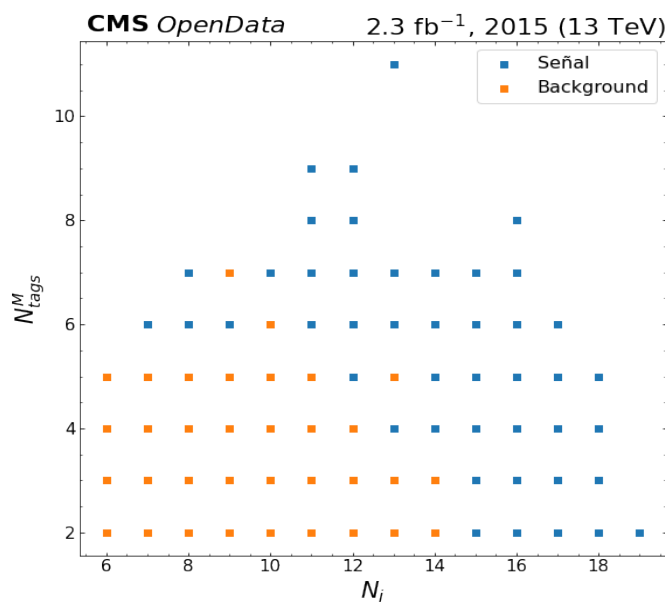


Figura 3.8: Distribuciones N_{tags}^M Vs. N_j para simulaciones. Se puede notar que un número de jets por encima de 14 corresponde a eventos de señal. Además, se corrobora que el numero de jets y b-jets son las variables que pueden ofrecer un mejor rendimiento para distinguir entre señal y background, ya que producen dos regiones claramente separadas.

Es necesario mencionar que el BDT es usado sobre el conjunto de simulaciones luego de haber aplicado el *baseline event selection*. Es decir, el conjunto de datos fue separado en 3 partes, i.e. señal, background y datos experimentales. De esta manera las simulaciones Monte Carlo se usan para entrenamiento

y los datos experimentales se usan para poner a prueba los resultados del entrenamiento.

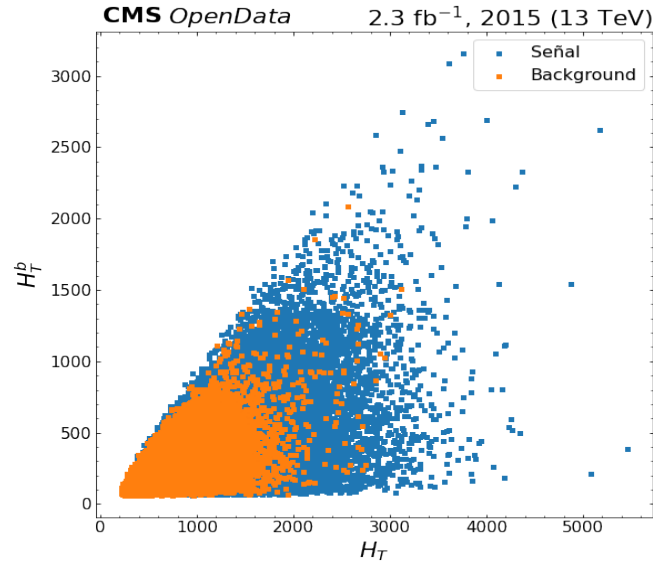


Figura 3.9: Distribuciones H_T^b Vs. H_T para simulaciones. Se pueden observar dos regiones, sin embargo, no se aprecia la misma definición o desempeño que N_{tags}^M Vs. N_j ya que existen eventos de background dispersos en la región donde se encuentran eventos de señal.

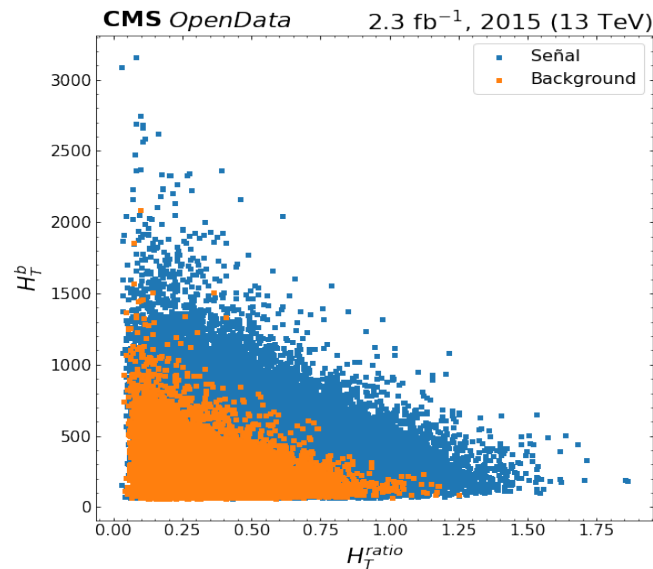


Figura 3.10: Distribuciones H_T^b Vs. H_T^{ratio} para simulaciones. Se observa un buen desempeño discriminatorio entre señal y background. Es decir, producen dos regiones visiblemente separadas.

3.3.2. Resultados de los clasificadores BDT

Los resultados de AdaBoost BDT y XGBoost BDT se encuentran en las figuras 3.11 y 3.12 respectivamente. Se observa que los datos experimentales siguen la forma de los datos clasificados como background. Sin embargo, la limitada cantidad de datos experimentales, no nos permite asegurar la presencia de algún evento real de $t\bar{t}t\bar{t}$.

En el eje x, cuando las distribuciones se encuentran más cerca de 1, existe mayor probabilidad de que se trate de un evento de señal, y, cuando se encuentran más cercanas a cero mayor probabilidad de que se trate de un evento de background. En la Figura 3.11 se observa que buena parte de las simulaciones de señal fueron clasificadas con 0.7 de probabilidad de ser eventos $t\bar{t}t\bar{t}$ claramente distinguibles. Sin embargo, los datos experimentales fueron categorizados casi en su totalidad como background. Es decir, con el número de datos experimentales recolectados, no se advierte, en este estudio, la posibilidad de haber observado eventos con cuatro tops.

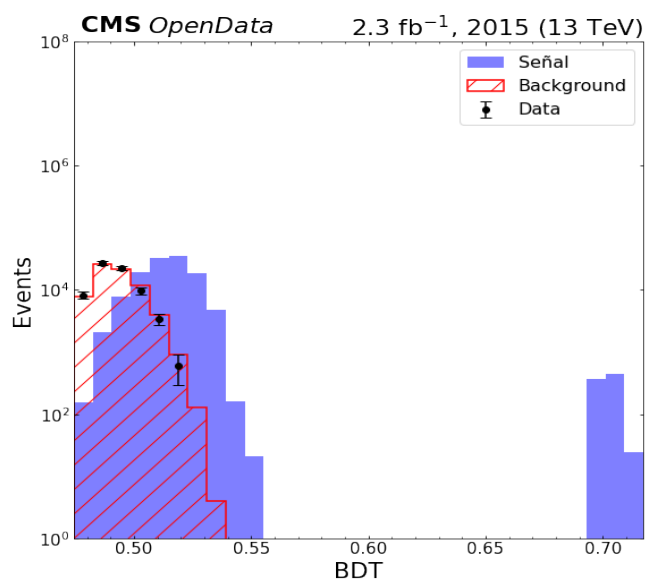


Figura 3.11: Distribuciones de las variables BDT para el canal electrón+jets obtenidas mediante AdaBoost

En la Figura 3.12, las distribuciones de background y señal ocupan todo el rango entre cero y uno. En consecuencia, no se puede aseverar la existencia de ningún evento de señal $t\bar{t}t\bar{t}$ en los datos experimentales. Por el contrario,

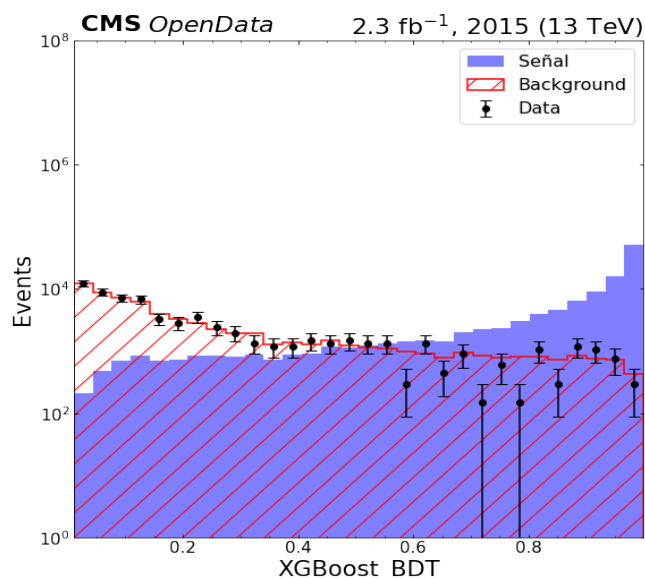


Figura 3.12: Distribuciones de las variables BDT para el canal electrón+jets obtenidas mediante XGBoost

al igual que con AdaBoost, los datos experimentales siguen la tendencia del background.

3.3.3. Rendimiento del clasificador BDT

Una herramienta para poder evaluar el rendimiento de un BDT se denomina curva *Receiver Operating Characteristic*, también conocida como curva ROC. Esto muestra la eficiencia del clasificador BDT sobre la señal (tasa de verdaderos positivos) dividida para la ineficiencia para remover la señal (tasa de falsos positivos).

En la Figura 3.13 se observa el resultado de las curvas ROC para ambos algoritmos BDT, AdaBoost y XGBoost. Hay que tener en cuenta que, mientras más cerca esté la curva de la esquina superior izquierda, significa que el clasificador correspondiente a esa curva es mejor. Otra manera de ver este resultado es calculando el área bajo la curva. Mientras mayor es el área, mejor es el clasificador BDT.

Tanto AdaBoost como XGBoost presentan un buen rendimiento al momento de distinguir entre señal y background. También se puede ver que XGBoost no presenta una mejora sustancial. Con estos resultados, figuras 3.11 y 3.12,

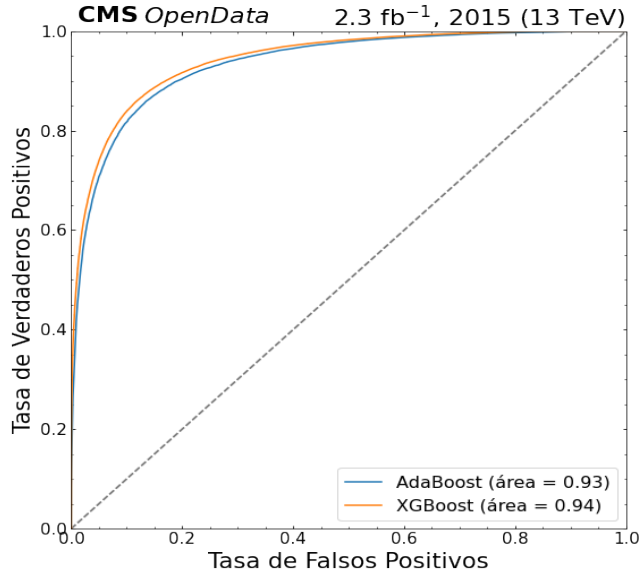


Figura 3.13: Curvas ROC para AdaBoost y XGBoost

se evidencia que no existe presencia de algún evento $t\bar{t}t\bar{t}$ en el conjunto de datos experimentales obtenidos por CMS en el año 2015, lo cual concuerda con lo mencionado en la publicación original [7].

3.4. Conclusiones y recomendaciones

Se realizó una búsqueda de eventos que contengan cuatro top quarks mediante el uso de datos abiertos (Open Data) del experimento CMS correspondientes al año 2015 con una luminosidad integrada de 2.3 fb^{-1} a $\sqrt{s} = 13 \text{ TeV}$. El análisis se enfocó en el estado final de un electrón+jets correspondiente al canal de un único leptón.

El trabajo fue desarrollado en 2 partes. La primera de ellas consistió en aplicar el *baseline event selection* (Sección 2.5.2) a cada evento con el objetivo de tener una primera aproximación de la supresión de fuentes de background. Con esto, se observó concordancia entre datos experimentales y simulaciones Monte Carlo en las variables de interés estudiadas.

La segunda etapa consistió en la aplicación de un algoritmo BDT simplificado para incrementar la sensibilidad en la distinción entre señal ($t\bar{t}t\bar{t}$) y background. Sin embargo, en este conjunto de datos reales obtenidos por CMS, no

se observó ningún evento congruente con la señal de interés.

Para asegurar que el entrenamiento del BDT no provocó un overfitting o sobreestimación, curvas ROC fueron analizadas las cuales mostraron que los algoritmos usados, AdaBoost y XGBoost, fueron eficientes. Sin embargo, es recomendable usar otros métodos adicionales que permitan corroborar que los resultados no son afectados por sobreentrenamiento del algoritmo.

Por otra parte, se hallaron inconsistencias con la normalización de las simulaciones en relación con los datos experimentales usados. Esto fue detectado solo en la parte final de este estudio y, por lo tanto, la determinación de un límite en la sección eficaz de $t\bar{t}\bar{t}$ no es recomendable. Sin embargo, el estudio de la forma de las distintas distribuciones y el código computacional desarrollado, contribuyen con el objetivo de brindar un primer enfoque al análisis de datos en Física de Partículas.

Finalmente, con el objetivo de mantener un análisis simplificado, en este estudio no se aplicaron correcciones a las simulaciones Monte Carlo para evitar efectos provocados por el Pile-Up. En efecto, es recomendable realizar tales correcciones junto con un tratamiento completo de incertidumbres sistemáticas. A pesar de estas simplificaciones, se logró reproducir, de manera general, la forma de las distribuciones de las variables más importantes del análisis original, incluyendo un estudio con BDTs. Se espera que este trabajo ayude a reducir la brecha técnica para un analista de datos (abierto) interesado en Física de Partículas.

Anexo A

Conjunto de simulaciones usadas en el análisis

Set de datos	Eventos	Orden generado	$\sigma(pb^{-1})$
/WJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/ RunIIFall15MiniAODv2-PU25nsData2015v1_76X_ mcRun2_asymptotic_v12_ext4-v1/MINIAODSIM[31]	19M	NLO	61526
/TT_TuneCUETP8M1_13TeV-powheg-pythia8/ RunIIFall15MiniAODv2-PU25nsData2015v1_76X_ mcRun2_asymptotic_v12_ext4-v1 [32]	19M	NLO	831
/TT_TuneEE5C_13TeV-powheg-herwigpp/ RunIIFall15MiniAODv2-PU25nsData2015v1_76X_ mcRun2_asymptotic_v12-v1 [33]	19M	NLO	831

Tabla A.1: Nombre del conjunto de datos y número total de eventos para las muestras sistemáticas simuladas. Todas las muestras utilizan el formato de datos MINIAOD y pueden ser encontradas en [CERN Open Data Portal](#).

Set de datos	Eventos	Orden	σ (pb^{-1})
/TTTT_TuneCUETP8M1_13TeV-amcatnlo-pythia8/ RunIIFall15MiniAODv2-PU25nsData2015v1_76X_mcRun2_asymptotic_v12_ext1-v1 [34]	960K $N_{eff} = 393K$	NLO	0,009
/TT_TuneCUETP8M1_13TeV-powheg-pythia8/ RunIIFall15MiniAODv2-PU25nsData2015v1_76X_mcRun2_asymptotic_v12_ext3-v1 [35]	97M	NLO	831
/WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/ RunIIFall15MiniAODv2-PU25nsData2015v1_76X_mcRun2_asymptotic_v12_ext2-v1 [36]	47M	LO	61526
/ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8TuneCUETP8M1/ RunIIFall15MiniAODv2-PU25nsData2015v1_76X_mcRun2_asymptotic_v12-v1 [37]	1M	NLO	35,6
/ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1/ RunIIFall15MiniAODv2-PU25nsData2015v1_76X_mcRun2_asymptotic_v12-v1 [38]	1M	NLO	35,6

Tabla A.2: Nombre del conjunto de datos y número total de eventos para las muestras simuladas. Todas las muestras utilizan el formato de datos MINI-AOD.

Referencias bibliográficas

- [1] ATLAS Collaboration. Evidence for $t\bar{t}t\bar{t}$ production in the multilepton final state in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *European Physical Journal C*, 80, 11 2020.
- [2] The CMS and ATLAS Collaborations. ATLAS and CMS observe simultaneous production of four top quarks. <https://home.cern/news/news/physics/atlas-and-cms-observe-simultaneous-production-four-top-quarks>. Accedido en Abril 01, 2023.
- [3] CMS Collaboration. Search for the production of four top quarks in the single-lepton and opposite-sign dilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Journal of High Energy Physics*, 2019, 11 2019.
- [4] Glennys R. Farrar and Pierre Fayet. Phenomenology of the production, decay, and detection of new hadronic states associated with supersymmetry. *Physics Letters B*, 76(5):575–579, 1978.
- [5] H.P. Nilles. Supersymmetry, supergravity and particle physics. *Physics Reports*, 110(1):1–162, 1984.
- [6] The CMS Collaboration. 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. *European Physical Journal C*, 80, 2 2020.
- [7] The CMS Collaboration. Search for standard model production of four top quarks in proton–proton collisions at $\sqrt{s} = 13$ TeV. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics*, 9 2017.<https://doi.org/10.48550/arXiv.1702.06164>.

- [8] CMS Collaboration. Observation of four top quark production in proton-proton collisions at $\sqrt{s} = 13$ TeV. 5 2023.
- [9] David Griffiths. *Introduction to Elementary Particles*. Second edition, 2008.
- [10] C. Patrignani. Review of particle physics. *Chinese Physics C*, 40(10):100001, oct 2016.
- [11] The CMS Collaboration. Measurement of the t-channel single-top-quark production cross section and of the $|V_{tb}|$ ckm matrix element in pp collisions at $\sqrt{s} = 8$ TeV. *Journal of High Energy Physics*, 2014, 2014.
- [12] Peter Barnreuther. *Top Quark Pair Production at the LHC*. 2012.
- [13] Andrea Giammanco. *Single top quark production at the LHC*. 11 2015.
- [14] Nicolas Stylianou. *A search for four top quarks production at the CMS experiment via the single lepton channel*. 2023.
- [15] Fabienne Marcastel. CERN's Accelerator Complex. La chaîne des accélérateurs du CERN, 2013.
- [16] The CMS Collaboration. The CMS experiment at the CERN LHC, 2008.
- [17] Tai Sakuma and Thomas McCauley. Detector and event visualization with sketchup at the CMS experiment. volume 513. Institute of Physics Publishing, 2014.
- [18] Izaak Neutelings. Grand challenges. https://tikz.net/axis3d_cms/, 2021.
- [19] The CMS Collaboration. Particle-flow reconstruction and global event description with the CMS detector. *Journal of Instrumentation*, 12, 10 2017.
- [20] CMS Collaboration. Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. *Journal of Instrumentation*, 16, 5 2021.
- [21] Gavin P. Salam. Towards jetography. *European Physical Journal C*, 67:637–686, 6 2010.

- [22] The CMS Collaboration. Identification of b quark jets at the CMS Experiment in the LHC Run 2, 2016.
- [23] The CMS Collaboration. Physics Objects Extractor Tool (POET) - 2015MiniAOD. <https://github.com/cms-opendata-analyses/PhysObjectExtractorTool/tree/2015MiniAOD>.
- [24] Elliott Kauffman. Produce-nanoAODs. <https://github.com/ekauffma/produce-nanoAODs.git>. Accedido en Enero 30, 2023.
- [25] Institute for Research and Innovation in Software for High Energy Physics (IRIS-HEP). Grand challenges. <https://iris-hep.org/grand-challenges.html>. Accedido en Enero 30, 2023.
- [26] CMS collaboration (2021). SingleElectron primary dataset in MINIAOD format from RunD of 2015 (/SingleElectron/Run2015D-08Jun2016-v1/MINIAOD). CERN Open Data Portal. DOI:10.7483/OPENDATA.CMS.29BN.FBTV.
- [27] Andrés Chicaiza. Four Tops CMS Open Data. <https://github.com/andresc27/FourTopsAnalysisCMSOpenData.git>. Accedido en Mayo 30, 2023.
- [28] Analysis essentials. Multivariate Analysis. <https://hsf-training.github.io/analysis-essentials/advanced-python/30Classification.html#>.
- [29] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55:119 –139, 1997.
- [30] Carlos Guestrin and Tianqi Chen. Xgboost: A scalable tree boosting system. 3 2016.
- [31] CMS Collaboration (2021). Simulated dataset WJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8 in MINIAODSIM format for 2015 collision data. CERN Open Data Portal. DOI:10.7483/OPENDATA.CMS.FLGU.07DD.

- [32] CMS Collaboration (2021). Simulated dataset TT_TuneCUETP8M1_13TeV-powheg-pythia8 in MINIAODSIM format for 2015 collision data. CERN Open Data Portal. DOI: [10.7483/OPENDATA.CMS.4BI9.J6N1](https://doi.org/10.7483/OPENDATA.CMS.4BI9.J6N1).
- [33] CMS Collaboration (2021). Simulated dataset TT_TuneEE5C_13TeV-powheg-herwigpp in MINIAODSIM format for 2015 collision data. CERN Open Data Portal. DOI: [10.7483/OPENDATA.CMS.KGUZ.PIUI](https://doi.org/10.7483/OPENDATA.CMS.KGUZ.PIUI).
- [34] CMS Collaboration (2021). Simulated dataset TTTT_TuneCUETP8M1_13TeV-amcatnlo-pythia8 in MINIAODSIM format for 2015 collision data. CERN Open Data Portal. DOI: [10.7483/OPENDATA.CMS.8TMX.NE3H](https://doi.org/10.7483/OPENDATA.CMS.8TMX.NE3H).
- [35] CMS Collaboration (2021). Simulated dataset TT_TuneCUETP8M1_13TeV-powheg-pythia8 in MINIAODSIM format for 2015 collision data. CERN Open Data Portal. DOI: [10.7483/OPENDATA.CMS.JJEM.1DKC](https://doi.org/10.7483/OPENDATA.CMS.JJEM.1DKC).
- [36] CMS Collaboration (2021). Simulated dataset WJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8 in MINIAODSIM format for 2015 collision data. CERN Open Data Portal. DOI: [10.7483/OPENDATA.CMS.YRJY.OK2Z](https://doi.org/10.7483/OPENDATA.CMS.YRJY.OK2Z).
- [37] CMS Collaboration (2021). Simulated dataset ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1 in MINIAODSIM format for 2015 collision data. CERN Open Data Portal. DOI: [10.7483/OPENDATA.CMS.EKZ2.EXSU](https://doi.org/10.7483/OPENDATA.CMS.EKZ2.EXSU).
- [38] CMS Collaboration (2021). Simulated dataset ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1 in MINIAODSIM format for 2015 collision data. CERN Open Data Portal. DOI: [10.7483/OPENDATA.CMS.JG3K.U1HH](https://doi.org/10.7483/OPENDATA.CMS.JG3K.U1HH).