



Universidad de Oviedo

UNIVERSIDAD DE OVIEDO

FACULTAD DE CIENCIAS

TRABAJO DE FIN DE GRADO

ESTUDIO DE PROCESOS CON LEPTONES Y USO
Y DESARROLLO DE TÉCNICAS DE INTELIGENCIA
ARTIFICIAL PARA SU IDENTIFICACIÓN EN EL
EXPERIMENTO CMS DEL LHC DEL CERN

Autor: ADRIÁN FERNÁNDEZ ROLDÁN

Tutor: FRANCISCO JAVIER CUEVAS MAESTRO

Cotutor: ALEJANDRO SOTO RODRÍGUEZ

13 de julio de 2023

Índice general

Introducción	III
1. Marco teórico	1
1.1. Modelo Estándar	1
1.1.1. Más allá del Modelo Estándar	5
2. Dispositivo experimental	7
2.1. El LHC	7
2.1.1. Luminosidad y Pile-up	9
2.2. Componentes del LHC	11
2.3. Detector CMS	13
2.3.1. Sistema de coordenadas de CMS.	14
2.3.2. <i>Tracker</i>	15
2.3.3. Calorímetros	17
2.3.4. Solenoide superconductor	20
2.3.5. Cámaras de muones	21
2.3.6. <i>Trigger</i>	22
3. Gestión de la información	25
3.1. Simulación de sucesos	26
3.2. Reconstrucción de sucesos	27
3.2.1. Vértice principal	28
3.2.2. Muones	29
3.2.3. Electrones	30

3.2.4. Hadrones y jets	30
3.2.5. Momento transverso faltante	30
4. Planteamiento del estudio	33
4.1. Objetivos del trabajo	33
4.2. Producción de quarks top	34
4.2.1. Proceso $t\bar{t}$ semileptónico	37
4.2.2. Proceso tW	38
4.3. Cut-Based	39
4.4. Lepton MVA	40
4.4.1. Descripción del Lepton MVA	41
4.5. Criterio de selección de eventos	45
5. Cut-based vs Lepton MV	49
5.1. Discriminación de leptones <i>non-prompt</i>	50
5.2. Efectos en la señal	52
5.3. Eficiencia en señal y fondo	54
5.4. Optimización de los resultados	57
5.5. Consistencia de los resultados	60
6. Conclusiones	61
Bibliografía	63

Introducción

En este trabajo se estudia la posible implementación de un algoritmo de *machine learning* para la reducción de una de las mayores fuentes de incertidumbre que se producen en la medida de la sección eficaz del proceso de producción asociada de un quark top y un bosón W realizado en el detector CMS del LHC del CERN en colisiones protón-protón a energías del centro de masa de 13 TeV descrito en [1]. La implementación de algoritmos como el usado en este trabajo abre un nuevo camino hacia el perfeccionamiento de las técnicas hasta ahora empleadas. El uso de la inteligencia artificial ofrece mejoras en la precisión de las medidas y con ello nuevos avances en investigación. Se espera que este estudio proporcione una visión clara de cómo estas técnicas pueden ayudar a mejorar las medidas actuales y futuras.

Se ha hecho uso de un algoritmo de aprendizaje automático como es el *Lepton MVA* para comprobar su desempeño y validez a la hora de reducir el número de leptones mal identificados del proceso $t\bar{t}$ semileptónico, que constituyen la segunda fuente de mayor incertidumbre en [1]. Para ello se ha empleado muestras de simulaciones MC del año 2018, que han sido tratadas siguiendo los pasos realizados en [1].

Este estudio, organizado en seis capítulos, presenta la siguiente estructura. En el capítulo 1, se realiza una introducción al marco teórico vigente actualmente en física de altas energías, el Modelo Estándar de Física de Partículas. En el capítulo 2, es descrito el dispositivo experimental que constituye una parte fundamental en este estudio, el gran colisionador de hadrones (LHC), así como su detector CMS. En este capítulo se describen ambos dispositivos y de esta manera se introducen los mecanismos que van a ser fundamentales en el estudio posterior. En el capítulo 3, se

trata la simulación y reconstrucción de los procesos que ocurren en estos dispositivos, así como la descripción de técnicas utilizadas para extraer la información física deseada de las colisiones. El capítulo 4 está dedicado a la explicación del estudio y la descripción de las herramientas utilizadas. En el capítulo 5, se muestran los resultados obtenidos así como la discusión de los mismos. Por último, el capítulo 6 incluye las conclusiones, con los principales resultados y un breve resumen del trabajo realizado.

Capítulo 1

Marco teórico. El Modelo Estándar.

1.1. Modelo Estándar

El Modelo Estándar es el marco teórico que describe las partículas fundamentales que componen la materia y sus interacciones. Agrupa las partículas elementales en dos grandes bloques: los *fermiones*, partículas que poseen espín semientero y los *bosones*, con espín entero. Dentro de la clasificación de los fermiones encontramos a su vez dos familias, los *leptones* y los *quarks*. Éstas son las partículas elementales que constituyen la materia y las distintas interacciones entre ellas se producen mediante el intercambio de bosones que actúan como mediadores de dichas interacciones. Las partículas llevan asociada cargas de distintos tipos que les confiere la propiedad de llevar a cabo ciertos tipo de interacción. Existen tres tipos de cargas: eléctrica, que está asociada a la interacción electromagnética; débil, asociada a la interacción débil; y carga fuerte o de color (rojo, verde y azul), que está asociada a la interacción fuerte. La intensidad de cada interacción viene dada por el valor de su constante de acoplamiento, α para la interacción electromagnética, α_W para la interacción débil y α_S para la interacción fuerte.

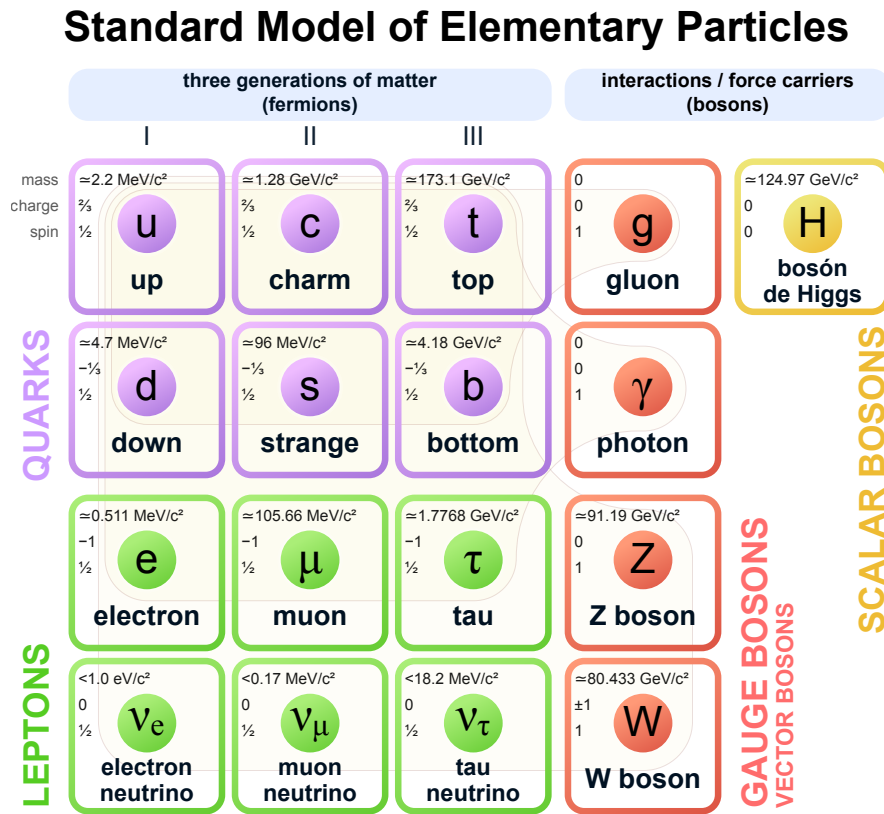


Figura 1.1: Clasificación de las partículas elementales según el Modelo Estándar.

Retomando la clasificación de los fermiones, tanto el grupo de leptones como el de quarks están formados por seis partículas agrupadas en parejas dando lugar a tres generaciones de partículas diferentes ordenadas de menor a mayor masa. La materia estable del universo está formada por partículas de la primera generación, quarks up, quarks down y electrones. Al ser de primera generación, son las partículas elementales más ligeras y, por tanto, las más estables. Las siguientes dos generaciones de leptones y quarks, corresponden a partículas más masivas que las anteriores y que decaen a partículas más estables.

Los quarks se clasifican en *up* (**u**), *down* (**d**) (primera generación), *charm* (**c**), *strange* (**s**) (segunda generación) y *top* (**t**), *bottom* (**b**) (tercera generación). Tienen carga eléctrica, carga débil y carga de color.

Los distintos quarks se unen formando *hadrones* de manera que la partícula re-

sultante no tiene carga de color. Los hadrones están formados por quarks del mar, quarks de valencia y gluones. A su vez, estos pueden ser de dos tipos: **bariones** o **mesones**. Los bariones son hadrones que están formados por tres quarks de valencia y los mesones son aquellos que están formados únicamente por dos quarks de valencia. En cuanto a la carga eléctrica, u, c y t tienen una carga de valor $+2/3$ (en unidades naturales de carga $c = 1$ y $\hbar = 1$) y d, b y s de $-1/3$. La partícula resultante de la unión de ciertos quarks de valencia tendrá de carga eléctrica la suma de la carga eléctrica de ellos. Otro dato importante es que los quarks nunca se observan en libertad, siempre se encuentran confinados en hadrones, ya que la fuerza que los une aumenta a medida que se separan. Todos los quarks tienen sus correspondiente antiquark \bar{u} , \bar{d} , \bar{t} , \bar{b} , \bar{c} y \bar{s} , igual en masa y opuesto en carga eléctrica y de color.

Los leptones se subdividen también en tres generaciones, *electrón* (e), *neutrino del electrón* (ν_e) (primera generación), *muón* (μ), *neutrino del muón* (ν_μ) (segunda generación) y *tau* (τ), *neutrino del tau* (ν_τ) (tercera generación). Los leptones poseen carga eléctrica y carga débil pero no carga de color. Cabe notar que las generaciones de los leptones se diferencian unas de otras principalmente por su masa, ya que poseen propiedades similares. Además, cada leptón posee su antipartícula correspondiente. Estas partículas presentan propiedades similares a su espejo, poseen misma masa y espín pero su carga eléctrica presenta signo contrario.

Por otro lado, existen partículas con espín entero, los bosones. Son partículas mediadoras de las distintas interacciones que existen entre la materia, es decir, son partículas de intercambio que aparecen cuando dos o más partículas interactúan. Existen cinco tipos de bosones, el **fotón** (γ), mediador de la interacción electromagnética, el **gluón** (g), mediador de la interacción fuerte, los bosones W^\pm y Z^0 , mediadores de la interacción débil y el **bosón de Higgs** (H), que mediante el mecanismo de Higgs dota de masa a las partículas.

La interacción electromagnética actúa entre partículas con carga eléctrica. Incluye la fuerza electrostática, que actúa entre cargas en reposo, y el efecto combinado de las fuerzas eléctrica y magnética que actúan entre cargas que se mueven una respecto a la otra. Tiene un campo de acción teóricamente infinito, pero es menos

significativo cuanto mayor es la distancia que las separa. Esta interacción es la que mantiene unidas las moléculas y átomos, ya que, a estas distancias de estructura atómica y molecular, la atracción y repulsión eléctrica es tan dominante sobre las otras fuerzas que estas pueden ser despreciadas.

La interacción débil, también conocida como interacción nuclear débil, se acopla a un tipo de carga llamada sabor, que la poseen los quarks y los leptones. Esta interacción es la causante de los cambios de sabor en estas partículas, en otras palabras, es la responsable de que quarks y leptones decaigan en partículas más ligeras. Su intensidad es menor que la intensidad de la interacción electromagnética y su alcance es menor que el de la interacción fuerte.

La interacción nuclear fuerte permite unirse a los quarks para formar hadrones. Estos son portadores de un tipo de carga, llamada carga de color, de distinta naturaleza a la carga eléctrica. El radio de acción de la interacción fuerte es mucho más pequeño que el de la electromagnética; sólo se le aprecia a muy cortas distancias tales como el radio del núcleo; su fuerza, sin embargo, es más intensa que la de cualquiera de las otras tres interacciones. En una colisión de altas energías cuando se producen partículas con carga de color, rápidamente se hadronizan. Este hadrón puede ser inestable, entonces se desintegra en otras partículas que a su vez pueden volver a agruparse formando hadrones. Esto produce un chorro de partículas que se denomina *jet*. La interacción fuerte entre los quarks de los protones y neutrones que constituyen los átomos explica la fuerza de atracción entre estas dos partículas dentro del núcleo atómico. Debido a la carga positiva de los protones, para que éstos se encuentren estables en el núcleo debía existir una fuerza mayor que la electromagnética para retenerlos. Esta se debe a la interacción fuerte.

Un concepto relacionado con las partículas mediadoras, es el de *partícula virtual*. En los procesos de interacción entre partículas, las partículas mediadoras viven por lo general un tiempo tan pequeño que, debido al principio de incertidumbre de Heisenberg, no es posible determinar sus características con total certeza, o lo que es lo mismo, no se pueden medir.

1.1.1. Más allá del Modelo Estándar

Es sabido que el Modelo Estándar es la teoría física más robusta hasta la fecha, pero hay ciertos conceptos que aún no ha sido capaz de explicar. Gran parte de experimentos realizados en física de partículas tienen como objetivo comprobar si los valores que predice el Modelo Estándar son ciertos, pero otros están enfocados en la búsqueda de fenómenos no descritos por el Modelo Estándar. Por otro lado, se avanza también en la parte teórica con el fin de englobar todo dentro de una misma teoría. Se está intentando incorporar por ejemplo, la teoría de cuerdas, la teoría de supersimetría y la gravitación cuántica como extensiones del Modelo Estándar. Algunos fenómenos que el Modelo Estándar aún no ha sido capaz de explicar son:

- La interacción gravitatoria: La descripción que se da de esta fuerza en *La Teoría de la Relatividad General*, no encaja con el formalismo empleado en el Modelo Estándar, en este las interacciones son llevadas a cabo mediante partículas mediadoras mientras que en la Teoría de la Relatividad se trata de una deformación del *espacio-tiempo*. Se ha propuesto la existencia de una partícula mediadora de esta interacción, el *gravitón*, pero han no ha sido probada su existencia.
- La materia oscura: No puede ser detectada por la observación de la radiación electromagnética. Se sabe de su existencia debido a los efectos que produce sobre objetos que sí se pueden observar directamente [2].
- La energía oscura: La existencia de energía oscura se infiere a partir de las observaciones cosmológicas, que indican que el universo se está expandiendo a un ritmo acelerado. Esta expansión acelerada sugiere que hay una forma de energía en el universo que está produciendo una fuerza repulsiva que está impulsando a las galaxias a alejarse unas de otras con cada vez más fuerza. Las teorías sobre la naturaleza de la energía oscura incluyen la posibilidad de que sea una propiedad intrínseca del espacio vacío, conocida como constante cosmológica, o que sea causada por una nueva forma de campo de energía que aún no se ha descubierto.

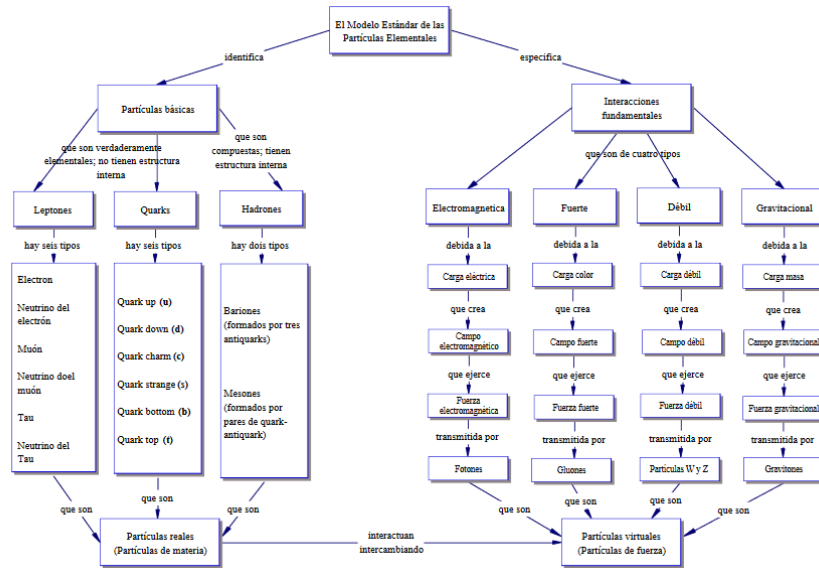


Figura 1.2: Cuadro resumen del Modelo Estándar. Partículas e interacciones. [2]

- **Asimetría materia-antimateria:** La asimetría materia-antimateria se refiere a la diferencia en la cantidad de materia y antimateria que existe en el universo observable. La materia y la antimateria son dos tipos de partículas subatómicas que tienen propiedades opuestas, como la carga eléctrica y el número cuántico de espín. No existe ninguna ley fundamental que se comporte diferente para partículas o antipartículas. Por lo tanto, en el momento del *Big Bang* se tendrían que haber creado en la misma cantidad de ambas. Se ha comprobado que el universo observable está compuesto casi en su totalidad por materia, lo que sugiere que hubo una asimetría materia-antimateria en el universo temprano.

Capítulo 2

Dispositivo experimental

2.1. El LHC

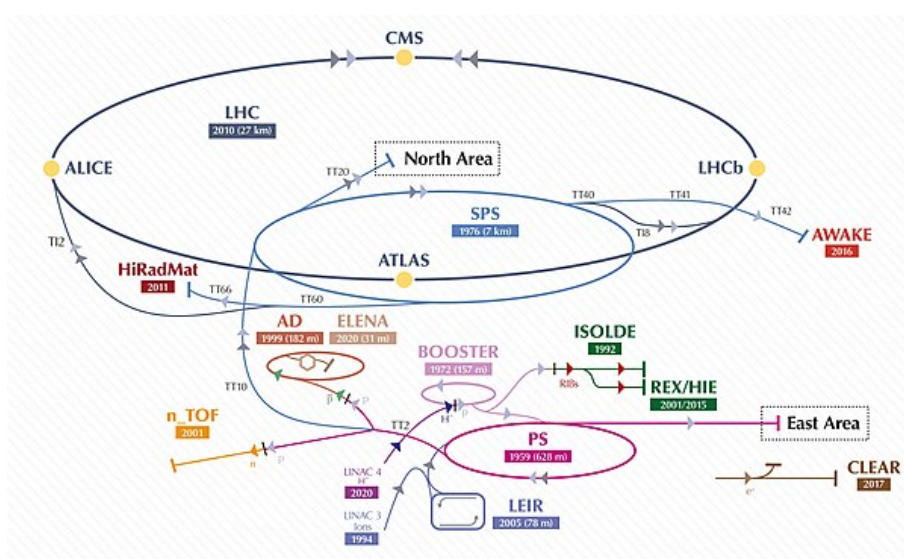


Figura 2.1: Estructura detallada de los aceleradores y experimentos del LHC.[3]

El LHC (*Large Hadron Collider*) es el mayor acelerador de partículas del mundo. Se trata de un acelerador de partículas circular con una longitud de unos 27 kilómetros. Se encuentra situado en la frontera entre Francia y Suiza a una distancia de la superficie que oscila entre los 50 metros y los 175 metros. El LHC tiene una estructura circular que consiste en dos anillos paralelos por los que circulan las partículas

y consta de imanes superconductores refrigerados mediante Helio-4 superfluido para poder alcanzar una temperatura de 1,9 K con el fin de provocar la mínima excitación molecular posible, mejorando así la conducción de las partículas y disminuyendo al máximo posibles interferencias.

Las partículas que se aceleran en el LHC son tanto protones como iones pesados, cuya velocidad llega a ser cercana a la de la luz. Estas partículas se agrupan en haces (*beams*). En el interior del acelerador, dos haces de partículas viajan en sentidos opuestos en dos anillos paralelos (el interior de estos tubos se encuentra al vacío). En este trabajo se estudiará la colisión protón-protón.

Los protones se obtienen en una primera fase después de la cual son acelerados por LINAC4, PSB, el Sincrotrón de Protones (PS) y el Super Sincrotrón de Protones (SPS). En esta primera fase a los átomos de H se les extraen los electrones de manera que se obtiene H^+ . Se hacen pasar estos iones por un cátodo de Cesio (un átomo con baja energía de ionización) por lo que los H^+ adquieren dos electrones convirtiéndose en ion hidruro H^- . Dada su carga negativa, el ion hidruro sale repelido de la superficie del cátodo para ser extraído finalmente de la fuente de iones en pulsos de 400 μs , formándose un haz de estos iones. Los iones H^- salen de esta fuente con una energía de 45 keV, y a través de la línea del Linac 4 de 86 m adquieren 160 MeV de energía final.

Los iones pasan a través de los electroimanes que cambian alternativamente de positivo a negativo. Primero atraen a los iones negativos (H^-) y a continuación los repelen impulsándolos hacia adelante. Los iones H^- llegan al PSB y se unen a los protones que ya circulan en este primer acelerador circular, formando paquetes compactos. Después, los dos electrones que conforman el ión hidruro son extraídos mediante una delgada hoja de carbón para que todo el haz que gira en el PSB esté formado solo por protones. Una nueva remesa de H^- llegan del Linac4 para repetir el proceso. A los protones circulantes en el PSB se les unirán más iones de hidrógeno cargados negativamente, que sufrirán el mismo proceso. Al entrelazar repetidamente iones negativos y positivos, se consigue crear paquetes o *bunches* de protones muy compactos [4]. Estos *bunches* poseen $1.15 \cdot 10^{11}$ protones y tardan alrededor de 25

ns. Tras su paso por PSB, llegan a PS, luego a SPS y finalmente alcanzan el LHC.

2.1.1. Luminosidad y Pile-up

Luminosidad

La luminosidad instantánea es una medida de la tasa de colisiones por unidad de tiempo y unidad de área. Es un parámetro importante para caracterizar la cantidad de datos que se pueden obtener tras la colisión de los haces de protones. Durante *Run 2* el LHC operó con una luminosidad instantánea promedio de $1,5 \cdot 10^{34} \text{ cm}^{-2}\text{s}^{-1}$. El cálculo de la luminosidad instantánea puede realizarse con la siguiente ecuación:

$$\mathcal{L} = \frac{N_1 N_2 f N_b}{4\pi\sigma_x\sigma_y} \quad (2.1)$$

N_1 y N_2 denotan el número de partículas de cada uno de los dos *bunches* de protones que van a colisionar. N_b hace referencia al número de *bunches* y f es la frecuencia de colisión entre *bunches*. En el denominador, el producto de σ_x y σ_y denota la sección eficaz del haz en el plano transversal, sobre el eje X e Y respectivamente..

Otra magnitud de importancia es la luminosidad integrada. Esta luminosidad hace referencia al número total de colisiones que se han producido en un tiempo t . Dicha medida se expresa en unidades de inverso de área como pb^{-1} (picobarns inversos) o fb^{-1} (femtobarns inversos). Se puede calcular de la siguiente forma:

$$L = \int_0^t \mathcal{L}(t') dt' \quad (2.2)$$

La luminosidad integrada permite tener un conocimiento sobre el número total de eventos N que se han producido en un periodo de tiempo t :

$$N = L \sigma \quad (2.3)$$

donde σ denota la sección eficaz.

Pile-up

El término *pile-up* se refiere al número de colisiones simultaneas que ocurren en cada cruce de paquete de protones. Cuando los haces se cruzan en el LHC, cada

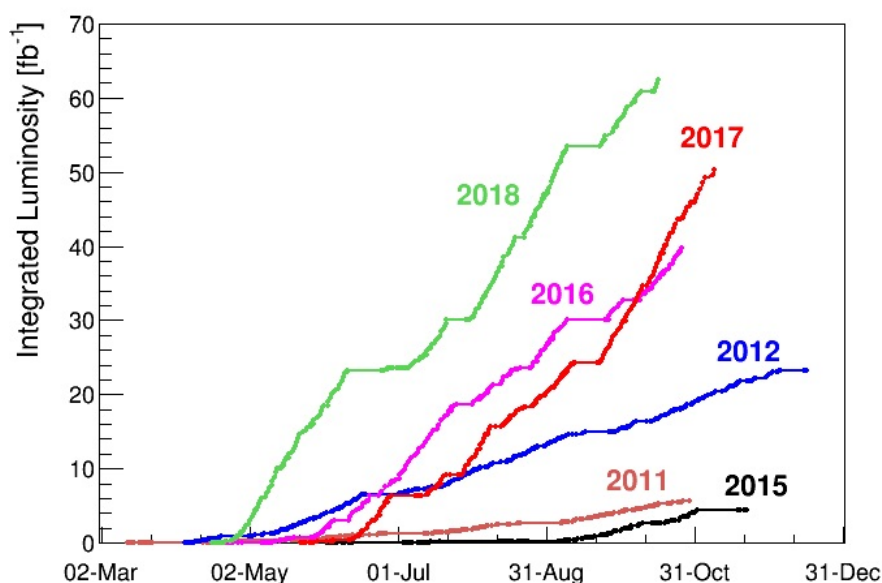


Figura 2.2: Luminosidad integrada con el paso de los años [5].

colisión produce un gran número de partículas secundarias que son detectadas por los distintos sensores del LHC.

Múltiples colisiones pueden ocurrir simultáneamente en el detector de partículas. Esto puede ser problemático porque los detectores tienen un tiempo de lectura finito, lo que significa que solo pueden procesar un número limitado de señales a la vez. Si hay demasiadas señales al mismo tiempo, algunas pueden perderse o ser mal interpretadas, lo que puede comprometer la calidad de los datos.

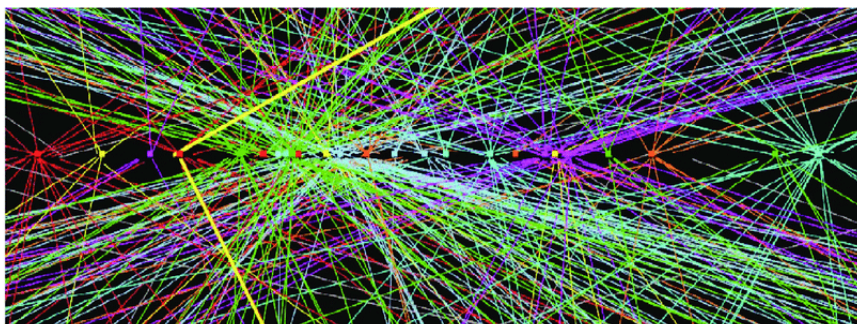


Figura 2.3: Efecto del *pile-up* en el tracker de ATLAS. Dos líneas amarillas gruesas muestran la trayectoria de los estados finales de un bosón Z [6].

El *pile-up* puede ser un problema en el LHC ya que se producen cientos de millo-

nes de colisiones por segundo. Para solventarlo, se diseñan detectores y algoritmos de análisis de datos que puedan manejar el alto número de colisiones por segundo y reducir el efecto del *pile-up* para garantizar que los datos sean precisos y útiles para la investigación.

2.2. Componentes del LHC

La energía máxima de centro de masas \sqrt{s} a la que ha operado el LHC durante la toma de los datos de este experimento ha sido 13 TeV, es decir, 6.5 TeV por beam. Esta alta energía se ha conseguido gracias a los distintos dispositivos que intervienen en cada experimento dentro del LHC, para una descripción más detallada se puede consultar [7].

Las **cavidades de radiofrecuencia**, véase Figura 2.4 son el dispositivo utilizado para acelerar las partículas en el LHC. Existen un total de 16 cuya finalidad es incrementar la energía de los protones desde los 450 GeV que proporciona el *SPS* hasta los 6.5 TeV. Las cavidades de radiofrecuencia se agrupan en 4 grupos que constan



Figura 2.4: Cavidades de radiofrecuencia en el LHC [8].

de 4 cada uno. Como ya se mencionó anteriormente, para el correcto funcionamiento de estas cavidades y, por tanto, para poder alcanzar la energía deseada, es necesario mantener una temperatura muy baja, de esta manera actúan como superconductores. Con este fin, se dispone de 4 refrigeradores en forma cilíndrica denominados

criomódulos. Las cavidades generan campos electromagnéticos que funcionan a una frecuencia de 400 MHz y están sincronizados con el paso del haz. Como resultado, los protones con energías distintas, superiores o inferiores, a la deseada son frenados o acelerados respectivamente y como resultado se obtiene un haz con la energía de 13 TeV deseada para la correcta realización del experimento.

Otro componente de vital importancia en el LHC son los imanes superconductores. Existen varios tipos dependiendo de su función. Los *imanes dipolares* tienen como objetivo el mantener una forma circular para la trayectoria de los protones a lo largo del anillo del LHC. Para mantener esta curvatura, se dispone de 1232 **imanes dipolares** [9] que son capaces de crear un campo magnético de hasta 8 T. Este campo está dirigido en la dirección vertical, perpendicular al plano que encierra el anillo del LHC, de manera que la fuerza de Lorentz es capaz de curvar la trayectoria de las partículas.

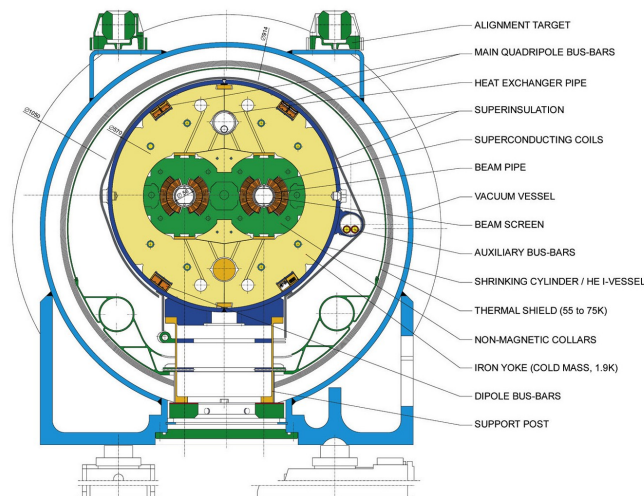


Figura 2.5: Esquema de la sección transversal de los anillos del LHC

[10].

Los **imanes cuadrupolares** se encargan de enfocar el haz antes de la colisión, lo que se traduce en un mayor número de colisiones cuando los dos haces chocan. Estos imanes poseen cuatro polos colocados simétricamente en torno al tubo por el cual circulan los protones. Este enfoque se consigue gracias a que los imanes ejercen fuerzas de compresión sobre los *bunches* que hace que estos se compacten. Tras el cruce de los haces, solo un pequeño número de partículas han colisionado. Las

partículas que no han interactuado, circularán de nuevo por el anillo si se desea reutilizar el haz o por el contrario éste se desvía hacia un bloque de grafito para detenerlo.

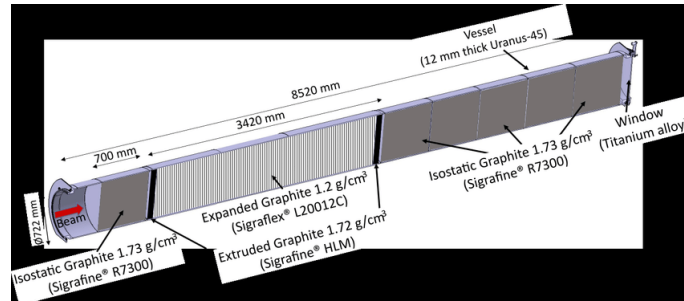


Figura 2.6: Esquema del sistema del sistema de bloques de grafito usados para detener los protones acelerados desechados [11].

En el LHC se realizan varios experimentos simultáneamente, siendo cuatro los detectores donde se hacen colisionar las partículas: CMS, ATLAS, ALICE y LHCb.

2.3. Detector CMS

Los datos empleados en este trabajo han sido recogidos por el detector CMS del LHC [12].

El detector CMS (*Compact Muon Solenoid*) es un detector de partículas de propósito general en el cual se estudian una gran variedad de procesos. Este detector, véase Figura 2.7, posee una forma cilíndrica en cuyo eje central se encuentra el tubo en el cual colisionan los protones acelerados. Posee varios equipos especializados cada uno en la detección de uno o varios tipos de partículas resultantes de la colisión. CMS se divide en dos partes principalmente, dos zonas de menor superficie que se corresponden con las tapas del cilindro (*endcaps*) y la zona más amplia que se corresponde con la superficie circular del cilindro, la cual rodea al solenoide superconductor que es capaz de generar un campo de $3.8 T$. Se trata de un detector gigantesco de 15 metros de alto y unos 21 de largo, que pesa alrededor de 14000 toneladas [13].

Los equipos que forman CMS son el detector de trazas o *tracker*, el calorímetro

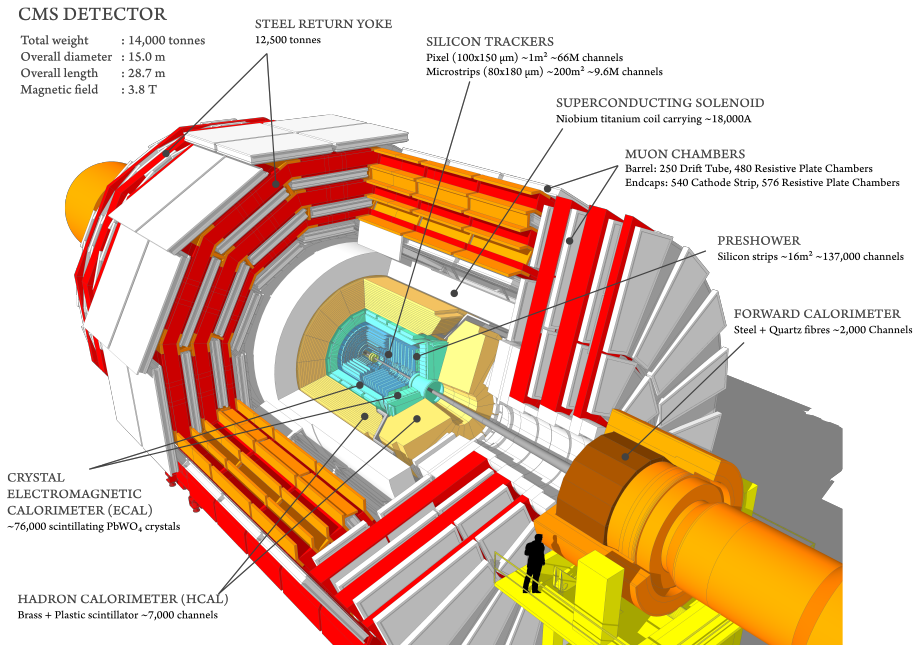


Figura 2.7: Componentes y subdetectores de CMS.[14]

electromagnético (ECAL), el calorímetro hadrónico (HCAL), el solenoide superconductor y las cámaras de muones.

2.3.1. Sistema de coordenadas de CMS.

Antes de explicar las partes del detector CMS, se verá cuál es la manera de indicar la posición de las partículas.

El sistema de coordenadas empleado en CMS tiene el origen en el punto de colisión de los haces denominado *vértice*. Se emplearán coordenadas cilíndricas y los ejes se colocan de manera que el eje Z se extiende a lo largo del haz que va a colisionar, mientras que los ejes X e Y se encuentran en el plano transversal con el eje X apuntando hacia el centro de la circunferencia descrita por el LHC y el Y hacia arriba.

Además de los ejes cartesianos, para describir la posición de las partículas en el plano transversal se usan coordenadas polares:

- El radio r se mide desde el origen hasta el punto en el que la partícula interactúa con el detector.

- El ángulo azimutal $\phi \in [0, 2\pi)$ se mide alrededor del eje z, en el plano transversal respecto del semieje positivo X.
- El ángulo polar $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ se mide respecto del eje Z.

Además de las coordenadas cilíndricas, CMS también utiliza la pseudorrapidez η para describir la posición de las partículas en la dirección del haz de protones. La pseudorrapidez se define como:

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right) \quad (2.4)$$

donde θ es el ángulo polar. La pseudorrapidez es una medida útil ya que en el caso de que la partícula no tenga masa (fotones) o si viaja a una velocidad próxima a la de la luz, la pseudorrapidez se aproxima a la rapidez. La rapidez es una cantidad que es invariante Lorentz y por lo tanto es muy interesante ya que no depende del sistema de referencia.

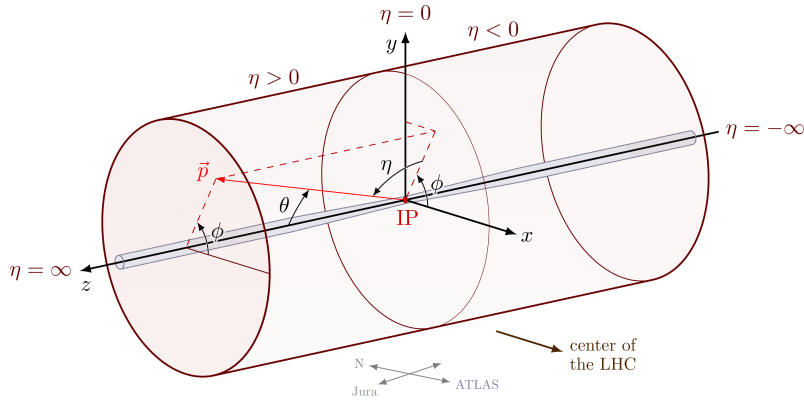


Figura 2.8: Sistema de coordenadas empleado en CMS [15].

Por otro lado se puede definir el momento transversal \vec{p}_T como la proyección del momento lineal de las partículas sobre el plano XY. Se denotará como p_T a su módulo, que viene dado por:

$$p_T = \sqrt{p_x^2 + p_y^2} \quad (2.5)$$

2.3.2. Tracker

Se trata del primero de los detectores con el que se encuentran las partículas tras la colisión. Con este detector se reconstruyen las trayectorias de las partículas que

poseen carga y de esta manera es posible calcular su momento y el lugar exacto en el que se produjo la colisión (*vértice*). Existen dos tipos de vértices; los primarios, aquellos en los cuales se ha producido la primera de las interacciones (colisión) y secundarios, correspondientes a las desintegraciones de partículas. Para hacer posible la reconstrucción de las trayectorias y el cálculo de su momento, el *tracker*, en Figura 2.9, presenta un campo magnético en su interior, de esta manera el camino seguido por las partículas cuyo momento sea menor presenta una mayor curvatura que aquellas que poseen un mayor momento. Dado que este detector no modifica apenas la trayectoria de las partículas, las mediciones posteriores no se verán alteradas y vendrán dadas con gran precisión. El *pile-up* dificulta esta reconstrucción de las trayectorias, ya que hace que el número de señales notadas por el *tracker* aumente y por tanto lo hará también el número de posibles trayectorias. Las señales que no se correspondan con sucesos de interés serán descartadas. Los algoritmos de reconstrucción de las trayectorias otorgan una precisión del orden de $10 \mu m$ [16].

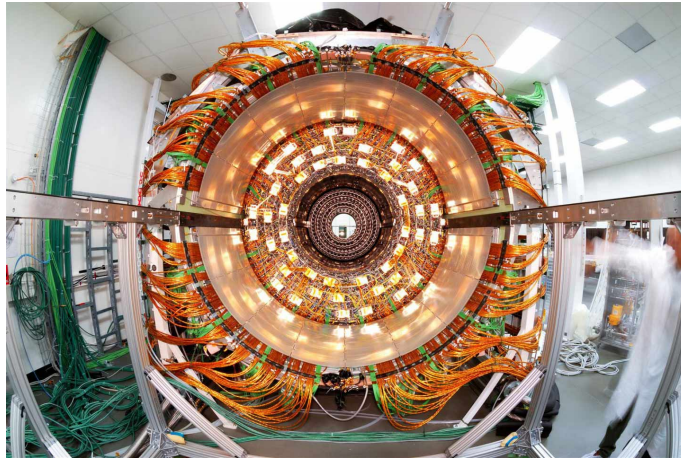


Figura 2.9: Foto del detector de trazas de CMS [17].

El detector de trazas se dispone en cilindros concéntricos con tapas alrededor del tubo donde se produce la colisión de las partículas cubriendo un área de cerca de $200 m^2$. Este dispositivo es capaz de detectar señales de partículas cuya pseudorapidez es $|\eta| < 2.5$. Posee dos elementos principales que emplea para la detección: los píxeles de silicio, ubicados en la parte más interna del *tracker* y las tiras de silicio. El *tracker* posee 124 millones de píxeles, lo que le confiere una gran precisión en la medición de las trazas y es de gran utilidad en la reconstrucción de las trayectorias de partículas

con un tiempo de desintegración muy bajo ya que se encuentra a escasos 3 *cm* del tubo donde colisionan las partículas. Cuando una partícula cargada atraviesa un píxel aporta la suficiente energía para hacer excitar los electrones más externos del silicio. Al aplicar un voltaje a estos sensores, se produce una pequeña corriente que será amplificada y de esta manera se puede realizar un seguimiento de los píxeles que han sido atravesados [18]. Las tiras de silicio están compuestas por dos sensores de silicio, una estructura que hace de soporte y los componentes electrónicos para la lectura de las señales. Tienen un funcionamiento análogo a los píxeles, encontrándose en la parte más externa del tracker y están dispuestas en capas hasta a 130 *cm* del tubo [19].

2.3.3. Calorímetros

Los calorímetros cumplen la función de medir la energía de las partículas. Al ser atravesados por una partícula ésta colisiona con el material denso que lo constituye. Este choque dará lugar a un chorro de partículas secundarias y la energía de la partícula original es absorbida por el calorímetro. Es por esto que los calorímetros rodean al detector de trazas, pues se determina la trayectoria de la partícula antes de que ésta sea absorbida. Existen partículas que se escapan a los calorímetros como los muones y los neutrinos, que continuarán hasta las cámaras de muones. En el detector CMS existen dos tipos de calorímetros: electromagnético y hadrónico [20].

Calorímetro electromagnético (ECAL)

Su principal función es la detección de electrones y fotones. Este calorímetro posee una gran resolución en energía, posición y ángulo, cubriendo hasta una pseudorrapidez $|\eta| < 1.48$ en la zona del barril (superficie circular del cilindro) y $1.63 < |\eta| < 3$ para las dos secciones en las *endcaps* (tapas). También ofrece una respuesta rápida de 25 *ns*. Está formado por 80000 cristales de tungstenato de plomo, material muy denso, que pesan alrededor de 1.5 *kg* [21]. Cuando un electrón o un fotón atraviesa uno de estos cristales se produce un destello. Los cristales están dispuestos junto a fotodetectores que captan dicho destello y lo transforman en una señal eléctrica.

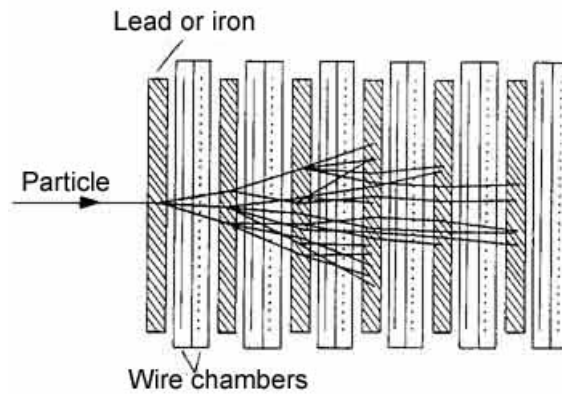


Figura 2.10: Descripción del funcionamiento de un calorímetro. Una partícula entra y colisiona con una placa de plomo o hierro. Así interactúa con los átomos de la placa y se crean partículas secundarias. Éstas se detectan en una cámara de trazas junto a la placa (tiras de silicio). Este proceso continúa a través de muchas capas de placas de plomo/hierro y cámaras de hilos hasta que un pequeño chorro de partículas secundarias han sido creadas. La energía de la partícula entrante puede calcularse midiendo la actividad en el interior de las cámaras de tiras de silicio [20].

Para ofrecer una mejor resolución espacial el ECAL, Figura 2.11 posee los *preshower detectors* situados delante de las dos zonas del *endcap*. Estos detectores permiten distinguir entre fotones aislados muy energéticos y pares de fotones poco energéticos. La función principal del ECAL Preshower es la de distinguir los fotones provenientes de la desintegración del Bosón de Higgs gracias a que su granularidad es mucho más fina que la que ofrecen los cristales del ECAL.

Calorímetro Hadrónico (HCAL)

Está situado en una capa aún más externa que el calorímetro electromagnético. Su principal función es la detección de los hadrones producidos en la colisión, tanto de aquellos que poseen carga como de los neutros. Se encarga de medir la energía de estos hadrones e indirectamente ofrece una medida de la presencia de partículas como los neutrinos [23]. La desintegración de estas partículas puede producir otras

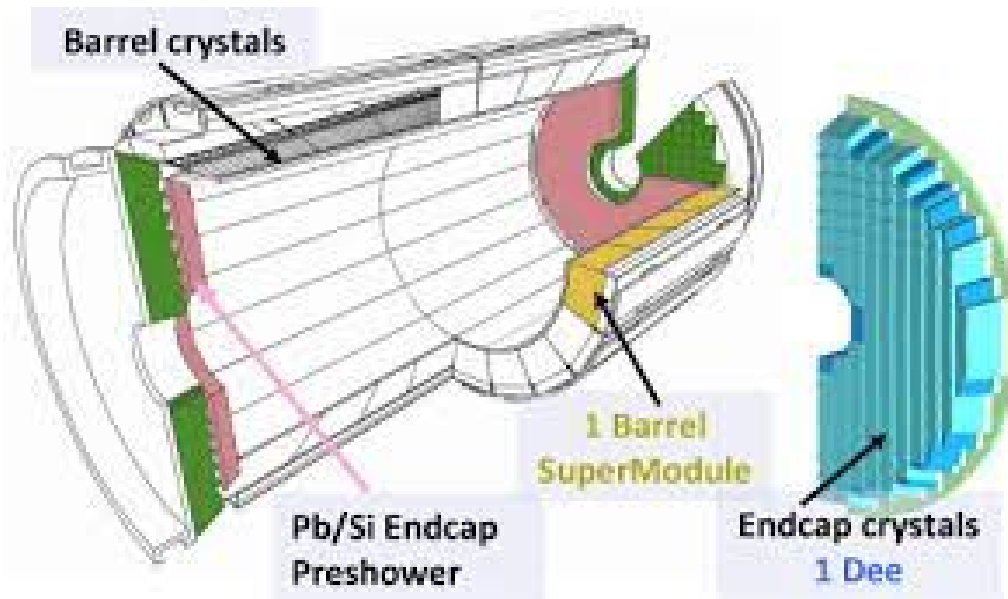


Figura 2.11: Esquema del calorímetro electromagnético [22].

nuevas que no dejan ningún rastro en los detectores de CMS, por ello HCAL ha de ser hermético para asegurarse de que captura toda partícula emergente de la colisión. Gracias a esto, se pueden medir desequilibrios en el momento y energía, es decir, que el momento y la energía no se conservan, de manera que es posible deducir que existen partículas “invisibles” que deben hacer que se cumpla la conservación de momento y energía. El HCAL se divide en dos partes: la zona del barril donde se encuentra el barril hadrónico (HB) y el calorímetro exterior (HO) y la zona de las endcaps donde se encuentran el endcap hadrónico (HE) y dos secciones frontales (HF). Cada una de estas zonas está formada por la superposición de capas de material absorbente y capas de material centellador. Cuando los hadrones alcanzan las capas de material absorbente, interactúan con éstas y se producen numerosas partículas secundarias. Éstas a su vez interactúan con otras capas produciéndose así una reacción en cadena que conforma un chorro de partículas que interactúan con el material centellador. Debido a ello, este material emite una señal luminosa que mediante fibras ópticas se transmite a dispositivos de lectura que se encuentran en el propio calorímetro. La superposición e interpretación de estas señales dará cuenta de la trayectoria de la partícula y de su energía.

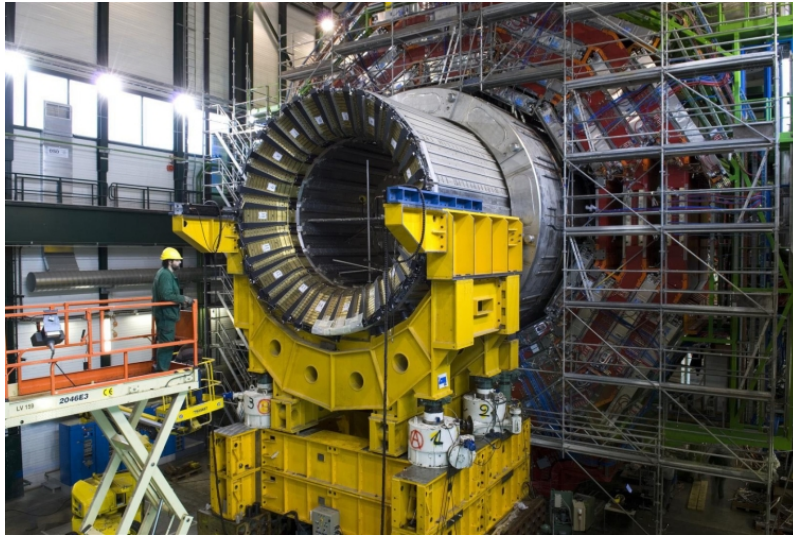


Figura 2.12: Foto del calorímetro hadrónico durante su montaje en CMS [23].

2.3.4. Solenoide superconductor

Se trata de un electroimán que genera un campo magnético de hasta $3,8 T$ y cuyo objetivo es curvar la trayectoria de las partículas emergentes tras la colisión. Al curvar la trayectoria de estas partículas es posible determinar su momento y su carga eléctrica. Dependiendo del signo de su carga eléctrica la partícula se curvará en sentido horario o antihorario y cuanto mayor momento posea la partícula menos curvatura tendrá su trayectoria [24].

El solenoide está formado por hilos de niobio-titanio (NbTi) superconductores. Este material de NbTi es capaz de transportar corriente eléctrica sin resistencia a temperaturas extremadamente bajas, lo que permite que las bobinas no ofrezcan resistencia y se generen campos magnéticos muy intensos y estables [25]. Para alcanzar estas temperaturas tan bajas, está dotado de sistemas de refrigeración que disminuyen su temperatura hasta unos $4 K$. El solenoide rodea a los calorímetros y al detector de trazas y éste a su vez está rodeado por las cámaras de muones. A su vez el solenoide se encuentra en el interior de un armazón de hierro que cumple la función de confinar el campo magnético y servirle como guía.

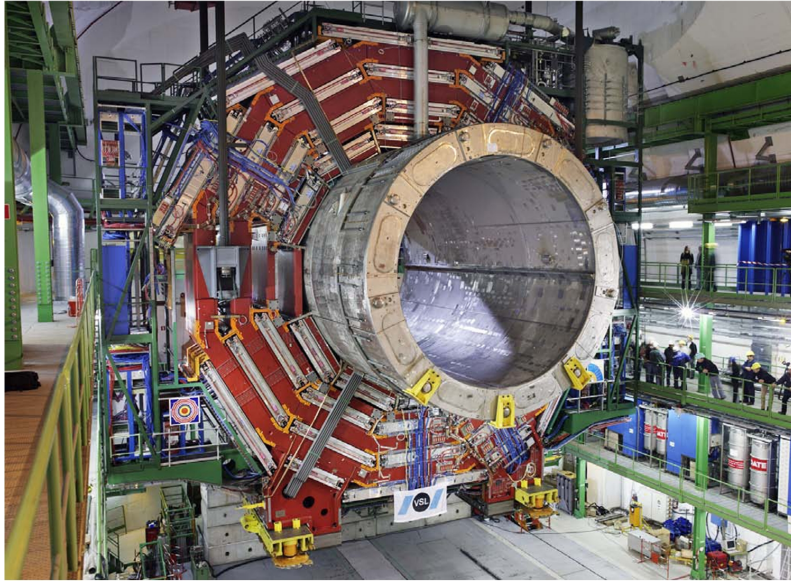


Figura 2.13: Imagen del solenoide superconductor que envuelve los calorímetros [26].

2.3.5. Cámaras de muones

Las cámaras de muones en CMS están ubicadas en la capa exterior del detector rodeando a todos los detectores anteriores y se utilizan para medir la trayectoria y la energía de los muones que pasan a través de ellas.

Existen tres tipos principales de cámaras de muones utilizadas en CMS: cámaras de deriva, cámaras de resistividad y cámaras de tiras catódicas.

Las cámaras de deriva o *drift tubes* (DTs) utilizan un gas noble, generalmente argón o xenón, para detectar los muones. Están situadas en el barril. Cuando un muón atraviesa la cámara de deriva, ioniza los átomos del gas, lo que produce electrones libres. Estos electrones se mueven hacia un alambre central cargado positivamente, donde se recogen y se miden para determinar la posición del muón.

Las cámaras de resistividad (RPCs), situadas tanto en el barril como en las tapas, también utilizan un gas noble, pero en este caso los átomos del gas se encuentran en un estado excitado. Cuando un muón atraviesa la cámara de resistividad, se produce una cascada de electrones que se mueve hacia los electrodos de lectura, que miden la posición del muón.

Las cámaras de tiras catódicas (CSCs), colocadas en las tapas, utilizan una tecnología diferente y más reciente que se basa en la detección de la luz. Estas cámaras utilizan tiras de plástico o de cristal recubiertas de un material sensible a la luz para detectar los muones. Cuando un muón atraviesa la tira, produce una emisión de luz que se detecta y se mide para determinar su posición.

CMS cuenta con 250 DTs que cubren un ángulo de $|\eta| < 1,2$, 500 CSCs que cubren un ángulo de $0,9 < |\eta| < 2,4$ y 610 RPCs que cubren un ángulo de $|\eta| < 1,9$.

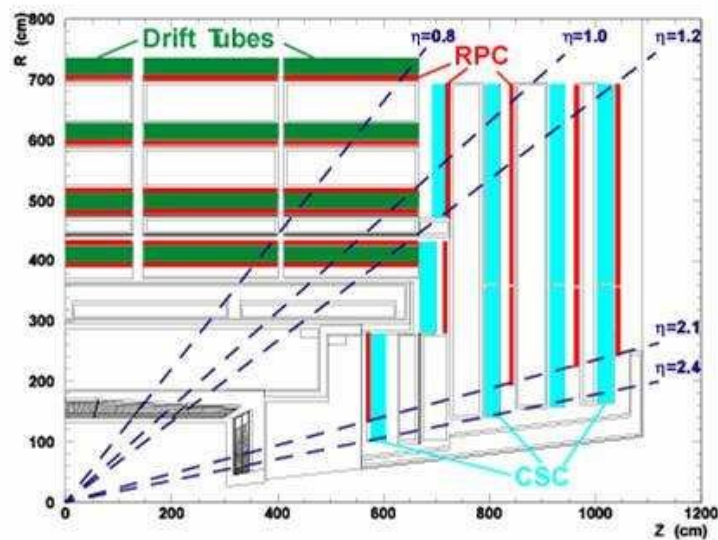


Figura 2.14: Esquema de la colocación de las cámaras de muones en un corte longitudinal de CMS [27].

Las cámaras de muones de CMS tienen una alta resolución espacial y temporal, lo que permite una medición precisa de la trayectoria y la energía de los muones. Estas propiedades son esenciales para la identificación de muones y la reconstrucción de eventos en el detector CMS [28].

2.3.6. Trigger

El trigger del CMS es un sistema para el procesamiento de los datos que provienen de los subdetectores de partículas del CMS y su principal característica es que decide cuáles deben ser almacenados para un análisis posterior. Dado que el LHC produce colisiones de protones a una tasa muy alta (alrededor de 40 MHz),

es imposible almacenar todos estos datos. En su lugar, el trigger del CMS tiene la tarea de seleccionar sólo los eventos que resultan de interés para su posterior estudio y análisis.

El trigger del CMS consta de dos niveles: el nivel 1 (L1) y el nivel 2 llamado *High Level Trigger* (HLT). El L1 es un sistema de hardware que procesa los datos de forma rápida y decide si un evento es interesante o no. El L1 tiene una capacidad de procesamiento de los datos de hasta 100 kHz y se basa en información de los detectores de partículas más rápidos del CMS, como los detectores de muones (cámaras de muones) y los calorímetros [29]. El L1 identifica rápidamente los eventos de interés y los guarda temporalmente en un búfer mientras espera la confirmación del HLT.

El HLT es un sistema de software más sofisticado que realiza una selección más fina de los eventos interesantes seleccionados por el L1. El HLT tiene una capacidad de procesamiento de alrededor de 1 kHz y utiliza información más detallada de los detectores de partículas del CMS, como el detector de trazas y los calorímetros. También usa algoritmos de filtrado más avanzados para seleccionar los eventos que van a ser guardados para su posterior análisis.

La selección de eventos interesantes mediante el trigger del CMS se basa en diferentes criterios, como la energía, la cantidad y la dirección de las partículas detectadas. El trigger del CMS debe ser lo suficientemente sensible para detectar partículas raras y eventos poco frecuentes, pero al mismo tiempo debe ser selectivo para evitar el almacenamiento de datos redundantes.

Capítulo 3

Gestión de la información

En el capítulo anterior se trató el tema de la recolección de los datos. En este capítulo se verá como estos datos son tratados para obtener la información que resulta de interés en un determinado estudio.

Tras la recolección de los datos en el LHC, el siguiente paso consiste en un proceso de reconstrucción, en el que la información previamente adquirida por cada subdetector es tomada en cuenta en su conjunto para diferenciar el suceso que se ha producido. Dado que hay diversos registros a tener en cuenta, este proceso se lleva a cabo una vez han sido almacenados los datos. Cuando se está estudiando un proceso físico en concreto, es necesario realizar una selección de los datos que resultan de interés. Estos datos constituyen la *señal* y el resto de datos que pueden ser confundidos con la señal se denominan *fondo*. Para la obtención de unos resultados con la mayor veracidad posible es importante maximizar la señal sobre el fondo. Con el objetivo de recrear los datos experimentales se llevan a cabo simulaciones de Monte Carlo de los procesos físicos con las mismas condiciones en las que se produciría en el LHC. Este proceso de simulación es conocido como *generación*. Tras la simulación de las colisiones, esta información se hace pasar por el detector (CMS) de la misma forma que lo harían las partículas en un experimento real. Por último, esta información sigue el mismo proceso de reconstrucción que los datos reales. Teniendo en cuenta la información obtenida de las simulaciones, es posible comparar los resulta-

dos esperados de las simulaciones con los datos así como las distintas eficiencias de ciertos procesos. De esta forma se comprueba que los resultados experimentales y los modelos teóricos son coherentes o por el contrario, se avanza en el descubrimiento de nueva física.

3.1. Simulación de sucesos

Como ya se ha comentado, la herramienta para la generación de datos procedentes de las colisiones de partículas de manera estocástica, son las simulaciones de Monte Carlo (MC). Las muestras MC son el resultado obtenido de una simulación del proceso y de la simulación del paso de las partículas por el detector.

Los programas más empleados en la simulación de sucesos son *MadGraph*, *Powheg* y *aMC@NLO*. Estos serán combinados con el software *Pythia* que se encarga de la construcción de la cascada de interacciones que se producen a raíz del suceso principal [30]. Por último el software *Geant4* simula las interacciones de los distintos procesos que se han producido con el detector CMS. Dado que la física del SM es de carácter perturbativo, las simulaciones se clasifican según el orden de perturbación que tienen en cuenta. El orden fundamental o *leading order* (LO) es el usado por *MadGraph* y primer orden o *next-to-leading order* (NLO) es el usado por *Powheg* y *aMC@NLO*.

Cuando se realizan las simulaciones, el número de eventos simulados es mucho mayor que el de datos recogidos en el LHC, esto es así ya que de esta manera se tiene información suficiente como para descartar resultados erróneos como consecuencia de una baja estadística. Es por ello que estos procesos simulados son dotados de un peso que tiene en cuenta el número total de sucesos de la muestra MC y la sección eficaz de dicho proceso a la energía en que se produce en el LHC.

Siguiendo lo antes comentado, tras la generación de los procesos y la simulación de su paso por el detector, se lleva a cabo la reconstrucción de los eventos que se hará de la misma forma que para los datos experimentales.

3.2. Reconstrucción de sucesos

El algoritmo que se encarga de la reconstrucción de los eventos es el *particle-flow* (PF) [31]. Este algoritmo usa de manera conjunta la información proporcionada por los subdetectores y la clasifica obteniendo posibles objetos físicos como jets, muones, electrones, hadrones cargados y neutros y sus características (p_T , ϕ , η , m , etc).

Las partículas que atraviesan CMS describen trayectorias y tienen alcances diferentes debido a su naturaleza. Las partículas con carga curvan su trayectoria debido al campo magnético del supersolenoide. Al atravesar el tracker, dejan una señal o *hit*. Los electrones son capaces de alcanzar el ECAL, al igual que los fotones, que es donde son detectados pero a diferencia de estos últimos, los electrones poseen carga y dejan señal en el tracker. Los hadrones cargados y neutros alcanzan el HCAL, pero los primeros curvan su trayectoria y los segundos no. Los muones por su parte, curvan su trayectoria e interactúan con el tracker ya que son partículas cargadas y atraviesan los subdetectores hasta llegar a las cámaras de muones y las atraviesan saliendo al exterior.

En primer lugar el algoritmo PF se encarga de la reconstrucción de las trazas mediante un proceso iterativo. Como ya se había comentado las partículas al atravesar el tracker y las cámaras de muones producen *hits*. El algoritmo PF en una primera iteración realiza la reconstrucción de las trazas que han sido medidas con una mayor precisión y en las iteraciones sucesivas, el criterio es cada vez menos estricto, de manera que el número de trayectorias recogidas aumenta en cada iteración. Otro elemento que se reconstruye son los *clusters* en los calorímetros. Este proceso consiste en identificar los depósitos de energía que se encuentran próximos y que habría realizado una sola partícula en su interacción con el detector.

Una vez reconstruida la traza y los clusters, PF trata de conectar las distintas características medidas con el fin de construir candidatos a objetos físicos.

Los primeros objetos que se reconstruyen son los muones, ya que resultan más fáciles de identificar pues son las únicas partículas que producen *hits* en las cámaras de muones. Seguidamente, los electrones y hadrones cargados ya que se conocen las

trazas y las depósitos en el ECAL y HCAL. Una vez se conocen los electrones, el resto de clusters del ECAL habrán sido producidos por los fotones y los que queden en el HCAL conocidos los hadrones cargados, se corresponderán con los hadrones neutros.

En la Figura 3.1 se pueden observar el funcionamiento antes descrito.

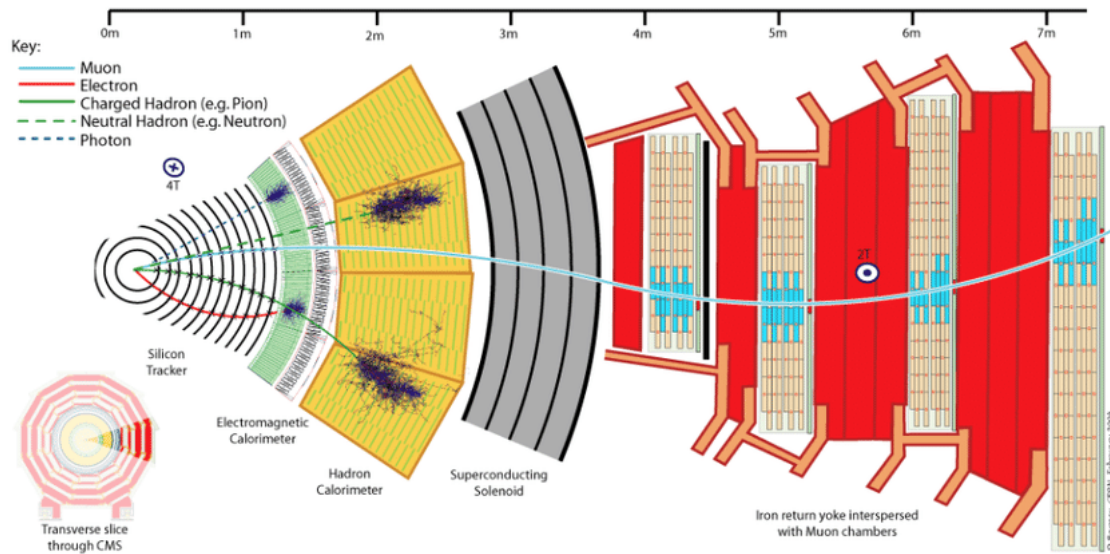


Figura 3.1: Plano transversal de CMS y sus subdetectores [32].

Tras la reconstrucción de los objetos físicos se procede a la reconstrucción del cuadrimomento de cada objeto con la información recogida por los subdetectores.

3.2.1. Vértice principal

El *vértice principal* (VP) es el punto del espacio en el que se produjo la colisión protón-protón del evento estudiado. El algoritmo *particle-flow* es capaz de identificar este punto reconstruyendo las trayectorias de las partículas que han sido medidas por los subdetectores hasta que éstas confluyen en un único punto en el interior del tubo donde se propagan los haces de protones. Debido al *pile-up*, hecho por el cual se puede detectar más de una colisión al mismo tiempo, se tomará como vértice principal aquel que tiene un mayor momento transversal. El algoritmo ha de ser lo suficientemente preciso como para diferenciar dos colisiones próximas pero no

demasiado ya que podría dividir las trayectorias de los productos de una colisión en varias distintas.

3.2.2. Muones

Para la reconstrucción de los muones, en primer lugar se reconstruyen sus trayectorias usando los *hits* producidos en las cámaras de muones y en el *tracker* y teniendo en cuenta las posibles pérdidas de energía que sufre el muón en su paso por el detector así como el cambio de dirección de la trayectoria al pasar por el solenoide. las partículas así reconstruidas se denominan *global muons*. Una vez se tiene reconstruida la partícula, para identificarla como muón, PF procede a examinar sus características aplicando criterios de calidad y aislamiento adicionales. En primer lugar cabe diferenciar entre *prompt* muón, aquel que proviene de las desintegraciones producidas en el vértice principal y *non-prompt* muón, producidos por las desintegraciones de hadrones en el paso de las partículas por los detectores. Para realizar esta clasificación se aplica un criterio de aislamiento relativo, I . Esta variable se define como la suma de los módulos de los momentos trasversos de las partículas que se encuentran en el interior del cono definido por $\Delta R := \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ y cuyo vértice es el muón.

$$I = \sum_{particulas} \frac{p_T}{p_T^{muon}} \quad (3.1)$$

Cuantas más partículas rodean al muón en cuestión mayor será esta variable. Posteriormente se construyen variables cuyo fin es permitir clasificar estas partículas que se consideraban muones de forma más o menos estricta. Se tienen, por ejemplo, los *loose* muones, aquellas partículas que se consideran muones pero no se tiene una certeza asegurada de ello y los *tight* muones que se conoce casi con certeza que son muones. Existen muchas otras variables que permiten realizar cribas para quedarse los objetos de interés, como el *Lepton MVA* (algoritmo de análisis multivariable) que se usará en este trabajo. Estas variables se construyen con el objetivo de que los diversos estudios que se realizan dentro de CMS puedan utilizar la información deseada.

3.2.3. Electrones

Para la reconstrucción de los electrones se procede de forma análoga a la de los muones. La diferencia radica en que a la hora de reconstruir las trayectorias de los electrones hay que tener en cuenta que debido a que su masa es mucho menor que la de los muones, éstos pierden energía en forma de radiación *Bremsstrahlung* lo que causa una pérdida de la forma helicoidal en la trayectoria. El algoritmo PF actúa como lo hacía para los muones, construye variables que asocia a cada electrón de la misma forma que lo hacía para los muones. Se tienen así electrones *prompt*, electrones *non-prompt* electrones, *loose*, electrones *tight*, etc.

3.2.4. Hadrones y jets

Debido a la interacción fuerte, los *quarks* producidos en una desintegración se hadronizan. Existe una excepción y se trata del *top*. Dado que presenta una vida media muy corta decae en otras partículas y no le da tiempo a hadronizarse. En estos procesos de hadronización se forman partículas y se desintegran en otras (leptones, hadrones, neutrinos, etc). Este conjunto de partículas es un *jet*. La reconstrucción de los *jets* se lleva a cabo a partir de los objetos reconstruidos por PF, que agrupa las partículas que pertenecen a cada uno. El algoritmo que se encarga de esta reconstrucción se llama *anti - k_T* [33]. No es posible distinguir el *quark* que originó el *jet* en la mayoría de ocasiones cuando procede de un b; dado que los hadrones que forma tienen una vida media mayor, se desplazan una cierta distancia desde el punto de colisión hasta que decaen, originando un vértice secundario. Gracias a la resolución del tracker es posible distinguir estos vértices secundarios y por tanto qué *jets* provienen de un b, *b-jets*. Los algoritmos que se encargan de este análisis se denominan algoritmos de *b-tagging* y usan técnicas MVA.

3.2.5. Momento transverso faltante

Como ya se ha comentado previamente, existen desintegraciones en las que se producen neutrinos. Estas partículas no interaccionan con ninguno de los subdetectores de CMS por lo que su energía no puede ser medida. Para determinarla se

aplican las leyes de la conservación de la energía y el momento. En una colisión pp al ser frontal, se espera que el momento total en el plano transversal tanto inicial como final sea 0. Cuando esto no se da se atribuye esta energía faltante a los neutrinos y se llama *Missing Transverse Energy* (MET). Se calcula como sigue:

$$\vec{E}_T^{MISS} = - \sum_{particulas} \vec{p}_T \quad (3.2)$$

Capítulo 4

Planteamiento del estudio

En este trabajo se propone la comparación de dos técnicas de distinta naturaleza, un algoritmo de BDT (*Boosted Decision Trees*) como es el *Lepton MVA* y el método *CutBased*, a la hora de reducir las contribuciones del canal $t\bar{t}$ semileptónico al fondo del estudio realizado en [1] para la medida de la sección eficaz del proceso tW .

Como se explicará más adelante, uno de los productos del proceso $t\bar{t}$ semileptónico son *jets* que pueden ser identificados erróneamente como leptones, los denominados leptones *non-prompt*.

Aplicando el *Lepton MVA*, se busca reducir el número de leptones *non-prompt* y de esta manera reducir las contribuciones del proceso $t\bar{t}$ semileptónico a la incertidumbre en los cálculos realizado en [1]. Esta tarea es interesante pues estas falsas identificaciones son las causantes de la segunda mayor fuente de incertidumbre que aparece en [1].

En los resultados presentados en este trabajo se han usado unidades naturales ($c=1$).

4.1. Objetivos del trabajo

Como ya ha sido comentado, este trabajo se centra en la reducción de las contribuciones del proceso $t\bar{t}$ semileptónico al fondo de eventos del proceso tW dileptónico.

Para ello se comparará la eficiencia de dos métodos de distinta naturaleza como son el método *cut-based* y el *Lepton MVA*.

La motivación de este estudio viene dada por la búsqueda de una reducción de la incertidumbre en la medida de la sección eficaz del proceso tW dileptónico que se realiza en [1].

La notación usada para hacer referencia al proceso $t\bar{t}$ semileptónico en [1] será *Non-W/Z* ya que estos eventos que contribuyen al fondo del tW únicamente tienen un leptón y el otro es un jet que ha sido reconstruido erróneamente como un leptón (también podría darse el caso de que fuera un leptón real proveniente de la desintegración de un hadrón pero que no ha sido incluido en el *jet* al que pertenece). A estos leptones se les pone la etiqueta de *non-prompt*. Es decir, se trata de un candidato a leptón que no proviene de la desintegración de un bosón.

En [1], se tiene que estos leptones erróneamente reconstruidos constituyen la segunda mayor fuente de incertidumbre en la medida de la sección eficaz del tW dileptónico, véase Figura 4.1.

Aplicando el *Lepton MVA* y el método *cut-based* se busca reducir el número de leptones *non-prompt* y de esta manera minimizar la incertidumbre que éstos causan.

4.2. Producción de quarks top

Tras la colisión de los protones, se producen diversas partículas objeto de estudio. Como se ha visto en las secciones anteriores, CMS recopila toda la información necesaria para la reconstrucción de los eventos, y lo hace a partir de los productos finales. Por esta razón, a la hora de estudiar cierto proceso es necesario saber cuáles son sus posibles estados finales.

Debido a su alta masa, el quark top decae débilmente en un bosón W y un quark tipo (d , s , b) [34].

$$t \rightarrow W + q \quad \text{con} \quad q = d (0.007\%), s (0.017\%), b (99.82\%). \quad (4.1)$$

Por tanto, casi en su totalidad se descompone en un bosón W y en un quark b. Debido

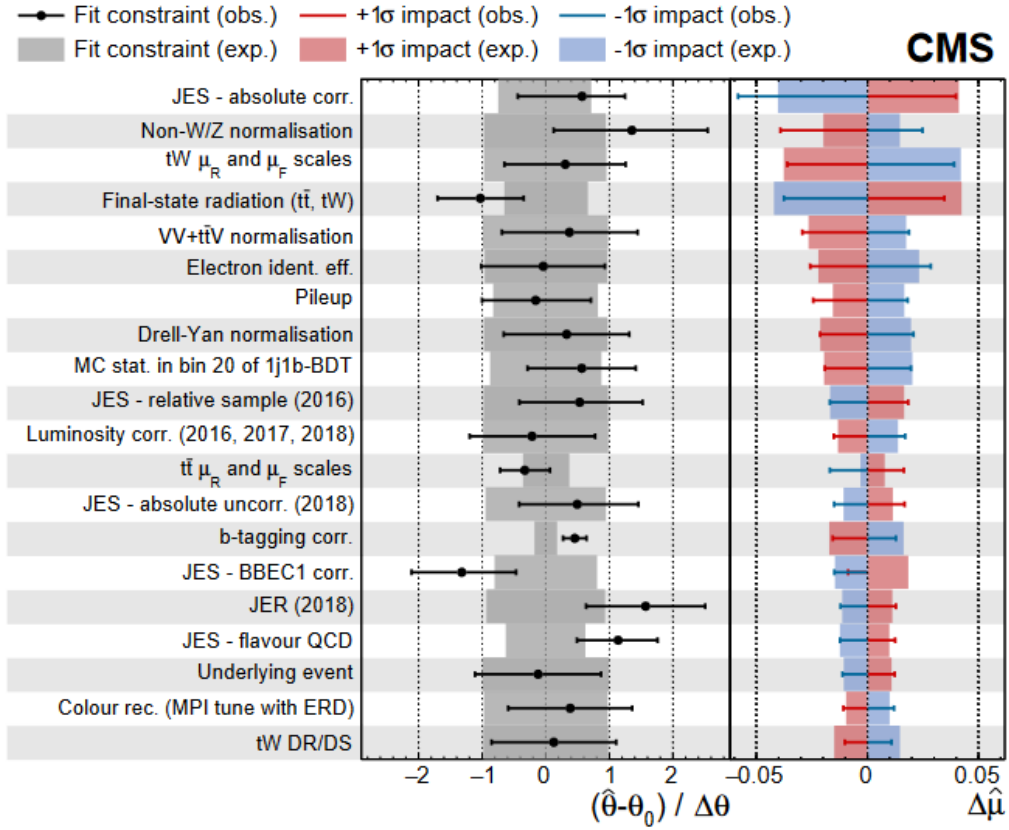


Figura 4.1: Los 20 mayores impactos $\Delta\hat{\mu}$ (columna derecha) y desviaciones $(\hat{\theta} - \theta_0)/\Delta\theta$ (columna central) de los parámetros de incertidumbre enumerados en la columna izquierda provienen del ajuste de *Maximum Likelihood* (ML) utilizado para determinar la sección eficaz inclusiva de tW en [1].

a su alta masa, su vida media es muy corta $\tau = 0.5 \times 10^{-24} s$ y como consecuencia el quark top decae antes de formar hadrones. Este hecho lo diferencia del resto de quarks del SM y de ahí la importancia de su estudio.

Para la producción de un quark top existen dos posibles mecanismos. El primero, mediante la interacción electro-débil. Se produce un único quark top por evento, en Figura 4.6, como es el caso de estudio en el artículo [1]. El segundo, mediante interacciones fuerte, Figura 4.2, se producen pares top anti-top $t\bar{t}$.

Dado que prácticamente el 100% de las veces el quark top decae en un bosón W y un quark b, un par $t\bar{t}$ decaerá en dos bosones W y dos quarks b [36]. Los quarks b

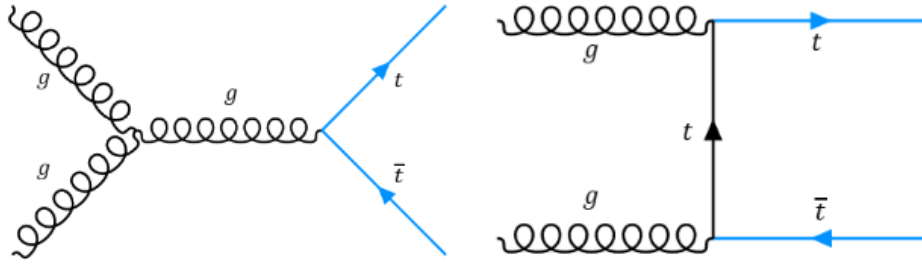


Figura 4.2: Diagramas de Feynman para la producción de pares $t\bar{t}$ mediante la fusión de gluones. Fuente [35].

se detectan experimentalmente como jets muy energéticos mientras que los bosones W pueden decaer o bien hadrónicamente, o bien leptónicamente.

- A. $t\bar{t} \rightarrow W^+ b W^- \bar{b} \rightarrow q \bar{q}' b q'' \bar{q}''' \bar{b}$,
- B. $t\bar{t} \rightarrow W^+ b W^- \bar{b} \rightarrow q \bar{q}' b \ell^- \bar{\nu}_\ell \bar{b} + \ell^+ \nu_\ell b q'' \bar{q}''' \bar{b}$,
- C. $t\bar{t} \rightarrow W^+ b W^- \bar{b} \rightarrow \ell^+ \nu_\ell b \ell'^- \bar{\nu}_{\ell'}$.

Figura 4.3: Resumen de las posibles desintegraciones del proceso $t\bar{t}$.

Se hablará de desintegración leptónica cuando ambos bosones lo hacen a leptón más neutrino. El decaimiento hadrónico se da cuando ambos se desintegran en un par $q\bar{q}$, dando lugar a un jet por cada uno de los quarks, ya que se hadronizan. Se hablará de decaimiento semileptónico, proceso que va a ser objeto de este estudio, cuando uno de los bosones W decae hadrónicamente y el otro lo hace leptónicamente. En la Figura 4.3, aparecen recogidos las posibles desintegraciones para el canal $t\bar{t}$.

Se diferencian también las posibles desintegraciones $t\bar{t}$ en cuanto al número de jets, en el decaimiento leptónico existen dos jets provenientes de los dos quarks b, en el semileptónico cuatro jets, dos b-jets y dos jets provenientes de los quarks en que decae el bosón W y en el hadrónico 6 jets, dos b-jets y un par de la desintegración de cada bosón W.

La probabilidad de desintegración en los diferentes estados finales se denomina *Branching Ratio* (BR) y para el proceso $t\bar{t}$ se presenta en la siguiente figura.

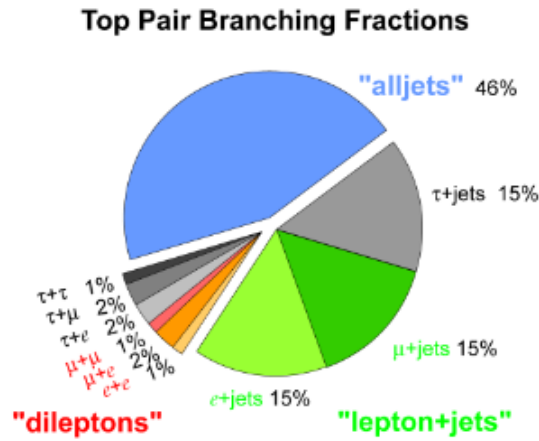


Figura 4.4: Contribuciones relativas de los distintos canales de desintegración del par $t\bar{t}$. Fuente [36].

4.2.1. Proceso $t\bar{t}$ semileptónico

Como ya se había comentado el modo de desintegración semileptónica consiste en la desintegración hadrónica ($q\bar{q}$) de uno de los bosones W y en la desintegración leptónica (e, μ o τ y su correspondiente neutrino) del otro. Los *jets* pueden ser ligeros

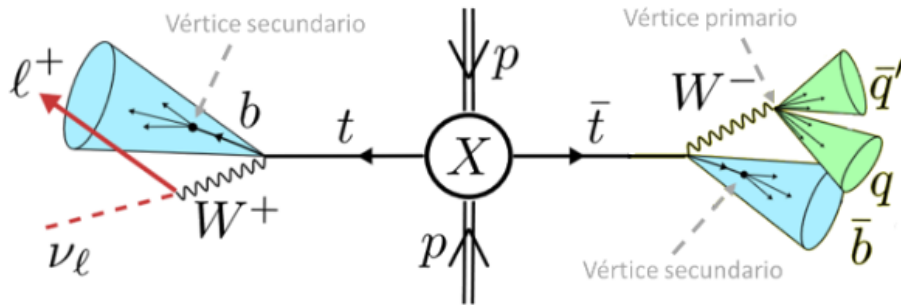


Figura 4.5: Esquema de la desintegración semileptónica $t\bar{t}$. Imagen extraída de [35].

(quarks u, d principalmente) o pesados (quark b). Generalmente los *jets* resultantes de la desintegración de bosones W son ligeros. Los quarks b al hadronizarse forman mesones B cuya principal característica es que son capaces de viajar una distancia resoluble por el detector antes de formar *jets*. Mientras que los *jets* ligeros se forman al hadronizarse el quark (vértice primario), los *jets* pesados surgen transcurrido un corto (pero medible) espacio, en el vértice secundario. Esta diferencia se puede ver

en el esquema de la desintegración semileptónica de $t\bar{t}$ de la Figura 4.5.

Algunas de las características cinemáticas más relevantes de este tipo de desintegraciones son: el alto momento transversal (p_T) del leptón (dada la baja masa del neutrino y la alta masa del bosón W , por conservación de la energía, el leptón es altamente energético); la energía faltante transversa (MET) se asocia a la energía de neutrinos al no ser detectados; la medida del vértice secundario respecto del primario sirve para identificar b-jets mediante lo que se denomina *b-tagging*.

4.2.2. Proceso tW

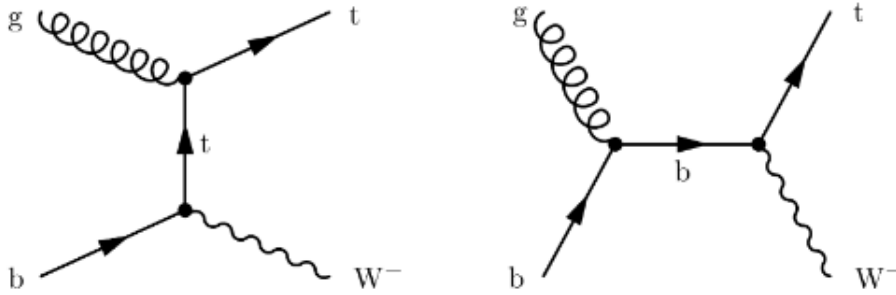


Figura 4.6: Diagramas de Feynman LO para la producción de quarks top en el canal tW . Fuente [1].

El objetivo de [1], que es el estudio en el que se basa este trabajo, es medir la sección eficaz del proceso tW con estados finales dileptónicos ($e^\pm\mu^\mp$). El proceso tW presenta una señal muy pequeña en comparación con el fondo que puede ser del 80 % en el mejor de los casos [1]. Dos de los procesos que contribuyen a este fondo son los ya mencionados $t\bar{t}$ y $t\bar{t}$ semileptónico. Las contribuciones al fondo de los procesos $t\bar{t}$ no pueden ser reducidas ya que, cuando se produce una desintegración leptónica, posee los mismos productos finales que el proceso tW . Por otro lado el proceso $t\bar{t}$ semileptónico contribuye a este fondo pues, aunque no tiene los mismos productos finales que el proceso tW , en la reconstrucción de los jets, estos pueden ser identificados erróneamente como leptones. A estos leptones mal identificados se les denomina leptones *non-prompt*. El objetivo de este trabajo es comparar la eficiencia de dos técnicas distintas en la tarea de discriminar estos leptones *non-prompt* y así reducir la contribución del proceso $t\bar{t}$ semileptónico al fondo de la medida del proceso

tW .

Una de estas técnicas se trata de una criba manual estableciendo criterios sobre la identificación de los leptones y jets, denominada *cut-based*, la otra, emplea un algoritmo de *machine learning* como es el *Lepton MVA* para ayudar con esta tarea.

4.3. Cut-Based

Como ya se ha comentado, en este trabajo se estudiará si el algoritmo *Lepton MVA* mejora la identificación de los leptones frente a la obtenida con un método *cut-based*.

Un análisis *cut based* (basado en cortes o selecciones) es un enfoque utilizado para identificar eventos de interés en datos experimentales. Se aplican una serie de criterios de selección denominados cortes, a las variables medidas en los detectores para distinguir entre señal y fondo. Los cortes se establecen de acuerdo con las características específicas de la señal buscada y los procesos de fondo que pueden imitarla. El objetivo de un análisis *cut-based* es reducir el fondo y aumentar la presencia de la señal en la muestra de datos seleccionada.

El método *cut-based* aplicado en este trabajo exige criterios de calidad sobre los leptones como son:

- $\text{Electron_lostHits} \leq 1$: Un *lost hit* se produce cuando un electrón deja de interactuar con uno o más de los detectores, lo que resulta en la pérdida de información sobre su trayectoria.
- $\text{Electron_cutBased} \geq 4$: Corte *tight*. Esta variable representa la eficiencia de identificación de un electrón basada en un conjunto de criterios de corte. El punto de corte *tight* tiene una eficiencia de alrededor del 70% para electrones genuinos y se utiliza cuando los fondos son mayoritarios [37].
- $\text{Muon_pfRelIso04_all} < 0.15$: Da cuenta de cómo de aislado se encuentra el muón, si está poco aislado probablemente provenga de un jet.
- $\text{Muon_tightId} = 1$: Corte *tight*. Esta variable representa la eficiencia de iden-

tificación de un muón basada en un conjunto de criterios de corte. La identificación de muones *tight* tiene como objetivo suprimir los muones provenientes de desintegraciones hadrónicas. Un muón *tight* es un muón *loose* con una traza medida en el *tracker* que utiliza hits de al menos seis capas del *tracker* interno, incluyendo al menos un hit de píxel. El muón debe ser reconstruido tanto como un muón del *tracker*, como un muón global. El ajuste del muón global debe tener un $\chi^2 < 10$ e incluir al menos un hit del sistema de muones. Un muón *tight* debe ser compatible con el vértice primario, teniendo un parámetro de impacto transversal $|d_{XY}| < 0.2$ cm y un parámetro de impacto longitudinal $|d_z| < 0.5$ cm [38].

Estas cuatro variables definen el método *cut-based* aplicado en este trabajo para la discriminación de leptones *non-prompt*. Se aplicarán criterios de selección adicionales sugeridos en [1] para la limpieza de las muestras.

A la hora de aplicar el *Lepton MVA* en la tarea de identificación de leptones, para electrones, se mantendrán los cortes antes mencionados y a mayores se le añadirá el corte del discriminante del *Lepton MVA*, mientras que para muones, se mantendrá el corte *tight* pero se le dejará realizar el corte del aislamiento al *Lepton MVA*.

4.4. Lepton MVA

En esta sección se desarrollará el funcionamiento del algoritmo de aprendizaje automático que será usado para minimizar el número de leptones *non-prompt* que pueden ser medidos en una muestra.

Como ya se comentó en el capítulo anterior, se establece una división entre los leptones (electrones y muones) *prompt* y los leptones *non-prompt*. Los leptones *prompt* son provenientes de la desintegración de W, Z, H o leptones τ y son normalmente identificados aplicando criterios optimizados de identificación y de aislamiento. Sin embargo, estos datos pueden estar contaminados por una fracción de leptones cuyo origen es otro y pasan esta selección, los leptones *non-prompt*. En los casos en los que la sección eficaz de la señal es menor que el proceso dominante en la colisión pp, la

contribución de los leptones provenientes de otras fuentes puede ser comparable con la señal. Por ello es importante la discriminación entre leptones *prompt* y leptones *non-prompt* [39].

Los leptones *prompt* al provenir de las partículas antes mencionadas, suelen estar aislados, aunque en algunos casos en los que partículas predecesoras a estos leptones tienen un momento transverso muy alto, los leptones *prompt* pueden perder este aislamiento y aparecen junto a ellas. Por otro lado, los leptones *non-prompt* provienen de la desintegración de un hadrón y por esta razón, aparecen rodeados de un *jet* de partículas. También, en un número muy alto de ocasiones, los leptones *non-prompt* son producidos por la desintegración de un quark *b*, cuyo *jet* podrá ser identificado usando técnicas de *b – tagging*.

El objetivo del *Lepton MVA* consiste en identificar las diferencias de las características entre los leptones *prompt* y los leptones *non-prompt* para así poder realizar la discriminación deseada.

4.4.1. Descripción del Lepton MVA

El *Lepton MVA*, que significa Análisis Multivariable de Leptones es algoritmo basado en BDTs (*Boosted Decision Trees*) usado para identificar y distinguir los leptones genuinos de los leptones *non-prompt*.

Los BDTs son algoritmos de *machine learning* utilizados para la clasificación y la regresión. Combinan la potencia de los árboles de decisión con la técnica de *boosting* para mejorar el rendimiento predictivo. Un árbol de decisión es una estructura de datos en forma de árbol que ayuda a realizar predicciones basadas en una secuencia de condiciones lógicas. Cada nodo interno del árbol representa una característica o atributo, y las ramas representan las posibles opciones o valores de esa característica. Las hojas del árbol contienen las etiquetas o las predicciones finales. El proceso de *boosting* consiste en construir una serie de árboles, donde cada nuevo árbol se ajusta a los datos de entrenamiento, poniendo un mayor énfasis en los ejemplos que fueron clasificados incorrectamente por los árboles anteriores. En cada etapa de este proceso, se asignan pesos a los ejemplos de entrenamiento, de tal forma que los

ejemplos más difíciles de clasificar reciban una mayor atención en la construcción de los árboles posteriores.

La idea detrás del uso de BDT es que la combinación de múltiples árboles de decisión que mediante el proceso de *boosting* puede llevar a un clasificador más poderoso y preciso. Cada árbol se construye de forma secuencial, y en cada paso, se ajusta a los errores cometidos por los árboles anteriores.

El MVA ha sido entrenado usando la implementación TMVA de un BDT con *boosting* de gradiente (BDTG). La señal en el entrenamiento son leptones *prompt* producidos en eventos $t\bar{t}H$ y el fondo leptones *non-prompt* producidos en eventos $t\bar{t}$.

El desempeño del discriminante para muones se muestra en Figura 4.7.

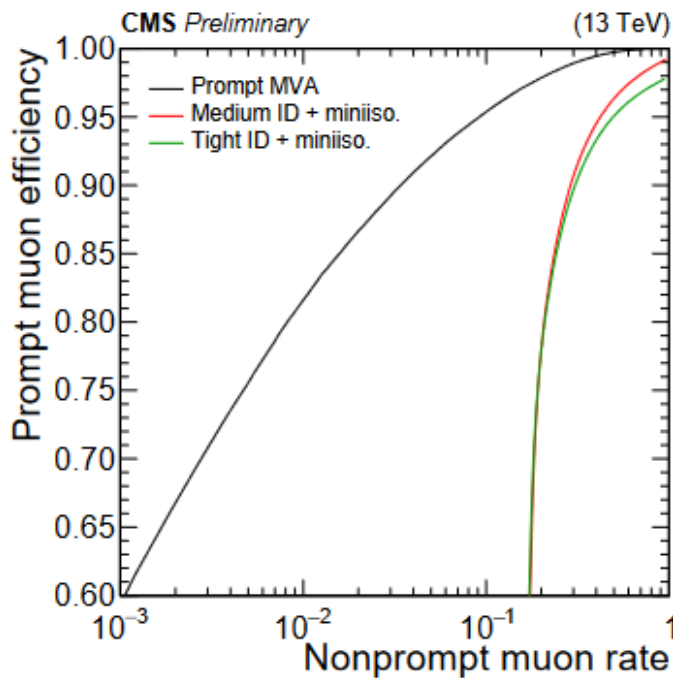


Figura 4.7: Desempeño del discriminante del MVA [39].

El lepton MVA es un algoritmo que utiliza múltiples variables de entrada relacionadas con las características de los leptones, como su energía, momento, aislamiento y otras propiedades, para construir un clasificador multivariable. Este clasificador se entrena utilizando técnicas estadísticas y algoritmos de aprendizaje automático como son los BDTs para distinguir entre leptones *prompt* y leptones *non-prompt*.

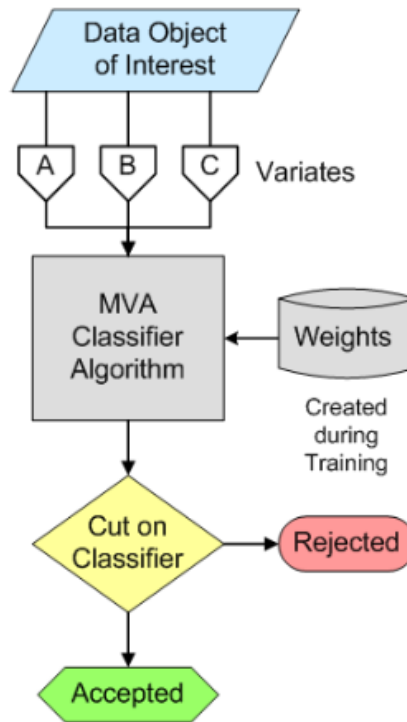


Figura 4.8: Esquema del funcionamiento del Lepton MVA. Imagen extraída de [40].

Las variables seleccionadas para el lepton MVA se eligen cuidadosamente para maximizar la capacidad de discriminación. Al combinar estas variables en un análisis multivariable, se busca obtener una mayor precisión en la identificación y clasificación de los leptones.

Algunas de las variables de entrada que emplea el *Lepton MVA* son [39]:

- p_T del leptón y $|\eta|$.
- Componente cargada de la variable de mini-aislamiento, definida como:

$$I_l^{\text{charged}} = \sum_{\text{charged}} p_T$$

- Componente neutral de la variable de mini-aislamiento, corregida por efectos de PU (*pile-up*) con el método de áreas efectivas, definida como:

$$I_l^{\text{neutrals}} = \max\left(0, \sum_{\text{neutrals}} p_T - \rho A \left(\frac{R}{0.3}\right)^2\right)$$

- Variable del ratio de p_T entre leptón y jet, p_T^{ratio} : la relación del momento transversal del leptón con el momento transversal del jet más cercano, p_T^l/p_T^{jet} . Si no hay un jet asociado al leptón presente, esta variable se establece en $\frac{1}{1+I_{\text{rel}}}$. Relacionada con esta variable está la variable de aislamiento relativo del jet, definida como $\frac{1}{p_T^{\text{ratio}}} - 1$, que está relacionada con el aislamiento del leptón.
- Variable de p_T relativo del leptón: la componente del momento del leptón en la dirección transversal al jet, $p_T^{\text{rel}} = p_l \sin \theta$, donde θ denota el ángulo entre los vectores de momento del leptón y del jet. Si no hay un jet asociado al leptón presente, esta variable se establece en cero.
- Puntuación de *b-tagging* del jet: el valor discriminante del algoritmo de *Deep Jet b-tagging* del jet correspondiente. Cuando no hay un jet de este tipo presente, la variable se establece en cero.
- Constituyentes cargados del jet: el número N_{charged} de partículas cargadas dentro del jet correspondiente. Las trazas asociadas a esas partículas deben estar dentro de un $\Delta R < 0.4$ del leptón y se requiere que provengan del vértice primario para entrar en el conteo. También se aplican criterios de calidad de traza mínima, p_T e impacto del parámetro. Si no existe dicho jet, esta variable se establece en cero.
- Parámetros de impacto: los parámetros de impacto de la traza del leptón con respecto al PV en la dirección transversal, d_{xy} , y en la dirección a lo largo de la dirección del haz d_z . El logaritmo de estas variables se utiliza como entrada para mejorar el entrenamiento del BDT.
- Significancia del parámetro de impacto: el parámetro de impacto, en tres dimensiones, de la traza del leptón con respecto al PV (punto de interacción primario), dividido por su incertidumbre, lo cual corresponde a su significancia $\frac{d}{\sigma_d}$.
- Compatibilidad del segmento del muón: la compatibilidad de los segmentos de la traza en el sistema de muones con el patrón esperado para una partícula de ionización mínima. Esta variable solo se utiliza para los muones.

Para la selección de partículas el discriminante del *Lepton MVA* viene recogido en las variables `Electron_mvTTH` y `Muon_mvTTH` que toman valores entre 0 y 1. El comportamiento de estas variables viene descrito por el de una curva como la de Figura 4.7, cuanto más próxima a la unidad sea la variable, mayor eficiencia tendrá, es decir, los leptones identificados como tal tienen una mayor probabilidad de que efectivamente lo sean, pero se asegura una mayor tasa de leptones *non-prompt*.

4.5. Criterio de selección de eventos

En [1] se mide la sección eficaz del proceso tW .

Como ya se ha comentado, el quark top decae casi el 100 % de las veces en un quark b y en un bosón W. En el estudio se analiza el modo de desintegración dileptónico con estados finales $e^\pm\mu^\mp$, el los que tanto el bosón W proveniente de la desintegración del quark top como el otro bosón W asociado a la producción del quark top, decaen leptónicamente, uno en muón y neutrino y el otro en electrón y neutrino. Esto lleva a un estado final en el que se hallan dos leptones con carga eléctrica opuesta, un jet proveniente de la fragmentación de un quark b y dos neutrinos.

Se usará la notación $njmb$ para representar eventos con exactamente n jets de los cuales m son identificados como jets b.

Es claro que la señal tW es máxima en la región $1j1b$ (Figura 4.9), pero dado que en el estudio llevado a cabo en este trabajo se analiza la contribución del proceso $Non-W/Z$, esta es mayor en la región $2j1b$ como puede apreciarse en la Figura 4.9.

Se comentan a continuación los cortes aplicados a las muestras siguiendo el trabajo realizado en [1].

Se exige a las muestras de leptones (muones y electrones) poseer $p_T > 20$ GeV y una pseudorapidez $|\eta| < 2.4$, ignorando, para los electrones, la región $1.444 < |\eta| < 1.566$ pues su reconstrucción no es óptima en esta región. Se seleccionarán eventos en los que haya al menos un leptón de cada tipo, muones y electrones. De

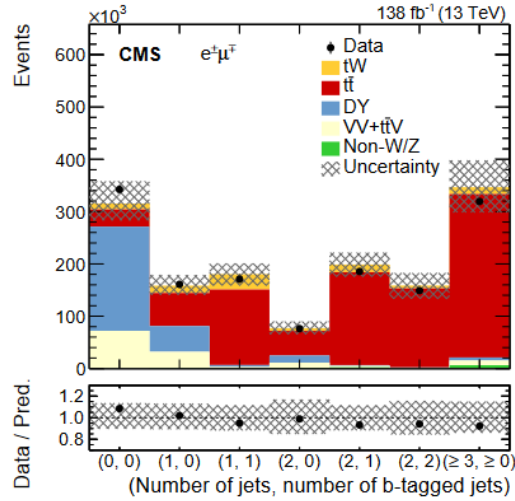


Figura 4.9: Número de eventos, datos (puntos) y predicciones de las simulaciones (histogramas) en función del número de jets y *b-tagged* jets. La barra vertical en los puntos muestra la incertidumbre estadística en los datos. La banda con tramas representa la suma de las incertidumbres estadísticas y sistemáticas. Los paneles inferiores muestran la razón de los datos frente a la suma de los rendimientos esperados. Imagen extraída de [1].

los candidatos en los eventos restantes se elegirá el muón y el electrón más energético (con mayor p_T) en cada evento. Éstas parejas de leptones han de satisfacer que su masa invariante sea superior a 20 GeV (para reducir la contribución de eventos *Drell-Yan* con baja masa invariante de dileptón) y se les exigirá que sus cargas eléctricas sean opuestas. Además, el *leading lepton*, que se define como el leptón con mayor momento transversal de cada evento, ha de tener un $p_T > 25$ GeV.

Esta selección anterior descarta eventos y dentro de aquellos que pasan la criba, se tienen solamente aquellos leptones que cumplen los requerimientos.

Paralelamente se aplican criterios de selección sobre los jets. Se exige un $p_T > 30$ GeV y $|\eta| < 2.4$ para ellos. Además, los jets han de pasar un criterio de calidad como lo exigido para los leptones ($\text{Jet_jetId} > 1$). Una vez se tiene que los jets cumplen estos requerimientos se procede a realizar una "limpieza de jets", serán seleccionados

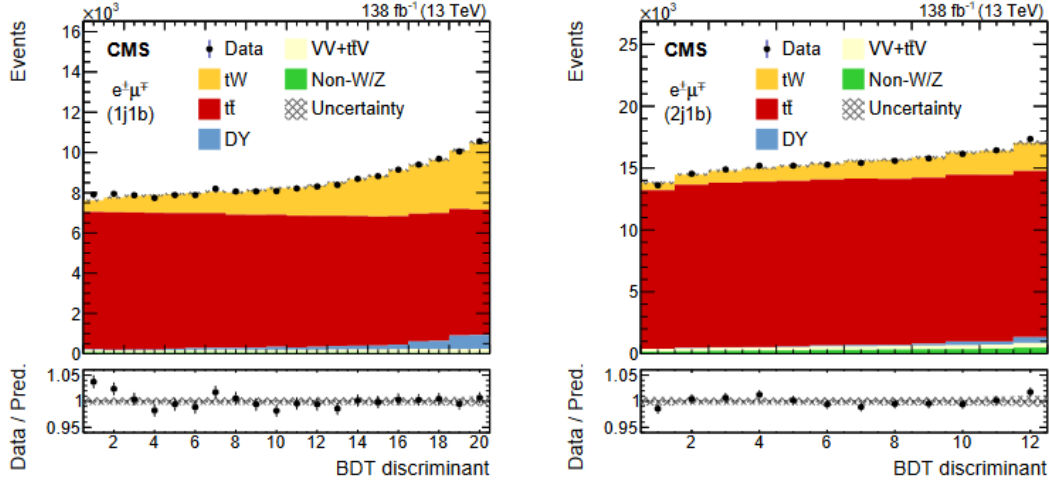


Figura 4.10: Las distribuciones de las salidas del BDT usado en [1] para eventos en las regiones 1j1b (izquierda) y 2j1b (derecha). Se muestran los datos (puntos) y las predicciones del MC (histogramas coloreados) después del ajuste. Las barras verticales en los puntos representan la incertidumbre estadística en los datos, y la banda sombreada representa la incertidumbre total en la predicción del MC. Los paneles inferiores muestran la razón de los datos frente a la suma de las predicciones del MC (puntos) después del ajuste, con las bandas indicando las correspondientes incertidumbres. Imagen extraída de [1].

aquellos jets que no se encuentren cerca de los leptones, para ello se define

$$\Delta R = \sqrt{(\eta_{Lepton} - \eta_{Jet})^2 + (\phi_{Lepton} - \phi_{Jet})^2}$$

y se exige que $\Delta R > 0.4$. Finalmente, pasarán la criba aquellos eventos con exactamente 2 jets y con uno de ellos proveniente de la desintegración de un quark b ($Jet_btagDeepFlavB > 0.2783$).

Todos los cortes anteriormente mencionados serán aplicados en la selección de eventos después de que los leptones hayan pasado el criterio de identificación de leptones *cut-based* o *Lepton MVA*.

Capítulo 5

Presentación de resultados.

Cut-based vs Lepton MVA

En esta sección se recogerán los resultados obtenidos tras el tratamiento de las muestras. El método *cut based* y *Lepton MVA* antes mencionados han sido aplicados con el objetivo de minimizar el número de leptones *non-prompt* en las muestras. Estos dos métodos serán comparados en cuanto a su eficiencia a la hora de discriminar los llamados leptones *non-prompt*.

Para la obtención de los resultados se ha empleado un código de elaboración propia escrito en *Python*. Los datos que se utilizan en un análisis en física de partículas tienen estructura de n-tupla plana y se almacenan en objetos llamados *trees*. Estos *trees* contienen diferentes *branches*, que se corresponden con diferentes variables de los sucesos. En un *tree* se almacena información de un conjunto de sucesos. Para el tratamiento de los *trees* se utilizó la librería *uproot* y para el tratamiento de las variables o *branches*, se ha usado el paquete *awkward-array*, que permite el manejo de *jagged arrays* pues la información de estas variables se recoge con esta estructura. Un *jagged array* posee una forma parecida a un *array* de *numpy* pero con un tamaño variable por cada entrada. Es decir, cada elemento del *array* tiene una dimensión diferente. Esta estructura de datos es adecuada para representar, por ejemplo, observables de objetos como muones, electrones o jets, cuyo número de elementos varía

en cada suceso.

Las muestras empleadas en este estudio se corresponden con datos simulados MC. El proceso tW , objeto de estudio de este trabajo, es simulado a Next-to-Leading Order (NLO) en QCD utilizando Powheg v2. Asimismo, el proceso $t\bar{t}$ es simulado también utilizando Powheg v2 [41, 42].

5.1. Eficiencia en la discriminación de leptones *non-prompt* del fondo

Para la realización del análisis *cut-based* se implementaron los requerimientos explicados en la sección 4.3, junto con los criterios de selección de eventos de [1] comentados en la sección 4.5. En cuanto al análisis *Lepton MVA*, se aplicaron los requerimientos mencionados en la sección 4.3 y exigiendo, en esta sección, que el parámetro discriminador para muones $\text{Muon_mvaTTH} > 0.85$ y de electrones $\text{Electron_mvaTTH} > 0.8$. En secciones posteriores estos valores van a ser modificados con el objetivo de mejorar aún más su desempeño.

Para el estudio de la discriminación de leptones *non-prompt* se empleó una muestra de fondo MC del año 2018 del proceso $t\bar{t}$ semileptónico con un tamaño de 724680 eventos. En la tabla de la Figura 5.1 que se muestra a continuación se presentan los resultados del análisis *cut-based* y *Lepton MVA*. Estos resultados han sido obtenidos aplicando ambos métodos a la selección de calidad de los leptones y tras el aplicado de los distintos cortes sugeridos [1]:

1. Selección exigiendo más de un leptón de cada tipo por evento
2. Corte sobre la masa invariante y momento transversal del *leading lepton*.
3. Selección exigiendo carga eléctrica opuesta de los leptones.
4. Selección de eventos pertenecientes a la región $2j1b$.

En las simulaciones MC no todos los eventos tienen el mismo peso estadístico, es por ello que para determinar el número de eventos que han pasado la selección hay

	<i>CUT-BASED</i>	<i>MVA</i>
Eventos tras selección 1	303711	11215
Eventos tras selección 2	297444	11034
Eventos tras selección 3	155163	6242
Eventos tras selección 4	29316	1140

Figura 5.1: Número de eventos en la muestra de fondo tras la aplicación de los criterios *cut-based* y *Lepton MVA* para cada una de las selecciones aplicadas en [1].

que reescalar los datos provenientes de las simulaciones del MC. Para realizar este reescalado, se suman los pesos de todos los eventos que han pasado la selección y se divide por el peso total de la muestra. Además como el proceso objeto de estudio es el $t\bar{t}$ semileptónico se debe multiplicar este valor por su sección eficaz $\sigma = 365.4 \text{ pb}$ y por la luminosidad del año 2018 $L = 59.74 \text{ fb}^{-1}$. De esta forma se obtiene el número de eventos esperados del proceso $t\bar{t}$ semileptónico para la luminosidad del año 2018.

Se tiene por tanto, después de la selecciones y del reescalado, que el número de eventos asociados al proceso $t\bar{t}$ semileptónico en la muestra de fondo es:

	<i>CUT-BASED</i>	<i>MVA</i>
Eventos	1343 ± 8	52 ± 2

Figura 5.2: Número de eventos finales tras la aplicación de los métodos *cut-based* y *Lepton MVA*. Las incertidumbres asociadas se corresponden con las estadísticas.

La diferencia en el número de eventos que pasan las selección *cut-based* y *Lepton MVA* es considerable. Esto da buena cuenta de cómo de útil es utilizar un algoritmo de *machine learning* a la hora de realizar la identificación de los leptones.

Se presentan los siguientes histogramas, en la Figura 5.3, con el objetivo de mostrar de una manera más visual esta reducción del fondo ya comentada.

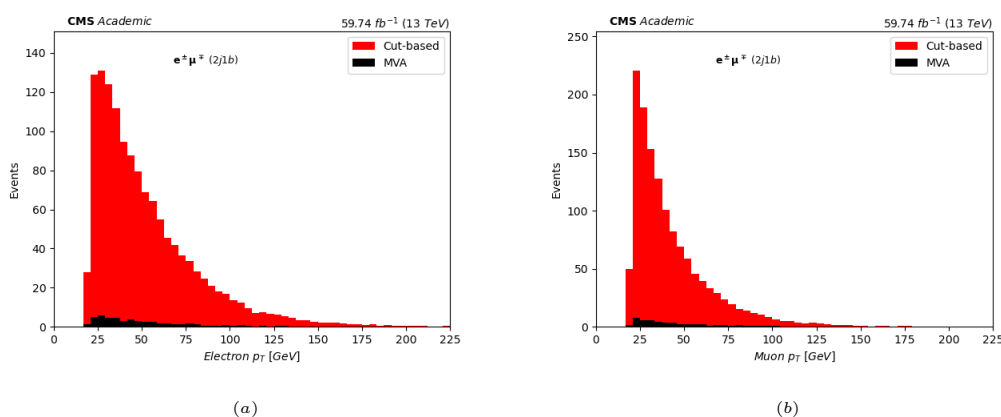


Figura 5.3: Momento transversal p_T de electrones (a) y muones (b) de la contribución del proceso $t\bar{t}$ semi-leptónico en la muestra de fondo.

A la vista de los resultados es claro que *Lepton MVA* presenta un criterio más estricto a la hora de discriminar leptones *non-prompt*. Más adelante se comprobará como afecta esta discriminación tan robusta a la muestra de señal tW .

5.2. Efectos en la señal

Se ha visto, en la sección anterior, la gran eficiencia del *Lepton MVA* a la hora de identificar y eliminar leptones que han sido erróneamente identificados como tal. Para que este sea un resultado válido, ha de preservar el máximo número de leptones en una muestra de señal del proceso tW , pues de lo contrario, si suprime demasiados eventos de una muestra de señal del proceso tW resulta no ser una buena solución para este problema a pesar de que haga una selección tan buena de los eventos en el fondo. De esta manera, lo esperado al aplicar este método sobre una muestra de señal es una reducción mínima de los eventos al pasar la selección. En los resultados mostrados a continuación en Figura 5.4, los valores de los discriminantes para los leptones fueron: $\text{Muon_mvaTTH} > 0.85$ y $\text{Electron_mvaTTH} > 0.8$. Con el objetivo de estudiar como afecta el *Lepton MVA* a la muestra de señal, se seleccionó una muestra MC del proceso tW que cuenta con 567042 eventos. Se procedió de forma análoga al análisis para la muestra de fondo.

Aplicando el mismo razonamiento a la hora de hacer el repesado, se presenta en

	<i>CUT-BASED</i>	<i>MVA</i>
Eventos tras máscara 1	166607	103517
Eventos tras máscara 2	165625	103062
Eventos tras máscara 3	163058	102816
Eventos tras máscara 4	26152	16726

Figura 5.4: Número de eventos en la muestra de señal del proceso tW tras el aplicado de los criterios *cut-based* y *Lepton MVA* para cada una de las selecciones aplicadas en [1].

Figura 5.5, el número de eventos finales obtenidos para la muestra de señal.

	<i>CUT-BASED</i>	<i>MVA</i>
Eventos	7130 ± 40	4560 ± 40

Figura 5.5: Número de eventos finales tras reescalado en muestra de señal tW después de la aplicación de los métodos *cut-based* y *Lepton MVA*. Las incertidumbres asociadas se corresponden con las estadísticas.

Se presentan en Figura 5.6 los histogramas del momento transversal para muones y electrones que han pasado la criba de la muestra de señal. Comparando estas

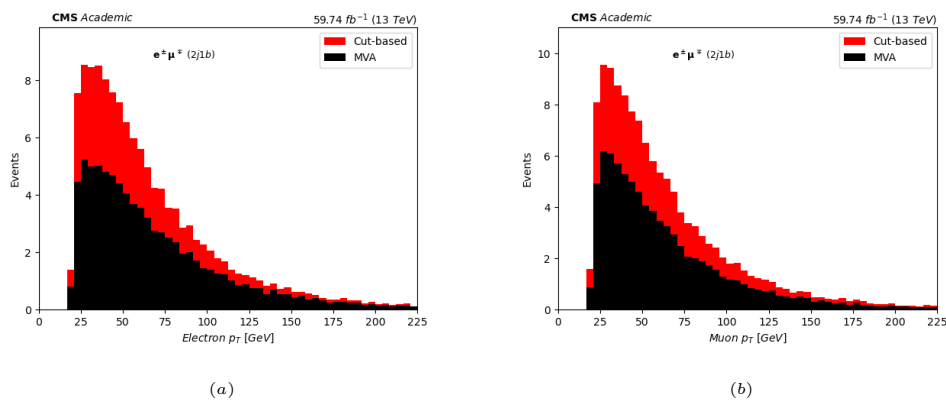


Figura 5.6: Momento transversal p_T de electrones (a) y muones (b) en la muestra de señal tW .

gráficas con las de la Figura 5.6, se observa como el poder discriminante del *Lepton MVA* frente al método *cut-based* es mucho mayor para los leptones de la muestra de fondo. Esto da cuenta de que en la muestra de fondo existía un mayor número de leptones erróneamente identificados frente a los leptones genuinos de la muestra de señal.

5.3. Comparación de eficiencia en señal y fondo

Dados los valores del parámetro discriminador para muones $\text{Muon_mvaTTH} > 0.85$ y de electrones $\text{Electron_mvaTTH} > 0.8$, se obtuvieron los resultados sobre el número de eventos de la contribución del proceso $t\bar{t}$ semileptónico al fondo del proceso tW , que se muestran en la Figura 5.2, y sobre el número de eventos del proceso tW en la muestra de señal para los métodos *cut-based*, sugerido en [1], y *Lepton MVA*.

Se tiene entonces que el método *Lepton MVA* reduce en un 96 % las contribuciones de los leptones *non-prompt* provenientes de la reconstrucción de eventos $t\bar{t}$ semileptónico al fondo del proceso tW pero esto conlleva una reducción también en la muestra de señal del 34 %.

Es claro que el objetivo de este trabajo se ha cumplido, pues la reducción de *non-prompt* es claramente notable, pero en la siguiente sección se muestra una configuración de los parámetros más eficiente en cuanto a que mantiene una discriminación de *non-prompt* muy elevada, pero afecta levemente a la muestra de señal.

Se presentan en Figura 5.7 y 5.8, los principales observables para los electrones y muones que han pasado las selecciones descritas en el capítulo anterior. En ellas se puede observar las actuaciones de los dos métodos presentados en este trabajo, como son *cut-based*, que es el sugerido en [1] y *Lepton MVA*, cuyo objetivo es reducir las contribuciones de leptones *non-prompt*.

Se observa a simple vista que el *Lepton MVA* ofrece una mejora sustancial a la hora de discriminar estos leptones frente al método *cut-based*, pero para esta configuración conlleva una cierta pérdida de señal.

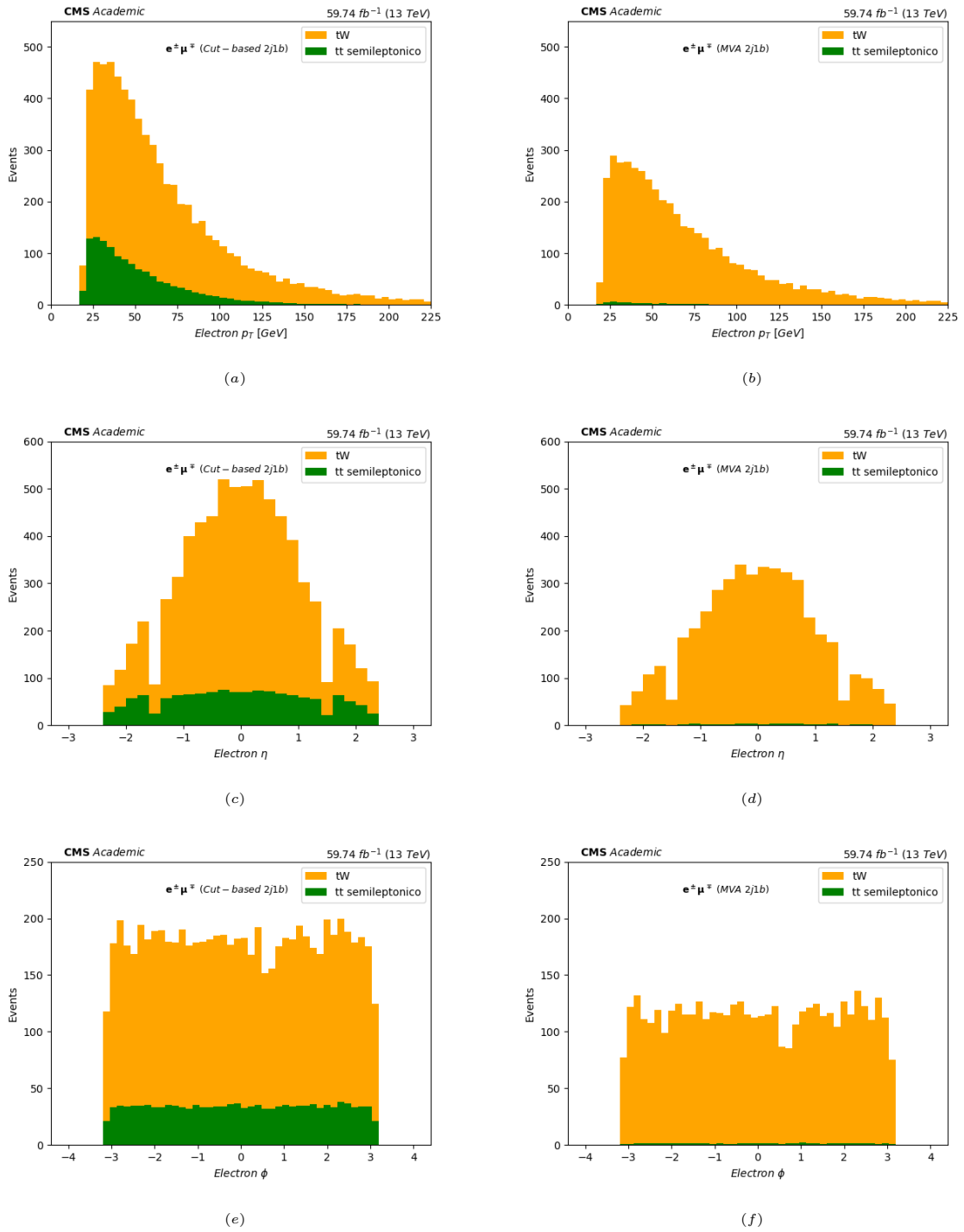


Figura 5.7: Momento transverso p_T ((a), (b)), pseudorrapidez η , ((c), (d)) y ángulo azimutal ϕ , ((e), (f)), de electrones tras aplicar método *cut-based*, en columna de la izquierda, y tras aplicar método *Lepton MVA*, en columna de la derecha.

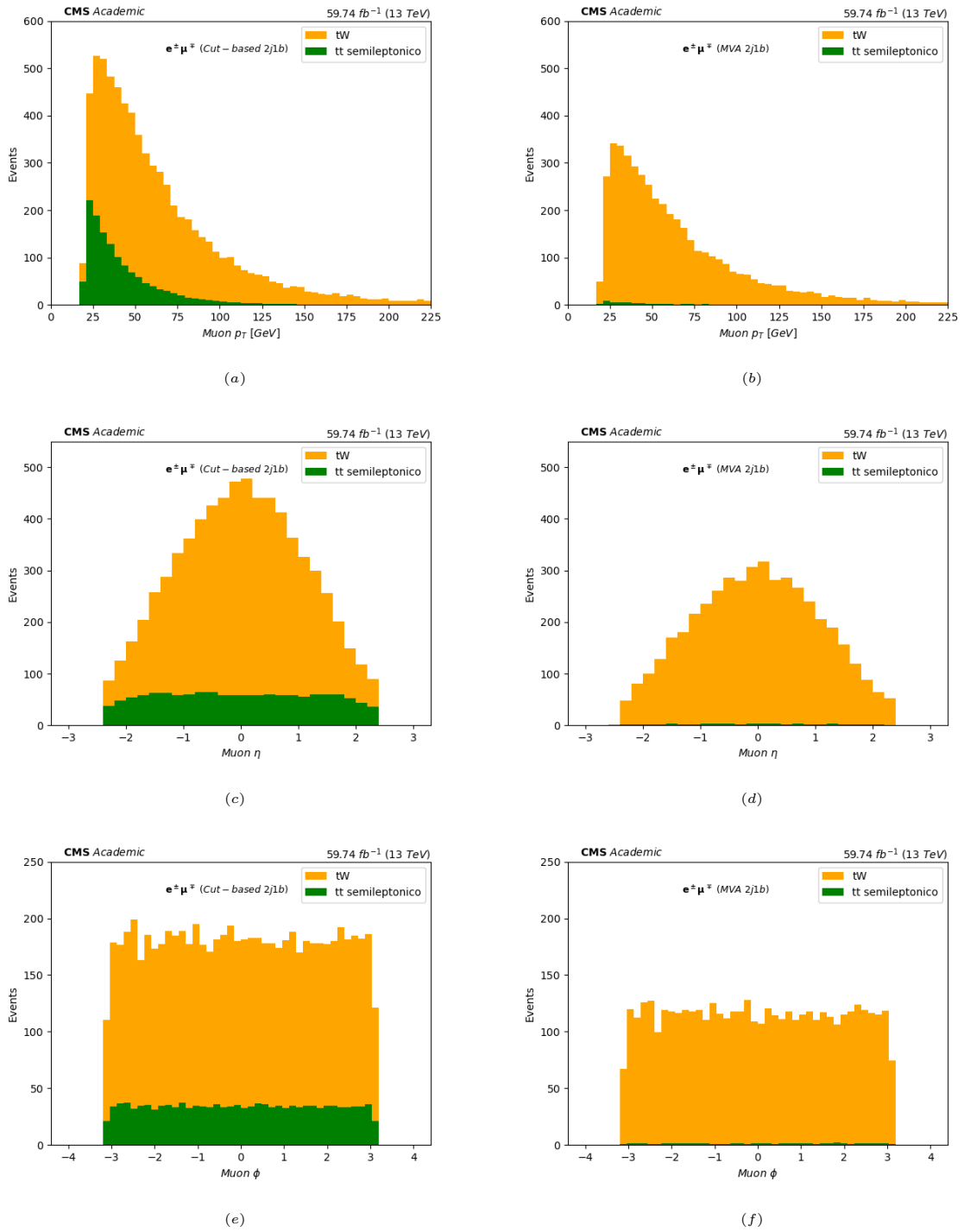


Figura 5.8: Momento transverso p_T ((a), (b)), pseudorrapidez η , ((c), (d)) y ángulo azimutal ϕ , ((e), (f)), de electrones tras aplicar método *cut-based*, en columna de la izquierda, y tras aplicar método *Lepton MVA*, en columna de la derecha.

5.4. Optimización de los resultados

En esta sección se muestran los resultados experimentales tras reducir los requerimientos impuestos sobre los parámetros para la identificación de muones y electrones con el objetivo de mejorar la eficiencia del proceso presentado en secciones anteriores. El objetivo es buscar unos valores para las variables `Muon_mvaTTH` y `Electron_mvaTTH` que mantengan una gran reducción del número de leptones *non-prompt* en el fondo y no afecten en gran medida a la muestra de señal.

Para esta sección se ha establecido $\text{Muon_mvaTTH} > 0.7$ y $\text{Electron_mvaTTH} > 0.5$. Para la muestra de fondo y tras el reescalado el número de eventos pertenecientes al proceso $t\bar{t}$ semileptónico y que contribuyen al fondo del proceso tW se muestran en Figura 5.9.

	<i>CUT-BASED</i>	<i>MVA</i>
Eventos	1343 ± 8	118 ± 2

Figura 5.9: Número de eventos finales pertenecientes al canal $t\bar{t}$ semileptónico tras la aplicación de los métodos *cut-based* y *Lepton MVA* para la muestra de fondo. Las incertidumbres asociadas se corresponden con las estadísticas.

Para la muestra de señal del proceso tW , una vez aplicado el reescalado han pasado las selecciones el número de eventos que se muestra en Figura 5.10.

	<i>CUT-BASED</i>	<i>MVA</i>
Eventos	7130 ± 40	6370 ± 40

Figura 5.10: Número de eventos de la muestra de tW tras la aplicación de los métodos *cut-based* y *Lepton MVA*. Las incertidumbres asociadas se corresponden con las estadísticas.

Se tiene por tanto, que el método *Lepton MVA* reduce en un 91 % las contribuciones de los leptones *non-prompt* provenientes de la reconstrucción de eventos

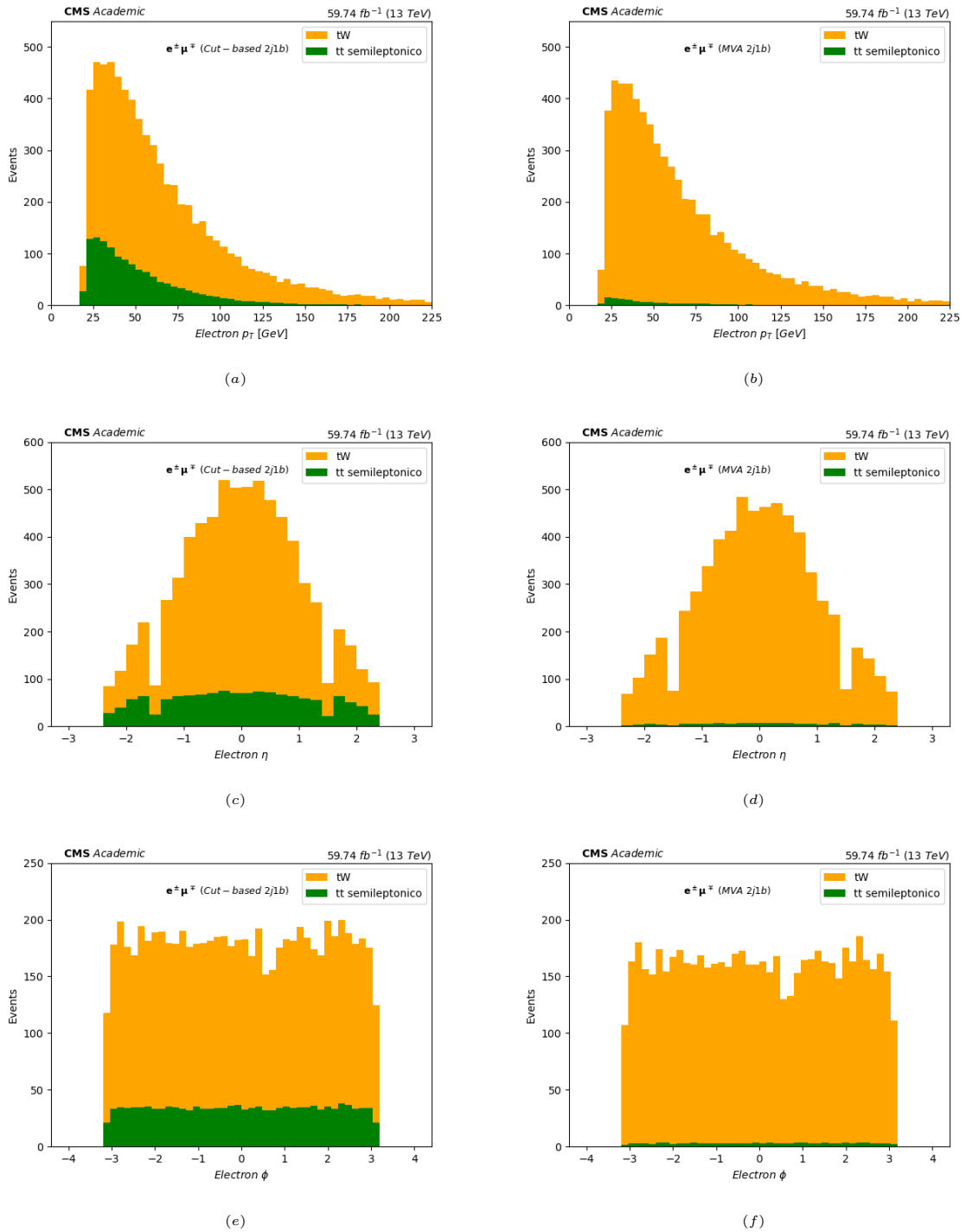


Figura 5.11: Momento transverso p_T ((a), (b)), pseudorrapidez η , ((c), (d)) y ángulo azimutal ϕ , ((e), (f)), de muones tras aplicar método *cut-based*, en columna de la izquierda, y tras aplicar método *Lepton MVA*, en columna de la derecha, con una configuración de los parámetros que mejora la fuerza de la señal para *Lepton MVA*.

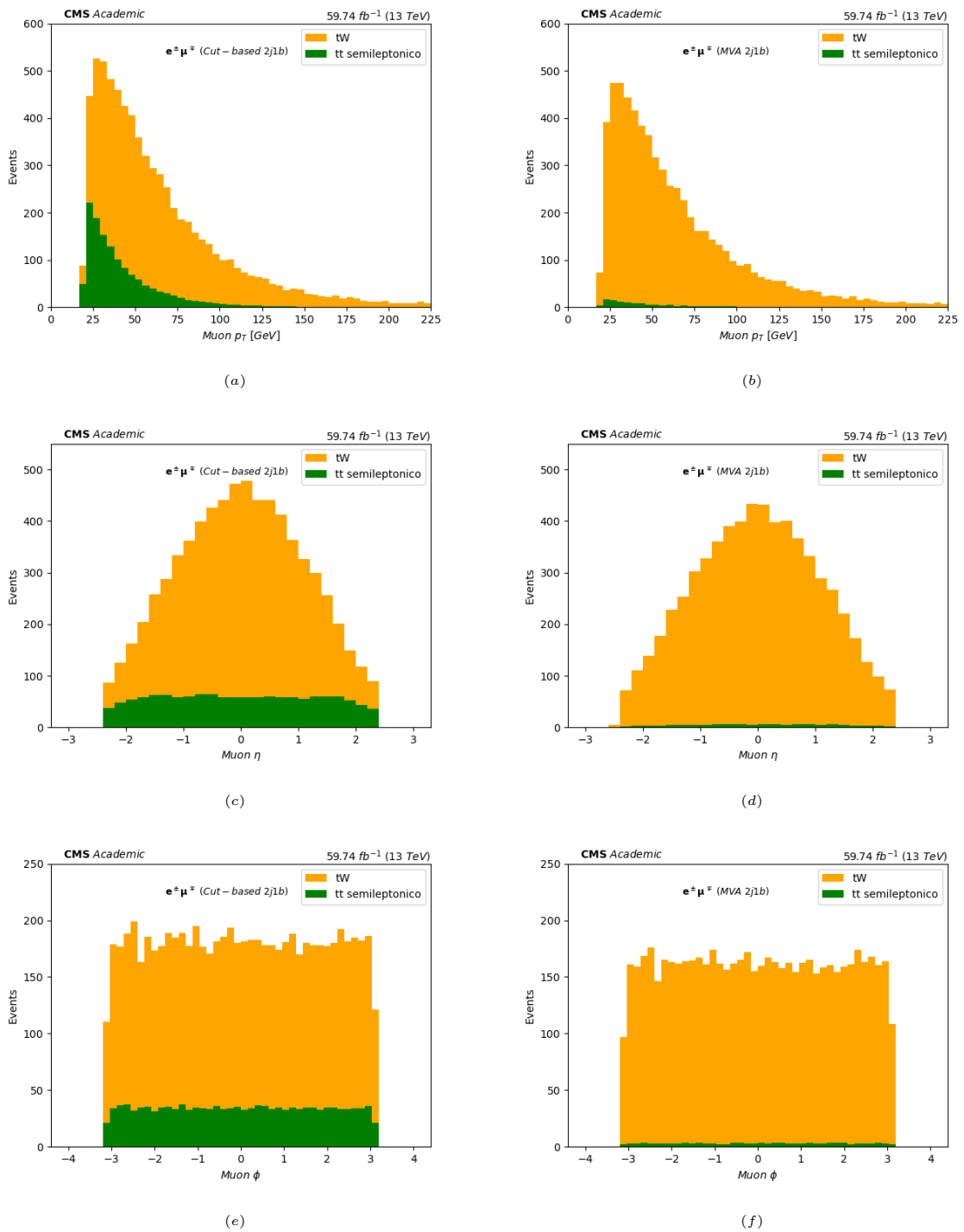


Figura 5.12: Momento transverso p_T ((a), (b)), pseudorrapidez η , ((c), (d)) y ángulo azimutal ϕ , ((e), (f)), de muones tras aplicar método *cut-based*, en columna de la izquierda, y tras aplicar método *Lepton MVA*, en columna de la derecha, con una configuración de los parámetros que mejora la fuerza de la señal para *Lepton MVA*.

$t\bar{t}$ semileptónico al fondo del proceso tW . Pero a diferencia de los resultados previamente mostrados, ahora la señal solamente se ha visto reducida en un 11%. Se presentan en Figuras 5.11 y 5.12, los principales observables de electrones y muones después de haber sido procesada la muestra. Observando los histogramas se ve claramente como *Lepton MVA* frente a *cut-based* realiza una mejor criba en los leptones de la muestra del fondo sin apenas afectar a la muestra de señal.

5.5. Consistencia de los resultados

Cabe destacar que los resultados obtenidos en este trabajo son robustos pues concuerdan con los presentes en el *paper* de referencia [1].

La aplicación del método *cut-based* en el *paper* devuelve unos 3600 eventos correspondientes al proceso $t\bar{t}$ semileptónico que pasan la selección. Hay que tener en cuenta la simulación de las muestras en [1] toma cuenta de los años 2016, 2017 y 2018, es decir, se tiene una luminosidad total de 138 fb^{-1} . En este trabajo se han utilizado muestras MC provenientes de simulaciones del año 2018, con una luminosidad de 59.74 fb^{-1} . Por tanto el número de sucesos obtenidos en [1] correspondientes a este año son

$$3600 \times \frac{59.74}{138} \approx 1500$$

frente a los 1300 obtenidos en el análisis de este trabajo y teniendo en cuenta que los 3600 eventos son después de hacer el ajuste a los datos experimentales (antes del ajuste habría un número menor de eventos).

Procediendo de forma análoga para el proceso tW , se tiene que en [1], el número de eventos correspondientes a este proceso es 7200 frente a los 7100 que ha sido obtenidos en este estudio.

Estos resultados dan cuenta de que el análisis aquí realizado es robusto y coherente con los resultados que han sido obtenidos en el estudio de referencia [1].

Capítulo 6

Conclusiones

En este estudio, se ha investigado el uso de un algoritmo de aprendizaje automático como es el *Lepton MVA* para reducir la presencia de leptones *non-prompt* del proceso $t\bar{t}$ semileptónico. Estos leptones, provenientes de fuentes distintas a la desintegración de partículas inmediatamente producidas, representan un desafío importante cuando se busca realizar una medida precisa de la sección eficaz del proceso tW , como la que se realiza en [1].

Se ha comprobado que *Lepton MVA* es capaz de reducir las contribuciones al fondo del proceso tW que producen estos leptones mal identificados en hasta un 91 % y con una asumible pérdida de señal del 11 %. Los resultados obtenidos demuestran que la utilización del *Lepton MVA* ha sido efectiva en la reducción de los leptones *non-prompt*, lo cual contribuirá a mejorar la precisión de la medida de la sección eficaz del proceso tW . Esta reducción de la contaminación de fondos no deseados es fundamental para obtener resultados más confiables y reducir las incertidumbres sistemáticas asociadas al análisis.

Tras el estudio realizado, se ha visto como la aplicación de algoritmos de aprendizaje automático proporciona una herramienta prometedora para abordar desafíos específicos y mejorar la precisión en la identificación y clasificación de partículas.

Por último, cabe destacar la coherencia del estudio mostrado en este trabajo con el análisis llevado a cabo en [1], pues los resultados aplicando las técnicas descritas

en [1], resultan ser del mismo orden que las que se han obtenido aquí. Los resultados obtenidos respaldan la validez y utilidad del enfoque de aprendizaje automático, específicamente el algoritmo *Lepton MVA*, en la reducción de los leptones *non-prompt* en el proceso $t\bar{t}$ semileptónico. Este enfoque ha demostrado ser una opción válida para mejora de la precisión de la medida de la sección eficaz del proceso tW , contribuyendo así al avance de la investigación en física de partículas.

Bibliografía

- [1] CMS Collaboration. Measurement of inclusive and differential cross sections for single top quark production in association with a W boson in proton-proton collisions at $\sqrt{s} = 13$ TeV. 2022.
- [2] La materia oscura. https://starchild.gsfc.nasa.gov/docs/StarChild_Spanish/docs/StarChild/universe_level2/darkmatter.html.
- [3] Qué es el LHC, qué es el CERN. https://www.sne.es/wp-content/uploads/2020/10/NT_01-20_CERN.pdf.
- [4] Producción de protones en el LHC. https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.linac4.
- [5] Luminosidad integrada. https://lhc-closer.es/taking_a_closer_look_at_lhc/0.luminosity.
- [6] Efecto del pile-up. https://www.researchgate.net/figure/A-visual-example-of-pile-up-in-the-ATLAS-tracker-a-Run-1-Z-event-collected-at-an_fig1_284913474.
- [7] Oliver Sim Brüning, Paul Collier, P Lebrun, Stephen Myers, Ranko Ostojic, John Poole, and Paul Proudlock. *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004.
- [8] Cavidades de radiofrecuencia en el LHC. <https://home.cern/science/engineering/accelerating-radiofrequency-cavities>.

-
- [9] Imanes superconductores. <https://home.cern/science/engineering/pulling-together-superconducting-electromagnets>.
- [10] Diagram of an LHC dipole magnet. <https://cds.cern.ch/record/40524>.
- [11] J. Maestre Heredia. Design and behaviour of the Large Hadron Collider external beam dumps capable of receiving 555 MJ/dump. 2021.
- [12] CMS Collaboration. The CMS experiment at the CERN LHC. <https://dx.doi.org/10.1088/1748-0221/3/08/S08004>, aug 2008.
- [13] Propiedades CMS. <https://cms.cern/detector>.
- [14] Esquema detector CMS. <https://cms.cern/detector>.
- [15] Sistema de coordenadas de CMS. https://tikz.net/axis3d_cms/.
- [16] Propiedades detector de trazas de CMS. <https://cms.cern/detector/identifying-tracks>.
- [17] Detector de trazas de CMS. <https://home.cern/news/news/experiments/cool-running-cms-tracker>.
- [18] Píxeles de silicio del tracker. <https://cms.cern/detector/identifying-tracks/silicon-pixels>.
- [19] Tiras de silicio del tracker. <https://cms.cern/detector/identifying-tracks/silicon-strips>.
- [20] Calorímetros. https://physicsmasterclasses.org/exercises/hands-on-cern/hoc_v21sp/main_frame/de_cal1.html.
- [21] Calorímetro Electromagnético de CMS. <https://cms.cern/detector/measuring-energy/energy-electrons-and-photons-ecal>.
- [22] Cristina Biino. The CMS Electromagnetic Calorimeter: overview, lessons learned during Run 1 and future projections. *Journal of Physics: Conference Series*, 587(1):012001, feb 2015.

- [23] Calorímetro Hadrónico de CMS. <https://cms.cern/node/1202/edi%3Fdestination%3D/node/1202>.
- [24] Solenoide superconductor en CMS. <https://cms.cern/detector/bending-particles>.
- [25] L. Rossi et al. The Compact Muon Solenoid: The CMS magnet project. 2003.
- [26] Solenoide superconductor de CMS. https://irfu.cea.fr/en/Phoce/Vie_des_labos/Ast/ast_visu.php?id_ast=3011.
- [27] Cámaras de detección de muones. https://www.researchgate.net/figure/Layout-of-different-types-of-muon-chambers-and-their-location-in-one-quarter-of-the-CMS_fig13_48410310.
- [28] S. Chatrchyan et al. The Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider. *Journal of Instrumentation*, 2008.
- [29] Triggering. <https://cms.cern/detector/triggering-and-data-acquisition>.
- [30] Christian Bierlich, Smita Chakraborty, Nishita Desai, Leif Gellersen, Ilkka Helenius, Philip Ilten, Leif Lönnblad, Stephen Mrenna, Stefan Prestel, Christian T. Preuss, Torbjörn Sjöstrand, Peter Skands, Marius Uthm, and Rob Verheyen. A comprehensive guide to the physics and usage of PYTHIA 8.3. <https://scipost.org/10.21468/SciPostPhysCodeb.8>, 2022.
- [31] A.M. Sirunyan, A. Tumasyan ski, and N. Woods. Particle-flow reconstruction and global event description with the CMS detector. *Journal of Instrumentation*, 12(10):P10003, oct 2017.
- [32] Eric Ballabene. *Measurement of the tt production cross section at 13 TeV in the all-jets boosted regime with CMS at LHC*. PhD thesis, 09 2017.
- [33] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-kt jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063, apr 2008.
- [34] R.L. Workman et al. (Particle Data Group). CKM Quark-Mixing Matrix. <https://pdg.lbl.gov/2020/reviews/rpp2020-rev-ckm-matrix.pdf>, 2022.

- [35] Conrado Muñoz Díaz. Top quark Physics: Top anti-top pair production with additional jets. 2022.
- [36] F. Maltoni (Univ. Catholique de Louvain) T.M. Liss (The City College of New York) and A. Quadt (Univ. Göttingen). THE TOP QUARK. <https://pdg.lbl.gov/2016/reviews/rpp2016-rev-top-quark.pdf>, 2015.
- [37] The CMS collaboration and A.M. Sirunyan et al. Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. *Journal of Instrumentation*, 16(05):P05014, may 2021.
- [38] A.M. Sirunyan and A. Tumasyan et al. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $s=13$ TeV. *Journal of Instrumentation*, 13(06):P06015, jun 2018.
- [39] Santiago Folgueras Clara Ramon Alvarez Andrea Trapote Fernandez Nicolo Trevisani Sergio Sanchez Cruz, Barbara Alvarez Gonzalez and Authors of HIG-19-0083. Multivariate techniques for prompt lepton identification. 2022.
- [40] K. Gobel J. Markert T. Galatyuk J. Stroth1 A. Rustamov M. Lorenz1, R. Holzmann and the HADES collaboration. Lepton identification using a multivariate classification method. 2009.
- [41] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *Journal of High Energy Physics*, 2007, nov 2007.
- [42] Carlo Oleari Simone Alioli, Paolo Nason. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *Journal of High Energy Physics*, nov 2010.