



UNIVERSIDAD DE SONORA
DIVISIÓN DE CIENCIAS EXACTAS Y NATURALES
DEPARTAMENTO DE FÍSICA

IDENTIFICACIÓN DE JETS USANDO MÉTODOS DE INTELIGENCIA ARTIFICIAL EN EL
CONTEXTO DEL EXPERIMENTO CMS DEL CERN

ALUMNO:
OLGA MARÍA FIMBRES MORALES

PROFESOR ASESOR:
DR. ALFREDO CASTAÑEDA HERNANDEZ

18 DE OCTUBRE 2020

Universidad de Sonora

Repositorio Institucional UNISON



**"El saber de mis hijos
hará mi grandeza"**



Excepto si se señala otra cosa, la licencia del ítem se describe como openAccess

Índice

1. Modelo Estándar.	4
1.1. Historia del modelo estándar	4
1.2. Tipo de fuerzas.	8
2. Origen de los Jets en la física de partículas	9
2.1. Principio de confinamiento.	9
2.2. Proceso de Hadronización	9
3. Observaciones experimentales de Jets	10
4. Gran Colisionador de Hadrones y experimento CMS	12
4.1. Detección de Jets en CMS	13
4.2. Procesos de estudio	14
5. Simulación	15
6. Análisis	20
7. Modelos de inteligencia artificial.	23
7.1. Redes neuronales para la clasificación de imágenes (CNN)	24
7.2. CNN para identificación de Jets	26
7.3. Parámetros configurables de la red CNN	28
8. Resultados	29
9. Conclusiones	30
A. Método de actualización de valores (backpropagation)	31
B. Gráficas de aprendizaje para las variaciones de la configuración de la red CNN	33
C. Algoritmo FasJet para la reconstrucción de Jets	35
D. Nodos y conexiones de la red neuronal en el modelo de Tensorflow	36
E. Definición de momento Transversal y Pseudo-rapidez	37

Introducción.

El presente trabajo de tesis tiene como principal objetivo el desarrollo de un modelo en base a redes neuronales para discriminar eventos correspondientes a dos procesos en la física de partículas. La relevancia de este trabajo es el mostrar la importancia de la implementación de tecnologías emergentes (como los algoritmos de inteligencia artificial) en la optimización de tareas propias de aquellos experimentos que producen una gran cantidad de datos, como lo son los experimentos ligados al Gran Colisionador de Hadrones del CERN. Lo anterior mediante la implementación de modelos de inteligencia artificial, en este caso específico utilizando una red neuronal propia para tareas relacionadas con la clasificación de imágenes (CNN) y ajustando su uso para reconocimiento de cierto tipo de partículas elementales, gran parte del trabajo se enfoca en la identificación de la mejor combinación de parámetros para obtener un aprendizaje óptimo de la red neuronal y con ello resultados más confiables. Otro objetivo a cumplir en este proceso fue el de adiestrarse en el uso de lenguajes de programación de alto nivel como python y entornos para el desarrollo de algoritmos de inteligencia artificial (machine learning) como es el caso de tensorflow, ambos permitirán en un futuro desarrollar proyectos mas avanzados.

La distribución del contenido se puede resumir de la siguiente manera: En el Capítulo 1 se encuentra una recapitulación de los descubrimientos más importantes en la física de partículas que llevaron a la construcción de lo que hoy se conoce como el modelo estándar, así como una descripción concisa de los tipos de fuerzas fundamentales que actúan sobre las partículas elementales.

En el Capítulo 2 se explica de manera concisa el principio de confinamiento de los quarks el cual lleva a la formación de partículas compuestas llamadas jets parte fundamental del desarrollo de este estudio, para después en el capítulo 3 hablar sobre las primeras observaciones experimentales de jets que llevaron al descubrimiento del gluon en 1979 con la observación de un evento compuesto con 3 de estas partículas.

El capítulo 4 presenta una descripción del experimento CMS del CERN ,localizado en el Gran Colisionador de Hadrones (LHC) y la manera en que se realiza de forma experimental la detección de jets, además se presentan dos procesos en física de partículas que contienen jets y que son usados durante el desarrollo de esta tesis, dichos eventos son W' y QCD , el primero correspondiente a una partícula hipotética y el segundo a un proceso ya medido y reportado en la física de partículas.

Los detalles de la simulación de producción de dichos procesos puede encontrarse en el capítulo 5, donde se habla de las paqueterías utilizadas y su secuencia en la generación de muestras.

En el capítulo 6 se muestran los análisis realizados usando las muestras generadas por simulación para los cuales únicamente se utilizaron los eventos con un momento transversal igual o mayor a 50 GeV, encontrarse dentro del rango de pseudo-rapidez de $[-4,4]$ y de ángulo acimutal de $[-\pi, \pi]$. Ambos procesos se visualizaron mediante la paquetería ROOT donde se obtuvo histogramas de momentos transversal, pseudo-rapidez, ángulo azimutal y masa invariante; de igual modo pueden observarse gráficos 2D y 3D de los eventos dentro del detector.

el capítulo 7 presenta una explicación de las redes neuronales utilizadas en la clasificación de las imágenes de procesos de jets; iniciando con la constitución de las CNN explicando de manera concisa las capas que las componen y la manera en que puede reconfigurarse para la identificación de jets.

Finalmente en el capítulo 8 se muestran los resultados obtenidos para el aprendizaje de la red neuronal en la identificación y clasificación de las imágenes de eventos de jets.

Se agregan apéndices con información complementaria. El apéndice A describe el método de propagación hacia atrás por el que la red neuronal actualiza sus parámetros. El apéndice B contiene las gráficas de aprendizaje para las variaciones de la red neuronal que se mencionan en el capítulo 7. El apéndice C contiene el algoritmo utilizado para la reconstrucción de jets. El apéndice D es una visualización de los nodos y conexiones de una red neuronal en TensorFlow. Terminando con el apéndice E que define el momento transversal y la pseudo-rapidez.

El contenido de esta tesis así como sus resultados fue obtenido con software de acceso libre respetando los derechos de autor e incluyendo las respectivas referencias en caso de haber usado un software de terceros.

1. Modelo Estándar.

A finales del siglo XIX dentro de la comunidad científica comenzaron a observarse una serie de resultados experimentales que finalmente condujeron a la búsqueda de la composición y comportamiento de la materia en su estado más elemental. Con el trabajo de hombres y mujeres de todas partes del mundo fue posible la recopilación de una colección de teorías, experimentos y resultados los cuales logran describir las interacciones de las partículas elementales conocidas, incluso permiten la búsqueda adecuada de nuevas partículas hasta el momento teóricas; tal colección de teorías es lo que se conoce actualmente como el modelo estándar.

Con el desarrollo tecnológico que permitió el aumento en la energía a la que trabajan los aceleradores de partículas, a partir de 1978 fue que el modelo estándar obtuvo reconocimiento como tal al lograr confirmar la existencia de un mayor número de partículas elementales. A continuación, a modo de contexto histórico, se presenta una introducción a los eventos más relevantes para la constitución de este modelo en la física de partículas. Iniciando desde el descubrimiento de la primera partícula elemental para finalizar con el estudio de los jets de hadrones, tema en el que se centra el presente trabajo.

1.1. Historia del modelo estándar

La historia del modelo estándar se puede resumir en varios acontecimientos de gran relevancia, los cuales se resumen a continuación y se describen con más detalle en [1]. Desde la década de 1870 comenzó el estudio de los rayos catódicos para los cuales se llegaron a considerar diversas teorías mas no fue si no hasta 1897 que J.J. Thomson llegó a la conclusión de que estos rayos eran partículas eléctricamente cargadas encontrando su velocidad y su relación carga-radio. En 1919 Milikan y su equipo lograron definir la magnitud de la carga eléctrica del electrón. Con este nuevo conocimiento sobre la composición del átomo, en 1914 Niels Bohr propone su modelo planetario con los electrones giran en órbitas definidas alrededor del núcleo; este último considerado como sólido hasta el descubrimiento del protón por el experimento de Rutherford en 1918. En 1932 Chadwick descubre el último componente del núcleo atómico, el neutrón. Finalizando con esto el periodo clásico de la física de partículas, mientras que de manera simultanea comenzaba a comprenderse el comportamiento elemental de las partículas eléctricamente cargadas.

Iniciado el siglo XX Planck sugiere una cuantización de la energía como solución al problema del cuerpo negro, para 1905 se completa la teoría por Einstein al proponer que la naturaleza de la energía cuantizada se debía al campo electromagnético mismo y no al mecanismo de emisión. En 1923 A.H. Compton, a partir de resultados experimentales, comprueba la dualidad del comportamiento de la luz a escalas macro y microscópicas; a la partícula de luz le denominó fotón. Años después, en 1927, Dirac propone su ecuación la cual describe energía de los electrones libres, dando un resultado dual donde por cada solución con energía negativa existe una con energía positiva. Prediciendo así la existencia de una partícula con la misma masa que el electrón pero magnitud de carga eléctrica

positiva. Tal partícula logró observarse hasta 1931 por Carl Anderson y recibió el nombre de positrón. Hacia 1940 Stueckelberg y Feynman realizaron una reinterpretación de la solución que dio Dirac a su ecuación, resultando en un principio de la teoría cuántica de campos donde por cada tipo de partícula debe existir su correspondiente antipartícula con la misma masa pero una carga eléctrica opuesta. En 1955 Berkeley Bevatron observa el antipositrón y un año más tarde el antineutrón, confirmando de este modo dicha teoría.

Hacia 1930, en el estudio del decaimiento beta podía observarse una discrepancia en la energía inicial y final del proceso; para tal problema Wolfgang Pauli propone que la energía faltante correspondía a otra partícula emitida durante el proceso. Además, propone que esta debía ser eléctricamente neutra para mantener la conservación de la carga y de este modo se explicaría la falta de una traza que delatará antes su presencia. Fermi presenta su nueva teoría sobre el decaimiento beta en 1933, esta vez incorporando la partícula propuesta por Pauli. Con el cálculo de la energía correspondiente al electrón podía seguirse que la partícula es cuestión debía ser extremadamente liviana, con una masa casi cero. Fermi le otorgó el nombre de neutrino.

En el año 1934 Hideki Yukawa comienza con el desarrollo de su teoría el campo que permite la unión de protones y neutrones dentro del núcleo atómico, el corto alcance en el que trabajaba sugería un mediador pesado, con una masa entre el protón y el electrón. Hacia 1937 dos equipos se encontraban trabajando en la búsqueda de tal mediador; ambos equipos lograron identificar la partícula en los rayos cósmicos que recibe la Tierra constantemente. Sin embargo, los resultados obtenidos discrepaban de la teoría de Yukawa en la vida media y la masa. En 1946, se descubren partículas de masa media en los rayos cósmicos, se les asignaron los nombres de π (pión) y μ (muón). Según los resultados obtenidos, la partícula a la cual se refería Yukawa en su teoría era el π , el cual se produce en la parte superior de la atmósfera; mientras que la partícula observada en 1937 era el μ , siendo este un producto del decaimiento del primero y el que puede identificarse a nivel del mar. Un año después, en 1947 Rochester y Butler descubren una partícula eléctricamente neutra con dos veces la masa del pión, el kaón K^0 . Y se desencadena el descubrimiento de diversas partículas extrañas en los años siguientes a 1950.

En el año 1962 se introduce el término Hadrón para referirse a aquellas partículas que sufren la interacción fuerte, como la que se presenta entre los elementos constituyentes del núcleo atómico. Para 1964 Gell-Mann y Zweig proponen una teoría en la que los hadrones se encuentran constituidos por tres elementos básicos, en distinta cantidad, a los que se les llamó quarks. Para estos nuevos constituyentes se consideró una carga eléctrica igual a una fracción de la carga del electrón y un espín también fraccionario. De modo que la suma de carga eléctrica corresponde a la carga del elemento que constituyen. Además de estas características O.W. Greenberg propone una nueva cualidad para los quarks la cual se denominó “color”, teniendo esta tres posibilidades (verde, rojo o azul); de este modo se limitaba a los hadrones a ser “incolores” o tener carga neutra de color debido a la suma de colores o que los tres colores se encuentren en igual cantidad. Para cada uno existe su

correspondiente antiquark con una carga opuesta, siendo así que existen dos diferentes tipos de hadrones; estos pueden ser mesones, constituidos por un par quark-antiquark, o bariones, que se componen por tres quarks. Del mismo modo puede considerarse a los antibariones como otro tipo de hadrones, los cuales se forman por tres antiquarks.

Existen tres diferentes generaciones de quarks:

Generación	Sabor	Carga
1ra.	arriba (u)	$\frac{2}{3}e$
1ra.	abajo (d)	$-\frac{1}{3}e$
2da.	encanto (c)	$\frac{2}{3}e$
2da.	extraño (s)	$-\frac{1}{3}e$
3ra.	cima (t)	$\frac{2}{3}e$
3ra.	fondo (b)	$-\frac{1}{3}e$

Cuadro 1: Generaciones de Quarks.

Finalmente, en el año 2012 los experimentos ATLAS y CMS, que se encuentran en el LHC, confirmaron la existencia de el último elemento del modelo estándar hasta la fecha, un nuevo bosón elemental [2]. Tal mediador difiere de los otros fermiones fundamentales y bosones al tratarse de una partícula escalar y poseer un espín 0. Además de esto provee el mecanismo por el cual todas las demás partículas y ella misma poseen masa. Una tabla que representa la línea del tiempo de los descubrimientos mas relevantes en los últimos dos siglos en la física e partículas se presenta en la Figura 1.

En un periodo de poco más de un siglo las investigaciones realizadas nos llevaron a obtener un mejor entendimiento de la forma en que se encontraba formado nuestro universo; conocer esto no resulta suficiente para terminar de comprender el porque la materia se comporta como lo hace, si no que es necesario conocer los medios o fuerzas mediante las cuales las partículas elementales interaccionan entre si. En conjunto con el descubrimiento de las diferentes partículas anteriormente mencionadas, se desarrollaron teorías que finalmente fueron comprobadas, sobre el mecanismo de interacción entre los diferentes elementos que componen a la materia; actualmente se conocen cuatro teorías diferentes de fuerzas elementales responsables del comportamiento del Universo, cada una con un mediador y un rango de alcance especial.

Key particle discoveries

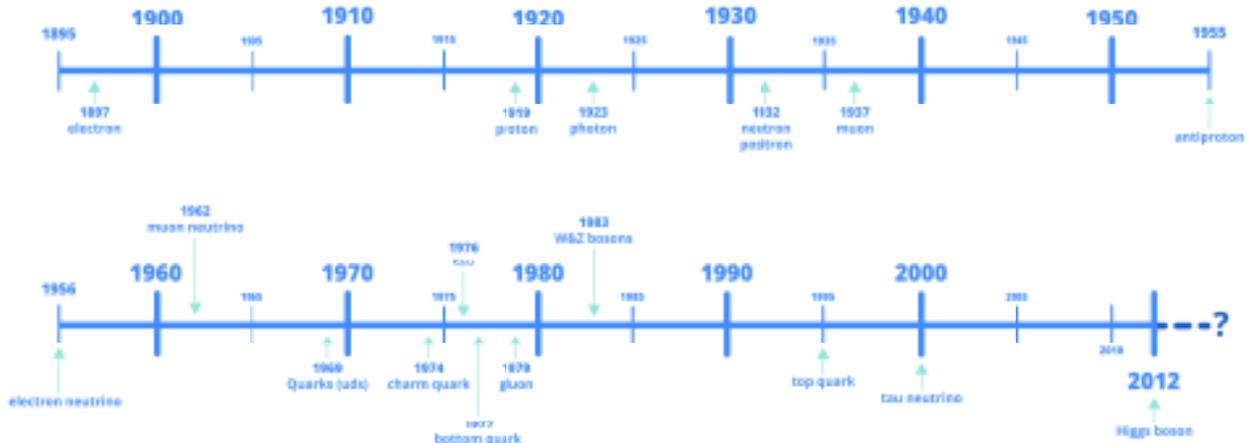


Figura 1: Línea del tiempo de los más importantes descubrimientos en física de partículas que llevaron a la creación del modelo estándar.

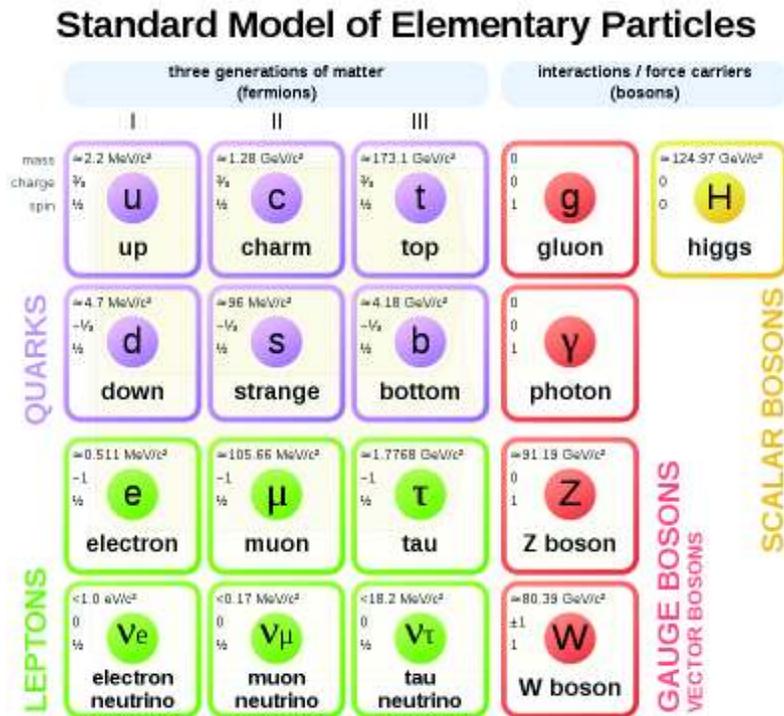


Figura 2: Modelo estándar.

1.2. Tipo de fuerzas.

Fuerza	Partículas que afecta	Mediador	Rango
Gravitacional	Toda partícula con masa	Gravitón	Infinito
Débil	Quarks y Leptones	W y Z	Corto alcance
Electromagnética	Toda partículas con carga	Fotón	Infinito
Fuerte	Quarks y Gluones	Gluón	Corto alcance

Cuadro 2: Fuerzas fundamentales y su partícula portadora (bosón), ordenadas de menor a mayor intensidad según su rango de alcance en escala de partículas.

La fuerza gravitacional siempre resulta atractiva entre los cuerpos y es la responsable de la estructura a gran escala del Universo. Al ser comparada su magnitud con otra fuerza esta resulta ser extremadamente débil, pero posee un rango de alcance infinito. Actualmente se considera que posee un mediador únicamente teórico denominado gravitón pues no ha sido posible obtener resultados observacionales o experimentales que permitan la confirmación de su existencia.

Siguiendo un orden ascendente en su magnitud; la interacción débil es la responsable de la formación de los núcleos atómicos y la interacción de estos con sus electrones orbitantes, por lo que se dice que afecta a quarks y leptones, siendo ambos grupos de partículas fundamentales que conforman la materia. Puede clasificarse en dos tipos dependiendo de si la naturaleza de esta resulta cargada, la cual se encuentra mediada por el bosón W, o neutra, mediada por el bosón Z. Todo proceso entre leptones y quarks referente a la interacción débil puede descomponerse en un conjunto de procesos fundamentales. El proceso fundamental cargado que involucra a los leptones consiste en la existencia de un leptón negativo (e^- , μ^- o τ^-) que emite o absorbe un bosón W^- o W^+ respectivamente; tras esto el leptón se convierte en su neutrino correspondiente. El proceso fundamental para la interacción neutra consiste únicamente en la emisión o absorción de un bosón Z por parte de los leptones, los cuales continúan inalterados, siendo este proceso idéntico para los quarks.

La interacción cargada entre quarks resulta un poco más compleja al tener ciertas limitaciones; el proceso conecta únicamente a quarks de la misma generación, convirtiendo un quark de carga $-1/3$ en su correspondiente quark con carga $2/3$ debido a la emisión de un bosón W^- . De esta forma el proceso realiza un cambio de “sabor” pero conservando la carga de color.

La teoría electrodinámica cuántica resulta ser la de estudio más antiguo y a diferencia de la gravitacional su mediador fue exitosamente identificado y estudiado. Tal fuerza afecta a toda partícula que posea carga eléctrica no nula, para las cuales el proceso elemental es la absorción o emisión de un fotón. Siendo esa la forma en que dos partículas cargadas interactúan entre si mediante el intercambio de fotones.

La cuarta fuerza elemental es la interacción fuerte, o cromodinámica cuántica QCD, la cual afecta a quarks y gluones, siendo estos últimos los mediadores con la singularidad de tener la capacidad de interactuar entre ellos. En el intercambio de gluones entre partículas se considera la característica de la carga de color, similar al intercambio de fotones. Se consideran tres cargas de color: rojo, verde y azul; para lo cual deben encontrarse en igual cantidad para mantener los procesos neutros. Mientras que los quarks poseen únicamente una carga de color, los gluones poseen una carga positiva y una negativa.

2. Origen de los Jets en la física de partículas

2.1. Principio de confinamiento.

Considerando la interacción entre quarks libres como el intercambio de gluones y debido a la interacción atractiva entre estos mismos se forma un campo “cilíndrico” de carga de color. A grandes distancias la densidad del campo es constante, de modo que el campo resulta proporcional a su separación. Por lo que para separar dos quarks una distancia macroscópica, o al infinito, sería necesaria una cantidad infinita de energía [3]. Es por esto que se considera que los objetos con carga de color se encuentran en arreglos neutros, hadrones, sin la posibilidad de propagarse libremente, siendo así que un quark no puede separarse de su hadrón correspondiente sin formar uno nuevo. Tal proceso de creación se describe a continuación.

2.2. Proceso de Hadronización

El proceso de hadronización se da en una colisión de partículas a altos niveles de energía tales que conlleven a la obtención de gluones y quarks, los cuales se reorganizan en hadrones que se alejan entre sí siguiendo su dirección inicial. Conforme aumenta su distancia y con ello la energía en el campo entre ellos comienza el proceso de hadronización que consecuentemente permite observar un conjunto de hadrones o jets. El proceso de hadronización puede describirse en los siguientes eventos y puede observarse una representación visual en la Figura 3:

- a. El quark y antiquark producidos se separan a altas velocidades.
- b. Conforme se separan el campo de carga de color entre ellos se limita a un tubo con una densidad de energía de aproximadamente 1 GeV/fm .
- c. Al medida que la separación continua, la energía contenida en el campo de color es suficiente para proveer la energía necesaria para formar nuevos pares quark-antiquark.
- d. Este proceso continua y nuevos pares quark-antiquark son producidos.
- e. Hasta que los quarks y antiquarks tienen suficiente baja energía para combinarse en arreglos de hadrones sin carga de color. Para resultar en dos jets de hadrones,

uno con la dirección del primer quark y otro con la dirección original del primer antiquark.

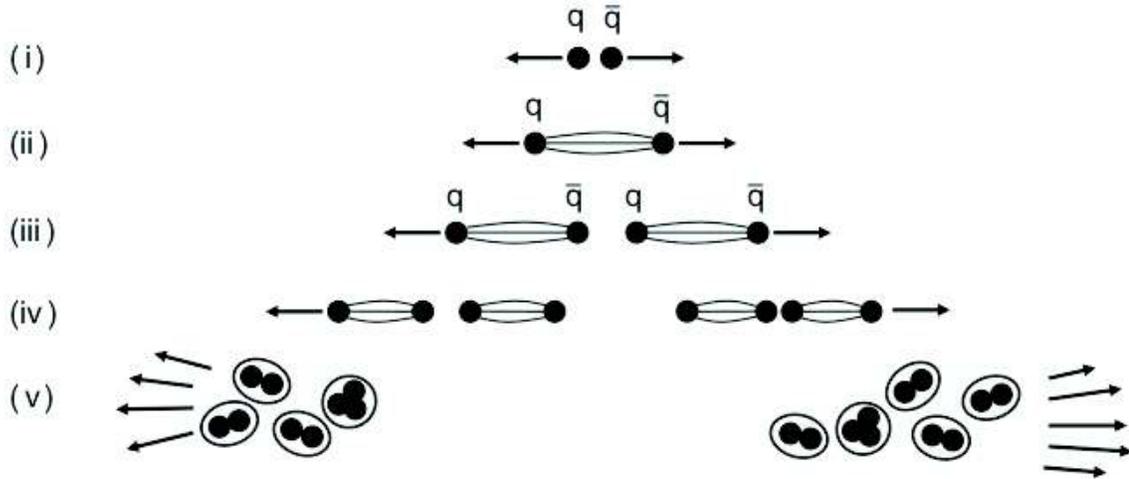


Figura 3: Hadronización, formación de jets.

3. Observaciones experimentales de Jets

El acelerador PETRA, Positron-Electron Tandem Ring Accelerator por sus siglas en inglés, en 1978-1979 cuya principal función era la búsqueda de quarks arriba (t) y leptones pesados, más tarde fue utilizado para comenzar la búsqueda de nuevos bosones gauges, siendo hasta entonces el fotón el único observado experimentalmente. Teóricamente se esperaba observar el gluón, siendo este el responsable de la interacción fuerte, pues se habían obtenido evidencia indirectas de su existencia en la dispersión del neutrino y la dispersión inelástica de electrones. De esta última se observaban resultados inconsistentes con el modelo del parton que involucraba únicamente quarks.

La teoría señalaba que la forma más sencilla de producir un gluón era mediante el proceso de colisión electrón-positrón obteniendo como resultado quark-antiquark-gluón. Ya conocido se tenía que un proceso de alta energía como la aniquilación electrón-positrón resultaba en dos jets de hadrones con direcciones opuestas. Al ser el gluón también un cuerpo con carga de color se esperaba observar un tercer jet como resultado de su hadronización.

A diferencia de un proceso de dos jets, para poder observar un tercero era necesario encontrar el plano formado por los tres momentos; no fue si no hasta Mayo de 1979 que Wu y Zobernig [4] realizaron la publicación de su algoritmo para encontrar el plano que permitiera la observación clara de los tres jets. En Junio de 1979 el experimento descubre un evento que contenía tres jets, una representación gráfica de las observación de dicho evento se pude ver en la Figura 4

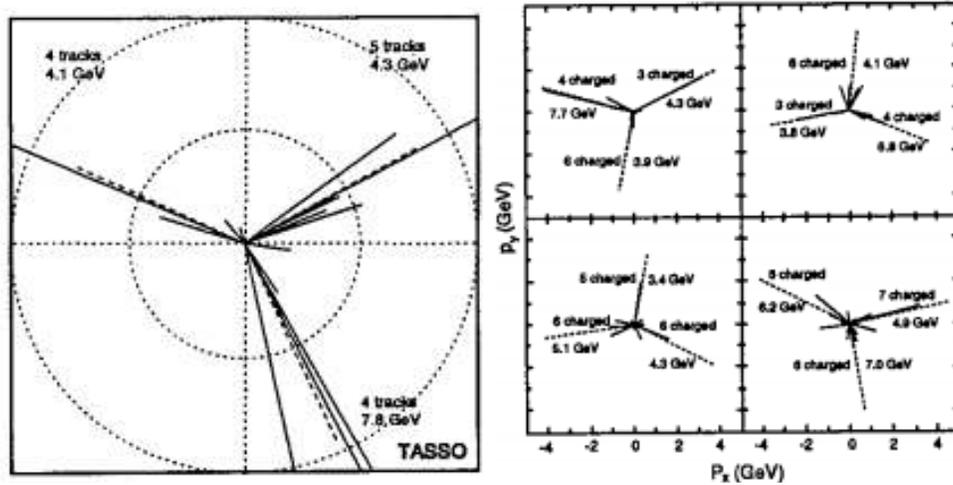


Figura 4: Primer evento de 3 jets observado por PETRA (Izquierda) y evento de 3 jets observado por TASSO (derecha) [5]

El análisis del resultado obtenido, la observación del evento requerido de tres jets, es que al ser los quarks y antiquarks fermiones, la colisión de dos fermiones (e^+ , e^-) no puede convertirse en tres; siendo así que la existencia de un evento con tres jets en tal aniquilación implica el descubrimiento de una nueva partícula. Similar a los quarks, esta nueva partícula tiene un proceso de hadronización que conlleva a obtener el tercer jet y por lo tanto no puede tratarse de una partícula sin carga de color. Siendo entonces explicados más naturalmente por radiación de frenado no colinear en QCD, la cual era la teoría que señalaba la existencia del gluón. De este modo puede explicarse la observación previa que se tenía de procesos con dos jets, pues al disminuirse el momento de uno de los jets en un proceso de tres jets, o bien el ángulo entre dos de ellos disminuye, lo que se observaba era un viento con un jet claramente más grande que el otro; siendo estos los eventos que se buscaban en los inicios de PETRA, una visualización de esto puede observarse en la figura 5.

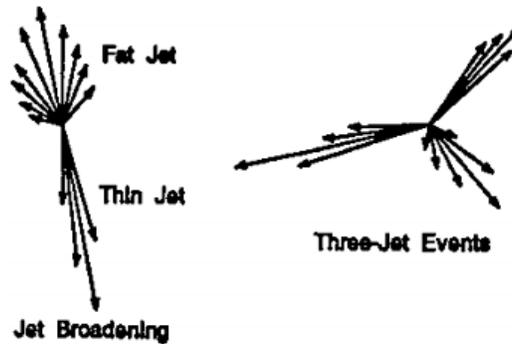


Figura 5: Comparación de un jet amplio y un evento de tres jets [5]

4. Gran Colisionador de Hadrones y experimento CMS

El Gran Colisionador de Hadrones, LHC por sus siglas en inglés, es el acelerador de partículas más grande del mundo y se localiza en la frontera entre Francia y Suiza a una distancia de 100 metros bajo tierra. En el haces de protones viajan en direcciones opuestas con una velocidad cercana a la de la luz y se hacen coincidir para colisionar en puntos determinados. En los puntos de colisión se encuentran detectores de partículas cuyo principal objetivo es el de recolectar el producto de las colisiones para su posterior análisis. Los dos experimentos de mayor tamaño son ATLAS (A Toroidal LHC Apparatus) [6] y CMS (Compact Muon Solenoid) [7] cuyo mayor logro hasta el momento ha sido el descubrimiento del bosón de Higgs en 2012 [8]. Después del descubrimiento del bosón de Higgs los experimentos del LHC han recolectado datos de forma continua hasta el año 2018 buscando realizar mediciones más precisas de procesos del modelo estándar y la posible detección de nuevas partículas las cuales puedan explicar fenómenos más haya del modelo actual como la composición de la materia oscura.

A partir del 2018 los experimentos entraron en una fase de actualización y mantenimiento para reanudar actividades en el año 2021. Los estudios de simulación presentados en esta tesis se basan en la geometría del detector CMS, sin embargo sus resultados y conclusiones pueden ser fácilmente adaptados al detector ATLAS ya que ambos poseen sistemas de detección similares. Una visualización del LHC y detector CMS se muestran en la Figura 6.

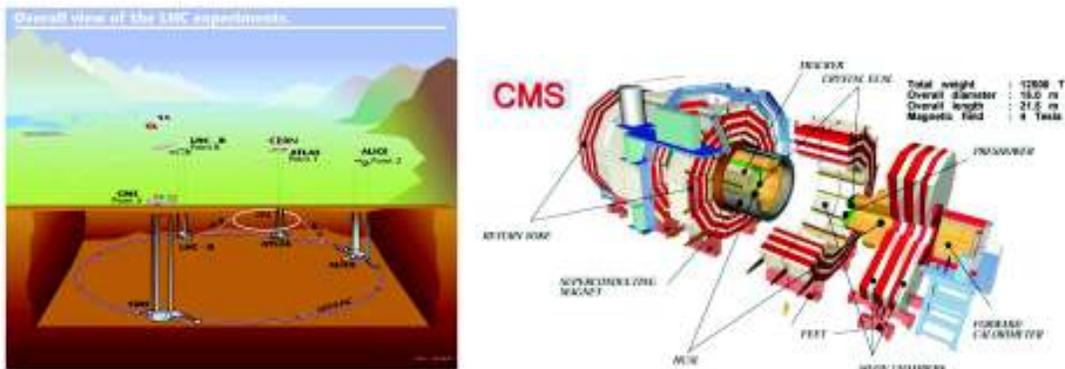


Figura 6: Gran Colisionador de Hadrones (izquierda) y detector CMS (derecha)

El experimento CMS esta equipado con subsistemas de detección cuyo principal objetivo es lograr la detección de todas las partículas del modelo estándar, presentadas en la Figura 2, además de buscar indicios de nuevos tipos de partículas y fuerzas. El concepto de construcción cilíndrica obedece a la necesidad de hacer un detector compacto que cuente con diferentes capas de materiales ordenadas de forma que aquellas partículas que interaccionan mas fácilmente y depositan su energía se identifiquen con aquellos sistemas mas cercanos al punto de colisión mientras que aquellas partículas que interaccionan débilmente con la materia, como es el caso de los muones, sean detectados por sistemas en la ultima capa del CMS. Esta configuración se puede visualizar en la Figura ??.

4.1. Detección de Jets en CMS

Para la identificación y reconstrucción de Jets el detector CMS cuenta con sistemas especiales, entre los cuales se encuentran los Calorímetros (Electromagnético y Hadrónico). Los Jets son formados debido al principio de confinamiento de los quarks descrito en la sección 2.2, las partículas constituyentes de dichos jets interactúan con los materiales de los calorímetros y depositan su energía.

El Calorímetro Hadrónico mide la energía de los hadrones, dicho dispositivo esta compuesto por capas de material fluorescente (centellador) que al interactuar con los hadrones produce un pulso de luz, dicha luz es transportada mediante fibras ópticas al lector (readout) donde foto-detectores amplifican la señal. Cuando la cantidad de luz es sumada sobre las diferentes capas que componen el calorímetro (o también llamadas torres) se puede identificar la presencia de un jet, una representación de esto se muestra en el esquema de la Figura 7. La intensidad de la señal recolectada es una representación de la energía de las partículas constituyentes del Jet, después por medio de criterios de agrupación, descritos en el apéndice C, se puede identificar donde con mayor probabilidad se observe un Jet. Una representación del calorímetro de hadrones utilizado en CMS y la visualización de las torres de Jets es presentado en la Figura 8

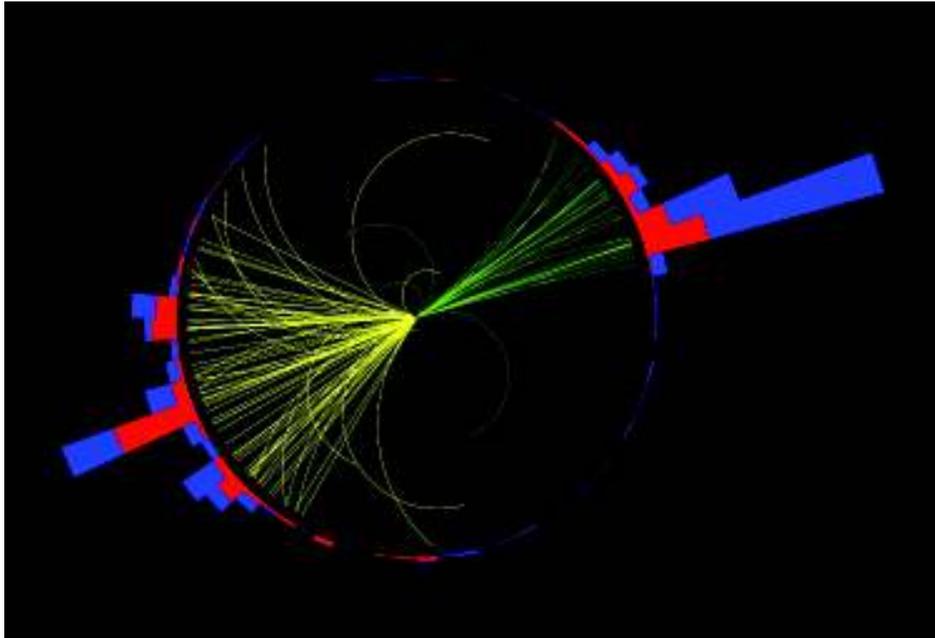


Figura 7: Esquema de la formación de jets

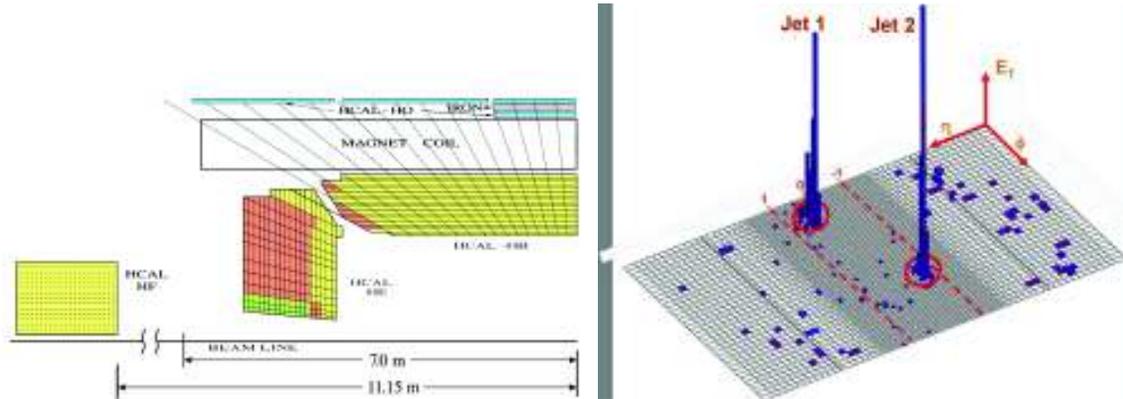


Figura 8: Representación de la estructura del Calorímetro de Hadrones (izquierda). Representación de la reconstrucción de jets (derecha)

4.2. Procesos de estudio

Los procesos simulados que serán utilizados como referencia para la identificación de jets son dos, QCD (cromodinámica cuántica) y W' (W prima). El primero, QCD , describe la interacción por medio del intercambio de gluones entre partículas que poseen carga de color, es decir quarks y gluones. Como se mencionó en la sección 1.2, debido a la interacción entre los mismos gluones es que se considera de corto alcance, pues estos evitan su separación.

Mientras que el bosón W' es una partícula hasta el momento hipotética el cual emerge del sector electro-débil [9] la cual podría ser detectada mediante su decaimiento a un leptón y un neutrino o un quark cima (t) y un quark fondo (b); en particular nos interesa el segundo decaimiento ya que los dos quarks (t , b) pueden ser identificados como jets. La teoría señala que la masa del W' puede estar en el rango de Tera-electronvolts por lo cual dicha señal está accesible a las capacidades del LHC. Diagramas de Feynman de ambos procesos pueden ser presentados en la Figura 9. Cabe resaltar que en la búsqueda de W' , QCD corresponde a uno de los procesos de ruido por lo que es de relevancia el desarrollo de técnicas para lograr diferenciar entre ambos procesos.

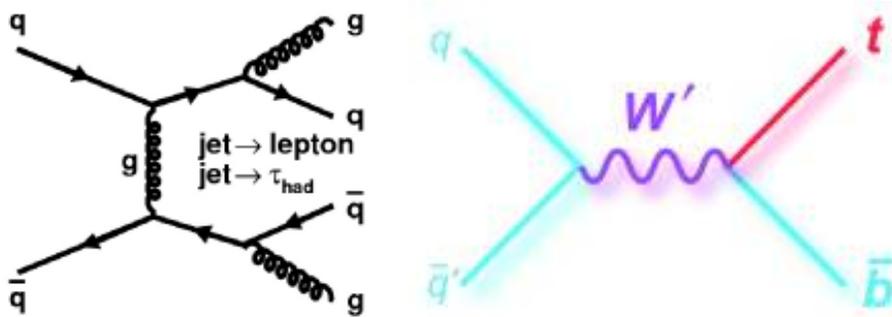


Figura 9: Diagramas de Feynman para los procesos QCD (izquierda) y W' (derecha).

5. Simulación

El proceso de simulación se realiza usando paquetes propios del área de altas energías, usualmente la simulación se da en una secuencia de procesos empezando desde la generación del proceso fundamental el cual está descrito por medio de los diagramas de feynman, probabilidades de decaimientos, recombinación de quarks hasta la simulación de la respuesta del detector y reconstrucción de partículas.

La simulación incluye una serie de procesos en forma serial los cuales están descritos visualmente en el diagrama de la Figura 10.

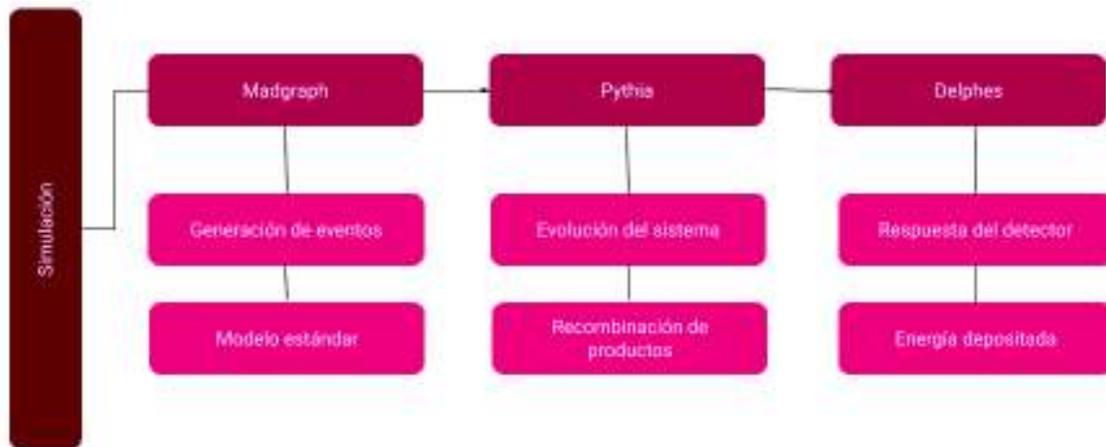


Figura 10: Esquema representativo del proceso de generación de muestras de simulación

Los tres componentes de la simulación y sus respectivas tareas se pueden describir de la siguiente manera: Madgraph [10] es un entorno que contiene los elementos para estudiar de forma fenomenológica procesos del modelo estándar y modelos que describan nueva física. Como cualquier programa de simulación la generación de eventos se realiza por medio del método de Monte Carlo, Madgraph es usado ampliamente para estudios de procesos en colisionadores u otro tipo de experimentos en física de partículas. Incluye elementos como estimación de las secciones eficaces (cross sections) de los procesos, propiedades de decaimiento de las partículas, leyes de física. Adicionalmente incluye una librería amplia para la simulación de una variedad de procesos. Los cálculos pueden desarrollarse con diferentes precisiones, es decir se cuenta con simulaciones a nivel "leading order" (LO) y "next-to-leading-order" (NLO), la diferencia entre uno u otro es el nivel de descripción del posible intercambio de partículas virtuales en los diagramas de Feynman los cuales tienen impacto en la precisión de la probabilidad de producción de dicho proceso.

Para tener una idea de la importancia de la sección eficaz en la Figura 11 se muestran secciones eficaces para diferentes procesos, en particular uno puede notar que los jets, los cuales corresponden a QCD, tienen un valor grande es decir la probabilidad de creación por colisión protón-protón es grande, mientras que el proceso W' aunque no está incluido en la figura corresponde a aquellos procesos con un valor bajo de sección eficaz, es decir

se requiere de un número de datos mucho mayor para tener posibilidad de observar dicho proceso.

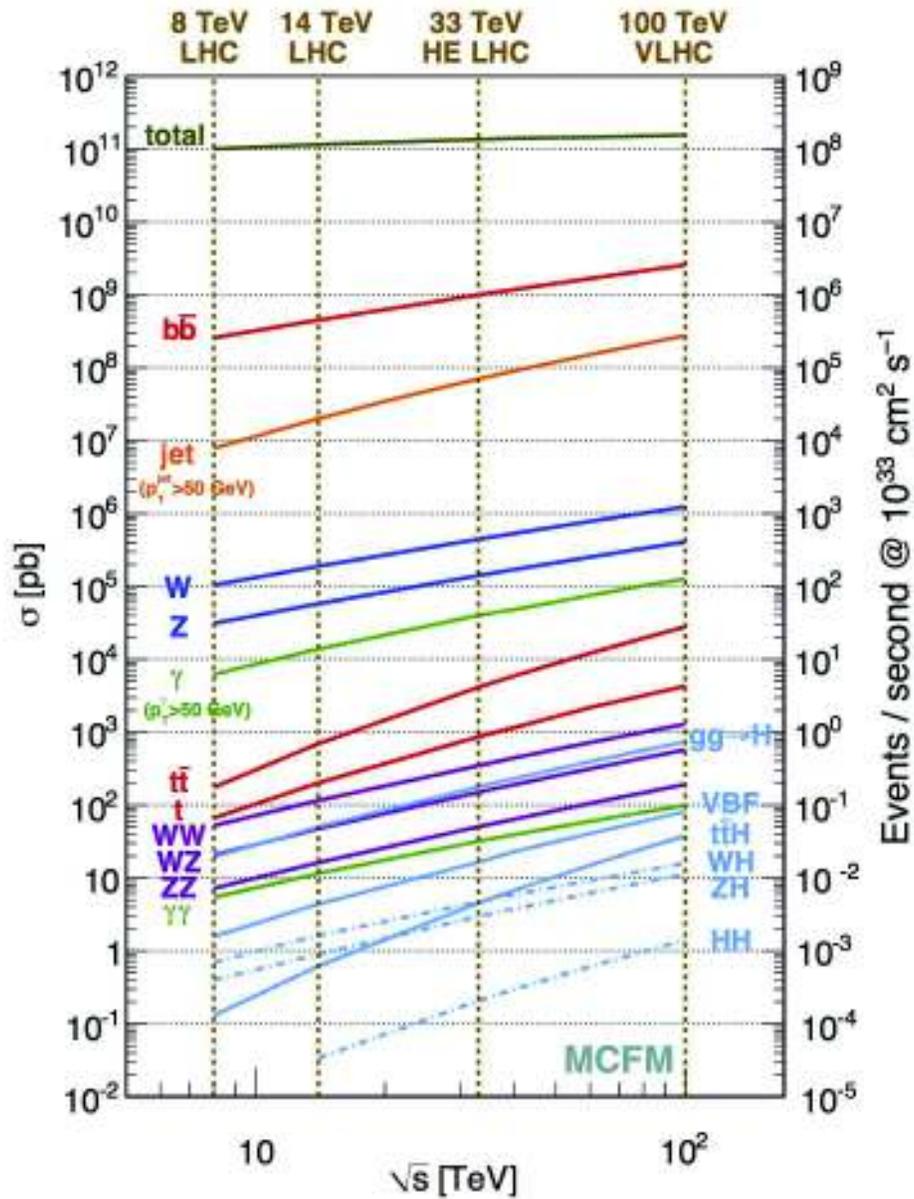


Figura 11: Secciones eficaces para diferentes procesos del modelo estándar.

Comúnmente Madgraph se usa en combinación con el programa Pythia8 [11]. Pythia8 se encarga de describir la evolución del sistema iniciando con un sistema de pocos cuerpos

hasta un sistema complejo de varias partículas producto de la evolución del sistema, recombinación de quarks y producción de diferentes partículas, este paso es fundamental en el proceso de formación de jets, una representación de dicha evolución puede verse en la Figura 12

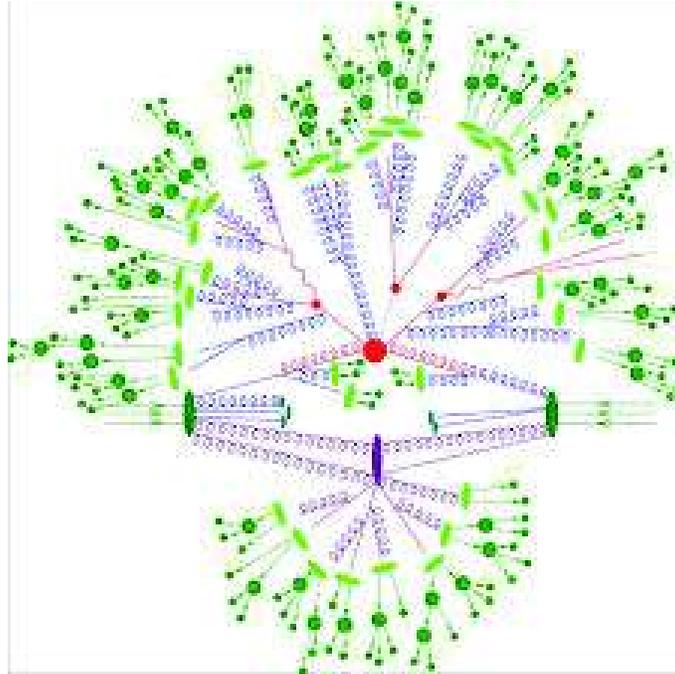


Figura 12: Representación de la evolución de un sistema por el proceso de hadronización

Finalmente DELPHES [12] se encarga de simular la respuesta del detector al paso de las partículas. En particular para la formación de los jets se usa un algoritmo conocido como FASTJET [13] el cual identifica y reconstruye las propiedades de los jets a partir de sus partículas constituyentes y la energía depositada en los calorímetros. Deben considerarse parámetros geoméricamente dependientes, es decir la distribución de las torres de energía juega un papel fundamental en la reconstrucción de las propiedades de los jets, estos parámetros dependen en gran medida de las propiedades del detector utilizado en el estudio. Delphes permite la simulación de un detector multipropósito, es decir consistente con el CMS pero a la vez con un detector con propiedades similares como el ATLAS, esto debido a que los elementos principales como el detector de trazas, calorímetro, campo magnético y detector de muones son comunes a ambos experimentos. La visualización del detector usado en Delphes se puede ver en la Figura 13 donde claramente se distinguen las diferentes capas de detección.

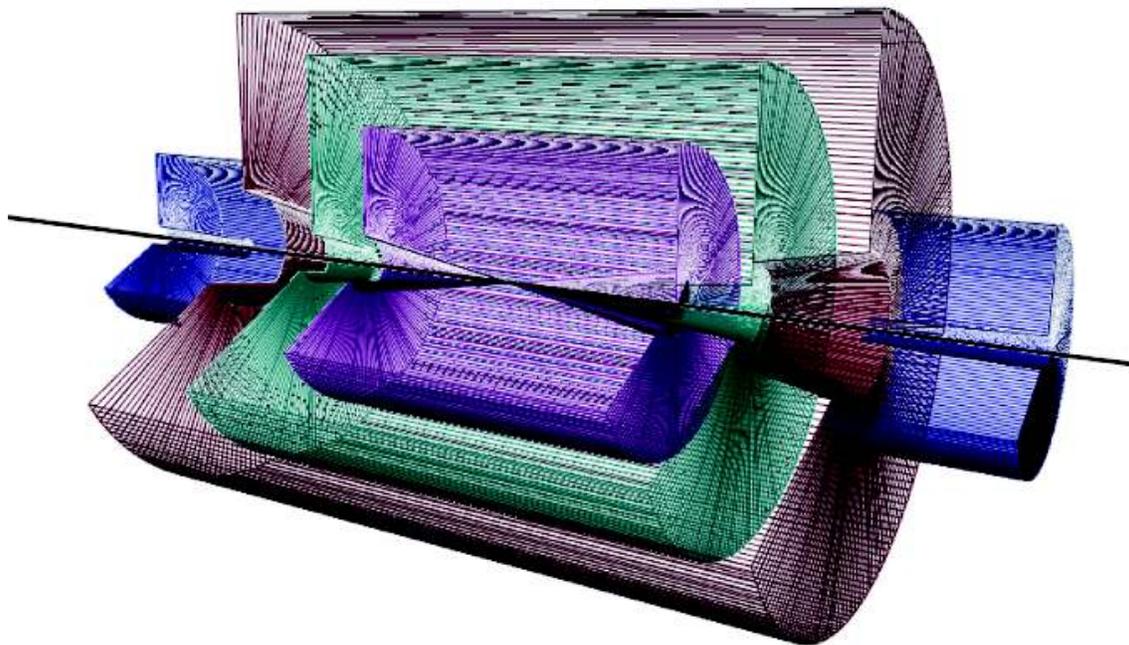


Figura 13: Geometría del detector usado por Delphes consistente con el CMS

Como resultado de la simulación se genera un archivo en formato ROOT[14] que es un paquete de análisis de datos desarrollado en el CERN y que permite la manipulación de los datos, generación de gráficos y selección de propiedades relevantes. El archivo de simulación consiste de contenedores, los cuales están asociados a un determinado tipo de partículas reconstruidas, en el caso de esta simulación el contenedor que almacena la información relevante es el de los Jets que contiene información relacionadas a propiedades de estas partículas y que se puede visualizar en la Figura 14.

Además de la información de Jets se tiene información del proceso de simulación en el contenedor llamado Events, e información sobre las torres de energía en el contenedor llamado towers. Todos estos módulos pueden ser accedidos mediante macros para realizar diferentes selecciones y cortes que permitan extraer la información de interés.

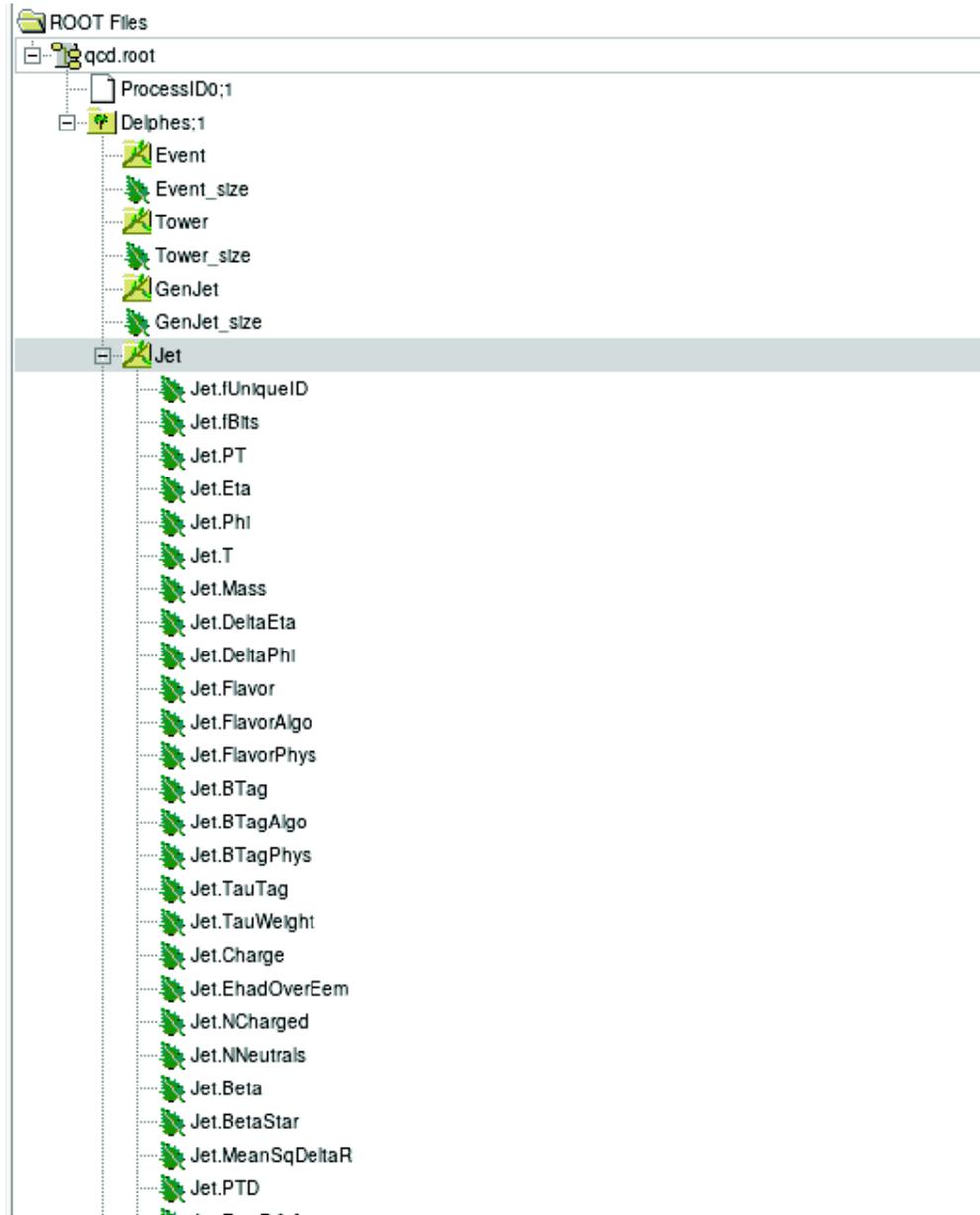


Figura 14: Archivo en formato ROOT que contienen la información de la simulación.

6. Análisis

Una vez generadas las muestras simuladas y teniendo disponibles los archivos en formato ROOT se procede al análisis y selección de eventos donde se pretende extraer información relevante a eventos que contengan jets y filtrado en base a las propiedades de los mismos. La selección de eventos consiste en la implementación de filtros en base a propiedades de los jets reconstruidos, dicha selección consta de los siguientes requisitos:

- Selección de eventos con al menos dos jets reconstruidos, lo anterior garantiza que los eventos de estudio corresponden a procesos reales y no a ruido de detectores.
- Los jets seleccionados deben de tener un momento transversal mayor o igual a 50 GeV, lo anterior garantiza mayor pureza en la selección de jets, ya que si el valor de momento es muy bajo es muy posible que dichas señales de jets puedan proceder de fallas de identificación en los calorímetros
- Los jets seleccionados deben de estar dentro de una ventana definida por el rango de pseudo-rapidez $[-4.0, 4.0]$ y ángulo azimutal $[-\pi, \pi]$, lo anterior garantiza cubrir el rango de detección posible de los calorímetros.

Al aplicar la selección a la totalidad de eventos simulados y para ambos procesos y comparar las distribuciones obtenidas para los dos procesos se puede distinguir claramente una diferencia en las propiedades de los jets generados en cada proceso (QCD y W'), algunas de las variables más relevantes para describir la dinámica de los jets son: el momento transversal (pT), la pseudo rapidez, el ángulo azimutal (plano transversal) y la masa invariante, como referencia las definiciones del momento transversal y pseudo-rapidez se pueden encontrar en el Apéndice E. Lo anterior se puede visualizar en los histogramas de la figura 15 los cuales muestran una comparación de las propiedades mas importantes de los jets, adicionalmente es posible reconstruir la información de los jets por medio de la energía depositada en los calorímetros y producir una imagen 2D como se muestra en la Figura 16. Las imágenes 2D se crearon a partir de la información de la energía depositada por los constituyentes de los jets, para convertir esa información a formato de imagen se realizaron una serie de operaciones y mapeos hasta lograr que cada píxel en la imagen represente una intensidad de energía, para realizar esta operación se utilizo parte de un algoritmo desarrollado por un tercero llamado delphes-gans [15]. De las imágenes 2D se pueden observar patrones que diferencian entre los dos procesos, estos patrones pueden ser utilizados para crear un modelo predictivo por medio de redes neuronales, el cual sea entrenado por medio de muestras de simulación y el cual sea capaz de clasificar a que proceso corresponde cada evento en base a las propiedades de sus jets.

Como información adicional y para efecto de validación de la simulación los eventos seleccionados también pueden ser visualizados en formato 3D por medio de un software especial, en estas visualizaciones se pueden reconocer los elementos generales del detector, la reconstrucción de los jets (representados como conos amarillos) y demás elementos

rojos y azules que representan las deposiciones de energía en los calorímetros, dichas visualizaciones se pueden observar en las figuras 17 y 18 para QCD y W' respectivamente.

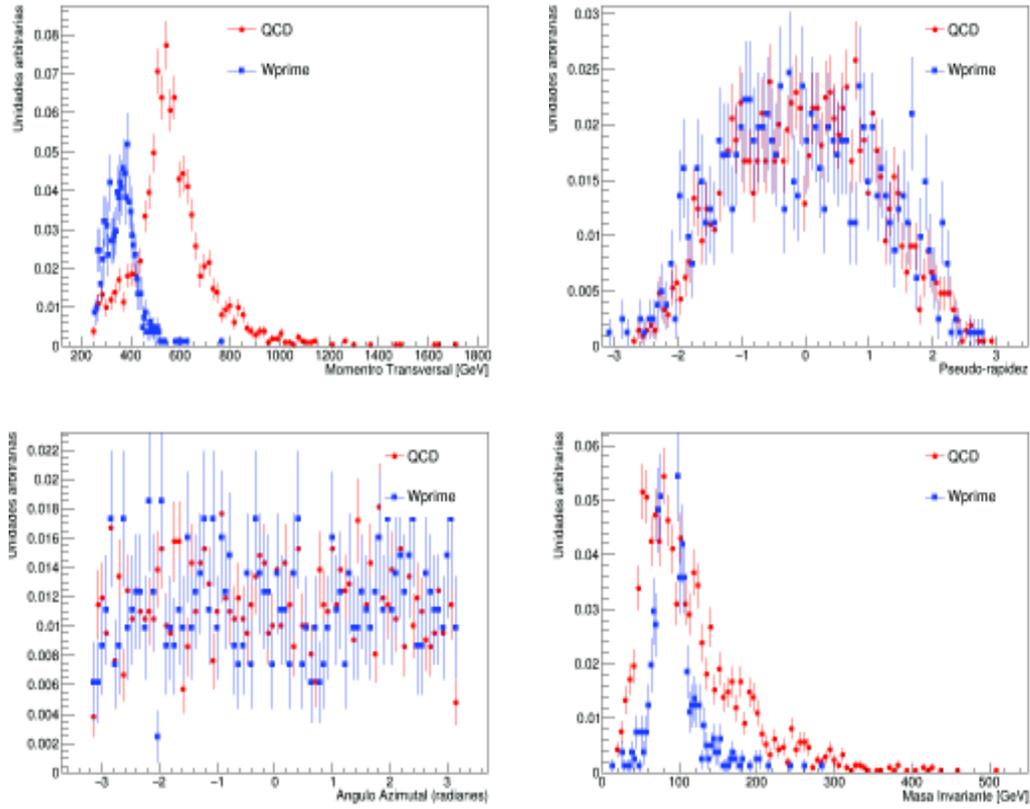


Figura 15: Histogramas que muestran las variables reconstruidas de las Jets comparando las muestras simuladas de QCD y W'

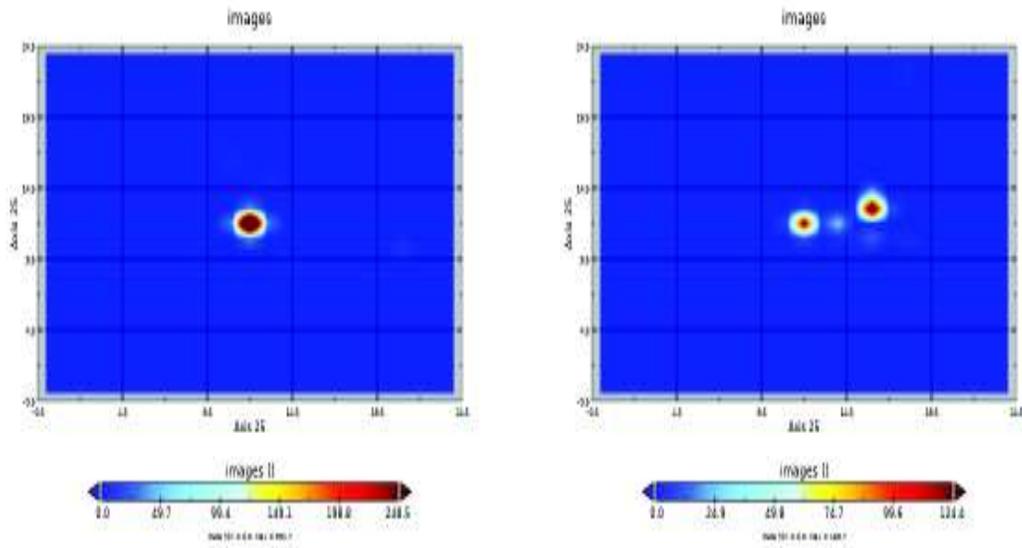


Figura 16: Gráficos 2D que muestran ejemplos de imágenes reconstruidas de Jets para procesos de QCD

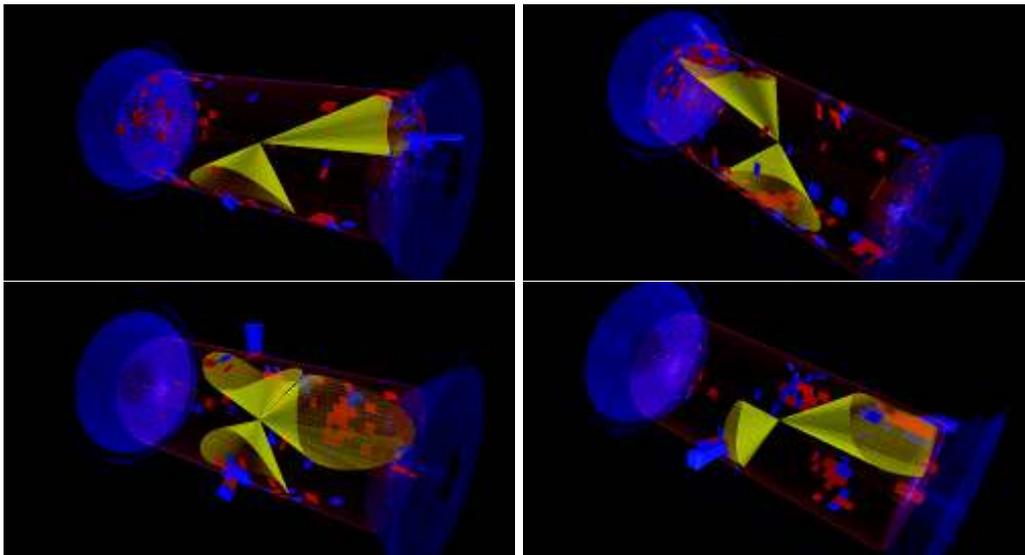


Figura 17: Visualización 3D de los eventos seleccionados en donde se muestra los jets reconstruidos para el proceso QCD.

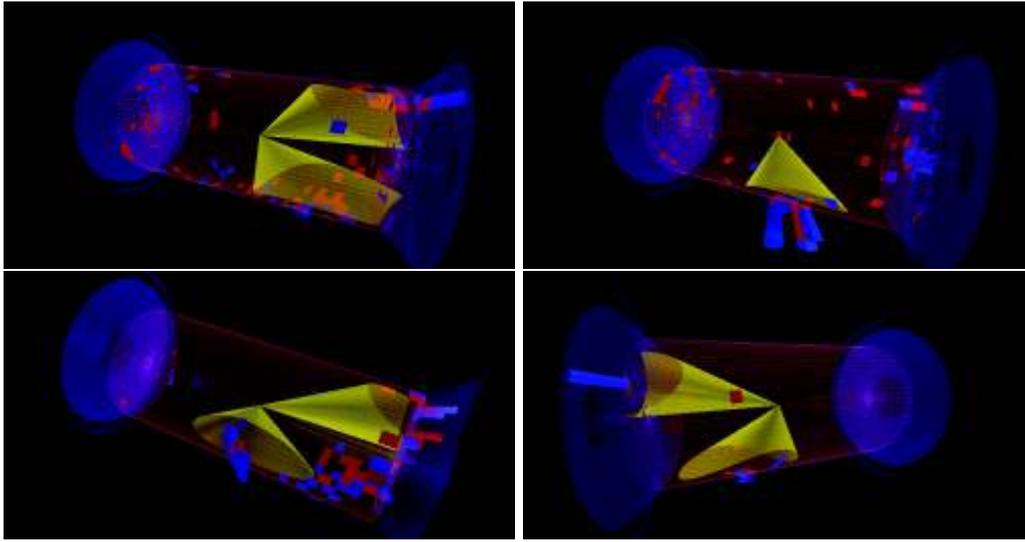


Figura 18: Visualización 3D de los eventos seleccionados en donde se muestra los jets reconstruidos para el proceso W' .

7. Modelos de inteligencia artificial.

Primeramente definamos de manera concisa lo que es la inteligencia artificial; es el estudio del diseño de agentes inteligentes, los cuales son aquellos que hacen lo apropiado para su ambiente y metas, son flexibles ante el cambio de ambos, aprenden con la experiencia y a partir de esto realizan elecciones apropiadas; todo, claro está, dentro de sus limitaciones perceptuales y computacionales [16]. Su principal objetivo es el de comprender los principios que permiten un comportamiento inteligente en sistemas naturales o artificiales. Estos son modelos computacionales inspirados en el comportamiento observado de las neuronas biológicas donde sus unidades interconectadas transmiten señales, las cuales trabajan con información de entrada, pasa por diversas operaciones y finalmente se obtiene un resultado de salida. Entre los enlaces que unen a las neuronas en estos modelos los valores de peso pueden incrementar o inhibir el estado de activación de las neuronas adyacentes. Por su capacidad de aprender que son utilizados en áreas donde la detección de soluciones o características específicas resultan complicadas con la programación convencional. Para llevar a cabo el proceso de aprendizaje los valores de peso entre las neuronas se ven modificados constantemente con el fin optimizar los parámetros de aprendizaje.

El aprendizaje automático, en inglés Machine Learning, es una rama de la inteligencia artificial que se enfoca en el estudio de algoritmos para que los programas aprendan de su experiencia. Su eficiencia depende de la representación de la información sobre la que se aplica; por lo que gran parte del desarrollo del aprendizaje automático es el de desarrollar métodos de transformación de la información que resulten en una representación

de la misma que sea óptima para la extracción de información útil para su clasificación. Entre los diversos métodos de transformación de la información existe la llamada aprendizaje profundo, en inglés Deep Learning, que se compone de múltiples transformaciones no lineales que tiene como fin una representación más flexible y abstracta; esto resulta beneficioso ya que mientras más abstracto sea el concepto este puede reconstruirse en términos de otros menos abstractos. Esto puede observarse en una subcategoría dentro del aprendizaje automático, las redes neuronales convolucionales, del inglés Convolutional Neural Network (CNN), donde los conceptos más abstractos generalmente permanecen invariantes ante los cambios de entrada y mientras mayor sea su abstracción se detectan categorías que cubren una mayor variedad de fenómenos [17].

7.1. Redes neuronales para la clasificación de imágenes (CNN)

Son útiles en la visión artificial, así como clasificación y segmentación de imágenes. La arquitectura de red neuronal llamada CNN es parte constituyente del aprendizaje profundo y comprende un rango grande de aplicaciones tanto en la visión computarizada, análisis de imágenes para seguridad, imágenes médicas, entre otras.

Una CNN es el fundamento principal de varios de las tecnologías de visión computarizada. Hace uso de dos operaciones llamadas *convolución*, en inglés convolution y *agrupación*, en inglés pooling, ambas operaciones con el fin de captar los patrones de las imágenes al mismo tiempo que se reduce la información. Un esquema de una CNN se puede observar en la figura 19.

A continuación se describe a detalle cada una de las capas usadas en una red CNN y como el procesamiento de la imagen original se descompone por medio de las diferentes capas y algoritmos, cabe aclarar que el reconocimiento de imágenes por medio de redes neuronales es un campo activo y de desarrollo, sin embargo existen reglas definidas de aprendizaje [18] las cuales pueden ser usadas para diferentes estudios.

- Capa de convolución (*Convolution layer*): Esta capa usualmente se compone de varias imágenes obtenidas de la imagen original. Estas imágenes se obtienen por medio de el uso de una serie de filtros en forma de matrices (kernel) y un valor bias, estos filtros son aplicados de forma secuencial sobre la imagen original para producir varias imágenes de mas baja dimensionalidad que capturan ciertos patrones. De forma matemática si la imagen original se llamara I , el conjunto de filtros K y el valor bias b . Si la imagen original es en la escala de grises entonces la operación de convolución se representaría por la siguiente ecuación:

$$(I * K)_{ij} = \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} K_{m,n} \cdot I_{i+m,j+n} + b \quad (1)$$

A los elementos en la capa de convolución se le aplica lo que se llama función de activación, uno de las capas de activación mas comunes es la conocida como ReLU

la cual aplica un mapeo a cada uno de los elementos constituyentes de la capa de convolución por medio de la siguiente función:

$$f(x) = \max(0, x) \quad (2)$$

Dicha función manda todos los valores negativos a cero y mantiene los positivos, lo cual tiene ventajas durante el proceso de convergencia de la red neuronal pues únicamente necesita dos ciclos de reloj.

- Capa de agrupación (*Pooling layer*) : La capa de agrupación se usa para reducir la dimensionalidad de la capa de convolución. La agrupación se realiza mediante una matriz la cual barre las imágenes de la capa de convolución y en cada elemento de esas imágenes se escoge el elemento (píxel) de mayor valor (a esto se le llama max pooling). Esto permite general la red de agrupamiento la cual contiene un número de elementos similar a la capa de convolución pero con elementos de menor dimensionalidad.
- Capa completamente conectada (*Fully Connected layer*): Los elementos de entrada son los que provienen de la red de agrupación, en este nivel la información que hasta ahora estaba estructurada en formato de matriz es descompuesta en forma unidimensional de tal forma que cada elemento representa una neurona, a partir de esta capa lo que sigue puede ser interpretado como un perceptron multicapas (Multilayer neural network) [19], la cual por lo general esta compuesta por al menos una capa oculta y una capa de salida la cual se ocupa de evaluar a que clase pertenece la imagen original. Una visualización del perceptron multicapas puede verse en la Figura 19

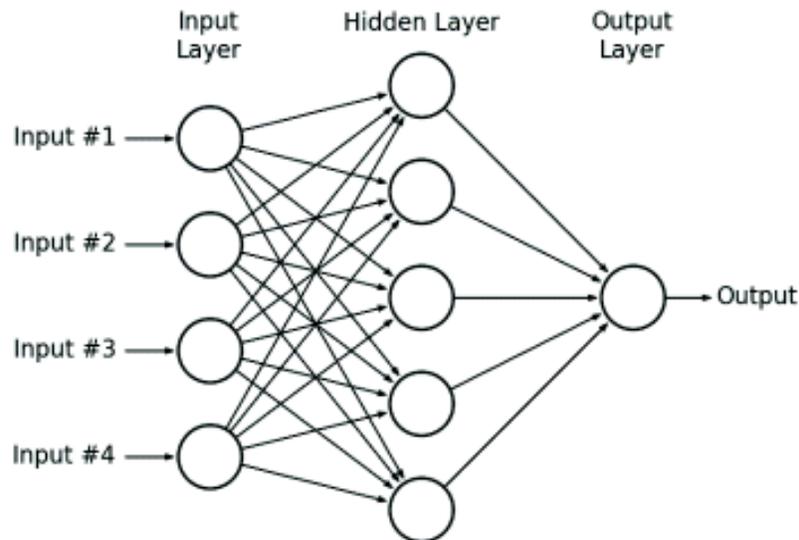


Figura 19: Visualización de un perceptron multicapas

7.2. CNN para identificación de Jets

En la sección anterior se describió a detalle los componentes de una red CNN general, en esta sección se darán detalles de la red neuronal que se uso para analizar las imágenes de Jets y entrenar un modelo para lograr en base a las propiedades de esas imágenes construir un modelo que permita clasificar las mismas en un proceso QCD o W' .

El tipo de capas es como la descrita en la figura 20, sin embargo hay diferencias en cuanto al orden de algunas de las capas y el número de capas de convolución y agrupamiento, a continuación se describe la secuencia utilizada:

- Capa de entrada: Corresponde a las imágenes originales las cuales son arreglos bidimensionales de 25x25 píxeles, como las que se muestran en la Figura 16.
- 1era Capa de Convolución: Se utilizan 6 filtros, cuyos valores (pesos) son iniciados de forma aleatorio, cada filtro con una dimensión de 5x5.
- 1era capa de agrupamiento: Donde se utiliza una matriz de 2x2 para seleccionar el elemento de mayor valor.
- Función de activación: Se usa la función ReLu.
- 2da capa de convolución: Con un número de 16 filtros con una dimensionalidad de 5x5.
- 2da capa de agrupamiento: Se utiliza una matriz de 2x2 para seleccionar al elemento de mayor valor.
- Función de activación: Con el uso de la función ReLU.
- Capa completamente conectada: Se reciben los elementos de la 2da capa de agrupamiento y se ordenan de forma unidimensional los píxeles, adicionalmente se le aplica la función de activación a dichos elementos .
 - Capa oculta: Con un número de 128 neuronas.
 - Capa de salida: Con dos neuronas las cuales reciben la información de la capa oculta y evalúa el procesamiento y decide a que categoría pertenece la imagen.

La red neuronal utilizada se puede observar gráficamente en la imagen del apéndice D.

En la ultima capa el resultado es evaluado por dos neuronas cada una de ellas encargada de una de las clases de los procesos QCD y W' . Cada neurona procesa un valor proveniente de la capa oculta y se aplica un proceso de normalización de este valor usando la función Softmax, la cual normaliza el valor en un intervalo entre 0 y 1, dicha función esta definida por la siguiente formula

$$\sigma(z)_i = \frac{\exp^{z_i}}{\sum_{j=1}^K \exp z_j} \quad (3)$$

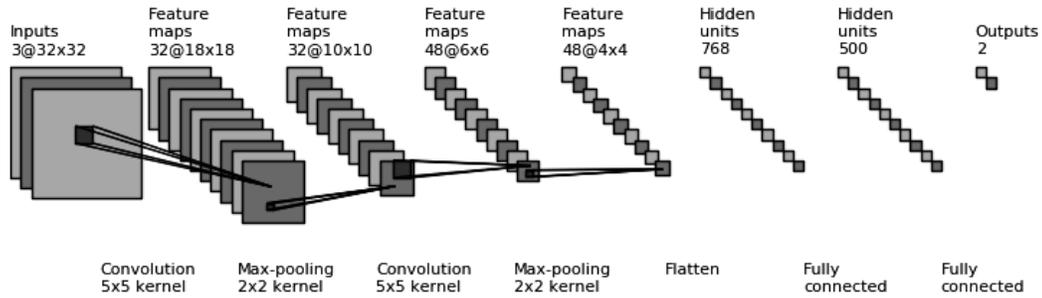


Figura 20: Construcción de una red CNN común

Después de aplicar esta función se toma el valor mayor entre estas dos neuronas para definir el resultado de la red neuronal. Las otras dos variables relevantes en el proceso de aprendizaje son la función de costo, la cual está definida por medio de la función de entropía cruzada, en inglés *cross entropy*, y la precisión (*accuracy*) que básicamente se define como el porcentaje de éxito en cada iteración del aprendizaje, es decir la comparación entre el resultado de la red neuronal y el valor esperado. El proceso de aprendizaje se basa en el concepto de que en cada iteración la función de costo o también conocida como error reduzca y que la precisión aumente, utiliza un algoritmo por medio del cual se logra el aprendizaje que se conoce como propagación hacia atrás e involucra un proceso de actualización de los pesos en las capas de convolución y en la red completamente conectada, el formalismo de dicho proceso se describe en el apéndice A.

El software que se utiliza para el entrenamiento de la red neuronal es TensorFlow [20]. TensorFlow es desarrollado por la compañía Google y se ha vuelto un referente en la industria de la inteligencia artificial. TensorFlow es intuitivo y permite crear arquitecturas de redes neuronales a partir de la definición de nodos y conexiones entre ellos por medio de operaciones. El proceso de aprendizaje de la red neuronal es iterativo, se inicia con el flujo descrito anteriormente, en donde los valores de los filtros usados en las capas de convolución son inicializados aleatoriamente, después en la capa completamente conectada las conexiones entre la 1era capa, la capa oculta y la capa de salida son igualmente inicializadas de modo aleatorio, debido a esto en la primera iteración se espera que la red neuronal no tenga el suficiente aprendizaje para clasificar de forma eficaz la imagen. Los parámetros que definen el proceso de aprendizaje están descritos en la siguiente sección.

7.3. Parámetros configurables de la red CNN

Aunque la estructura de una red neuronal CNN contiene elementos característicos y que no pueden faltar (capas de convolución, agrupamiento, completamente conectada), se tiene una lista de parámetros que pueden ser configurados durante el proceso de aprendizaje, de la selección de estos parámetros dependerá la rapidez en el proceso de aprendizaje e incluso la precisión, dichos elementos se conocen como hyper-parámetros y algunos de ellos son los siguientes:

- Factor de aprendizaje: Usualmente este es un número fijo, y está relacionado al proceso de minimización de la función de costo o error, en este proceso de minimización usualmente se utiliza un gradiente, dicho gradiente debe moverse en función de este parámetro de aprendizaje, es decir si el factor de aprendizaje es un número grande la velocidad del gradiente será mayor pero su precisión menor y viceversa, no hay una receta para lograr optimizar este valor más que prueba y error ya que usualmente las funciones de error están descritas por un hyper-volumen con máximos y mínimos difícil de predecir su comportamiento.
- Función de activación: La función de activación usada por default es el ReLU ya que es simple y proporciona buena convergencia en el proceso de aprendizaje, sin embargo se tiene la posibilidad de jugar con otras funciones tales como: Sigmoid y Tanh, una representación gráfica de algunas de las funciones de activación más usadas se presenta en la figura 21
- Función de pérdida: La función de pérdida o error puede ser definida por medio de diferentes algoritmos, el usado en este estudio es el conocido como entropía cruzada.

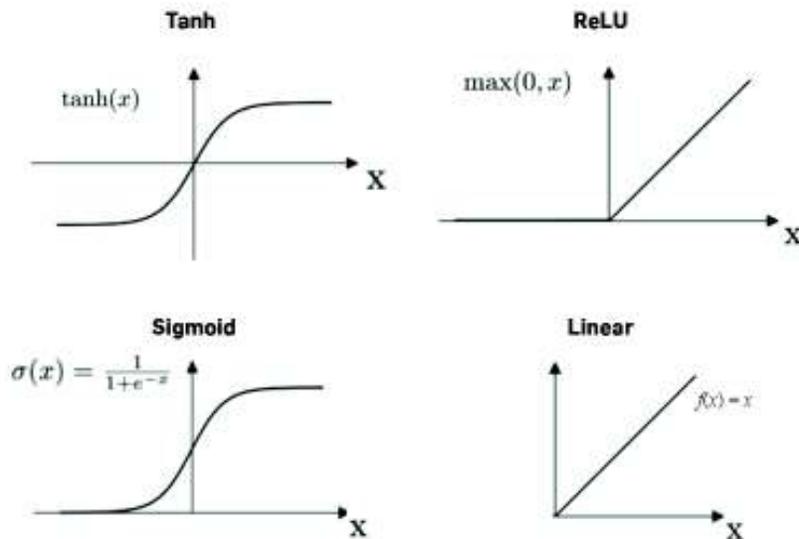


Figura 21: Representación gráfica de algunas de las funciones de activación usadas para redes CNN

8. Resultados

La red neuronal fue entrenada con las imágenes obtenidas por simulación y la configuración descrita en la sección 7.2; el número de imágenes utilizadas para el entrenamiento fue de 10497, mientras que las utilizadas para la validación fue de 2625. El número de iteraciones usadas en el proceso de aprendizaje se configuró de tal manera que se lograra una precisión mayor a 95% y reducir lo mas posible el error tal como se puede observar en la Figura 22.

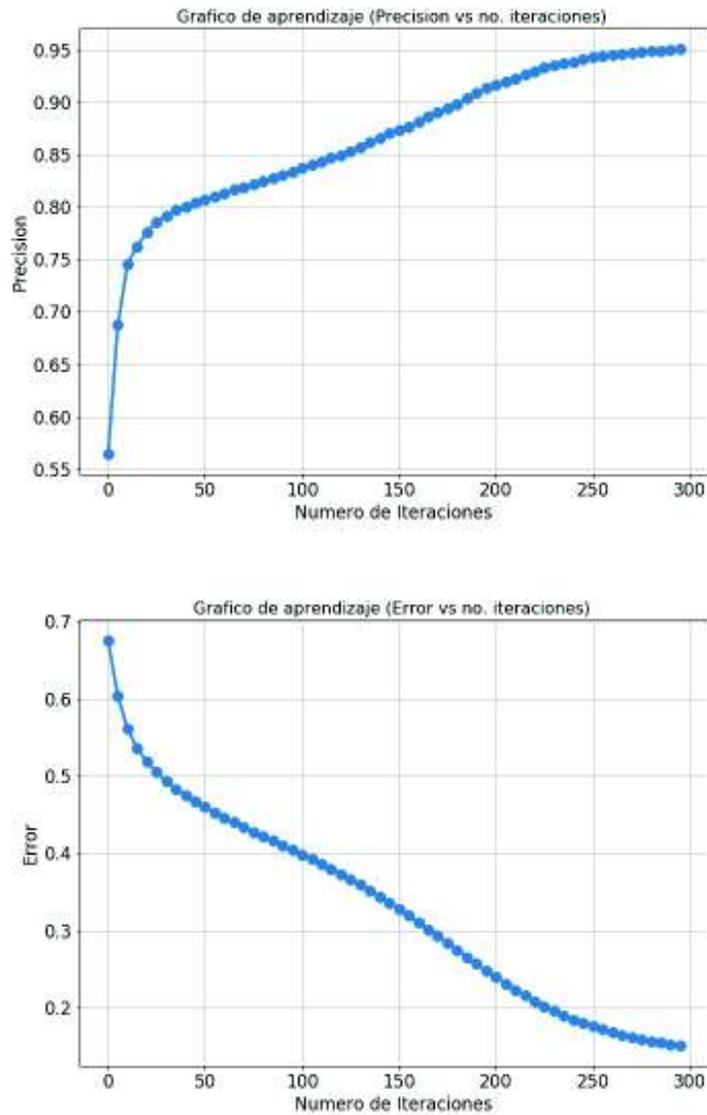


Figura 22: Gráfica de aprendizaje para la configuración estándar de la red CNN

Se utilizaron otras variaciones de función de activación y factor de aprendizaje como se muestra en la Tabla 3 con el fin de observar e identificar la combinación óptima para la resolución de nuestro problema. Las gráficas de aprendizaje para cada una de estas variaciones pueden observarse en el Apéndice B.

Variación	Función de Activación	Factor Aprendizaje
default	ReLU	1e-4.
var1	ReLU	1e-3.
var2	ReLU	1e-2
var3	Tanh	1e-4
var4	Tanh	1e-3
var5	Tanh	1e-2

Cuadro 3: Variaciones de los parámetros configurables de la red CNN

9. Conclusiones

Se ha desarrollado un modelo en base a redes neuronales CNN que es capaz de clasificar imágenes características en la reconstrucción de jet y evaluar a que proceso pertenecen entre QCD y W . Para la obtención de las imágenes se uso una simulación por el método de Monte Carlo la cual describe los procesos de física fundamental y la respuesta el detector al paso de las partículas, el detector que se uso es consistente con el CMS del CERN, en donde se simulan las colisiones de protones ultrarelatvitas y la producción de partículas. Después de obtener los resultados de la simulación, la energía depositada de los jets se convierte a formato imagen (2D) en donde estas imágenes se incorporaron a un algoritmo de red neuronal CNN para lograr construir un modelo productivo que lograra clasificar las imágenes si pertenecen a uno de los dos procesos. Se logro una eficacia mayor a 95 % y un error decreciente por lo que se puede declarar al modelo entrenado. Se variaron los parámetros de la red CNN para verificar su estabilidad y comprar los resultados en cuento a eficacia y error, de tal comparación se puede observar que la configuración que logra una convergencia mas rápida es la que usa una función de activación ReLu y un factor de aprendizaje de 1e-3 (var1). El modelo desarrollado en esta tesis puede ser refinado o misma técnica desarrollada para la clasificación de otros procesos en el área de física de partículas.

Si desea reproducirse los resultados obtenidos puede descargarse MadGraph [21] y encontrarse los códigos utilizados referentes a la conversión de los archivos root a archivos h5 [22] y al entrenamiento de la red CNN [23] en los repositorios. Esto no se encuentra limitado a los procesos utilizados en el trabajo de tesis pues como ya se mencionó en la sección 5 MadGraph tiene información referente cualquier proceso del modelo estándar e incluso modelos de nueva física.

A. Método de actualización de valores (backpropagation)

Propagación hacia atrás es el algoritmo por el cual la red neuronal de una forma iterativa va actualizando los diferentes parámetros para lograr un aprendizaje, es decir reducir la función de costo y aumentar la precisión (accuracy). Como se explica en el capítulo 7 en donde se detallan los elementos esenciales de una CNN los valores que se tienen que actualizar corresponden a los asociados a la capa de convolución (matrices de filtros) y en la capa completamente conectada, dejando fijos la función de activación y los parámetros de la capa de agrupación.

La función utilizada para la optimización en nuestra red es la de entropía cruzada, por lo que la función de error esta descrita como sigue

$$H(p, q) = - \sum_x p(x) \log q(x) \quad (4)$$

Esto mide, entre dos distribuciones de probabilidad, la medida de bits necesarios para identificar un evento en específico dentro de un conjunto de posibilidades. La probabilidad \mathbf{p} es la distribución verdadera y la probabilidad \mathbf{q} es la distribución obtenida de un modelo estadístico construido; donde la primera corresponde al valor obtenido de los datos y la segunda al obtenido por la red neuronal. Minimizar la entropía es equivalente a minimizar la divergencia KL

$$H(p, q) = H(p) + D_{KL}(p||q) \quad (5)$$

donde

$$H(p) = - \sum_x p(x) \log p(x) \quad (6)$$

$$D_{KL}(p||q) = - \sum_x p(x) \frac{\log p(x)}{\log q(x)} \quad (7)$$

Usualmente este procedimiento se realiza por medio de un algoritmo que considera el gradiente de descenso de la función de error. Este gradiente puede ser interpretado como la medida de cambio de la función de error conforme se actualiza alguno de los pesos representativos de los filtros en la capa de convolución por ejemplo. Gráficamente esto se representaría como en la figura 23. Eventualmente el peso actualizado tendrá un efecto en el resultado de la evaluación de la red neuronal, esto será cierto para cada uno de los pesos en las diferentes capas.

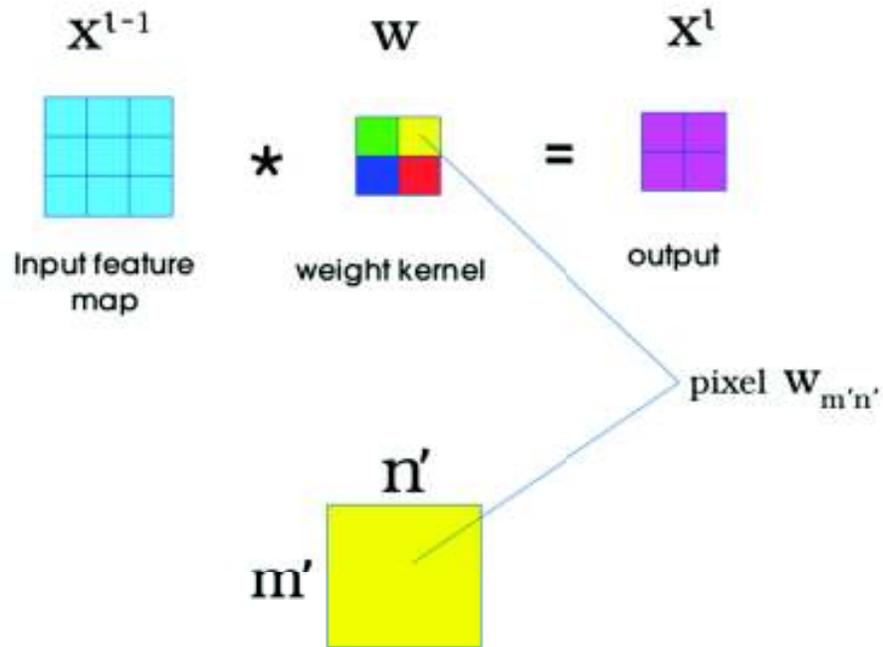


Figura 23: Representación del proceso de propagación hacia atrás para una CNN

El procedimiento completo involucra una serie de derivadas parciales las cuales para efectos de este estudio no se mencionan debido a la complejidad de la notación y la poca relevancia dentro de los resultados, si se requiere conocer la expresión exacta de dichas relaciones se puede consultar bibliografía especializada en el tema.

B. Gráficas de aprendizaje para las variaciones de la configuración de la red CNN

A continuación se presentan las gráficas de aprendizaje de la red CNN para las variaciones descritas en la tabla 3

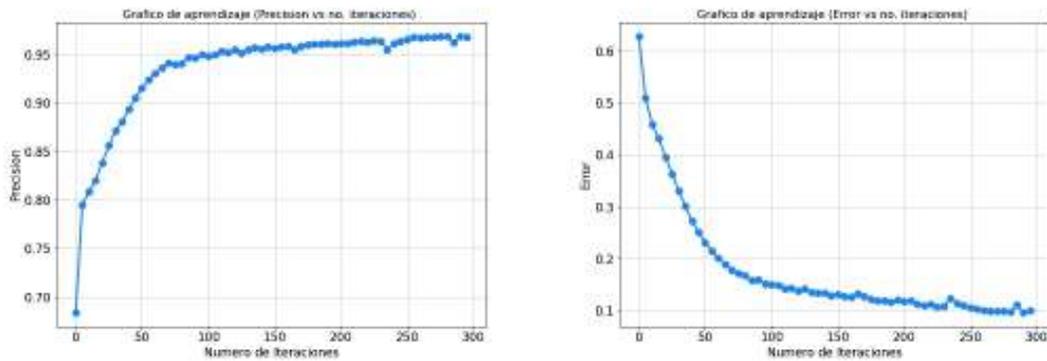


Figura 24: Gráfica de aprendizaje para la variación 1 de la red CNN

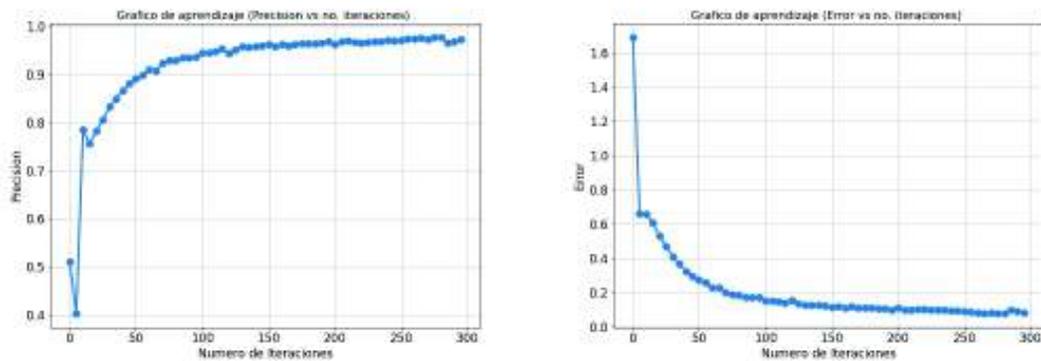


Figura 25: Gráfica de aprendizaje para la variación 2 de la red CNN

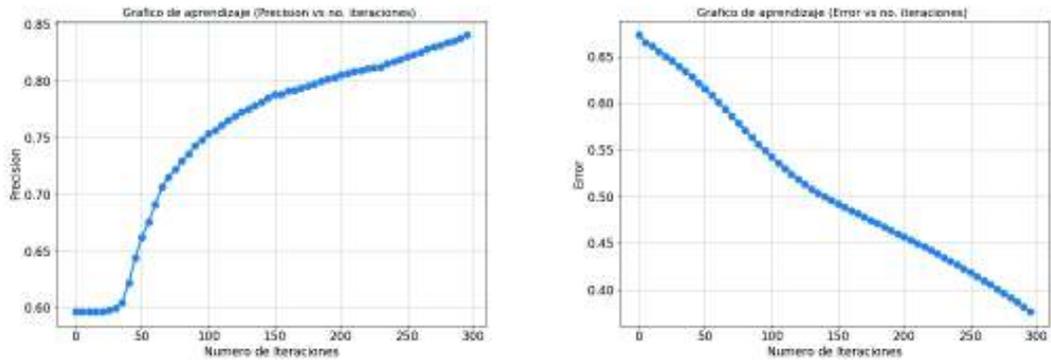


Figura 26: Gráfica de aprendizaje para la variación 3 de la red CNN

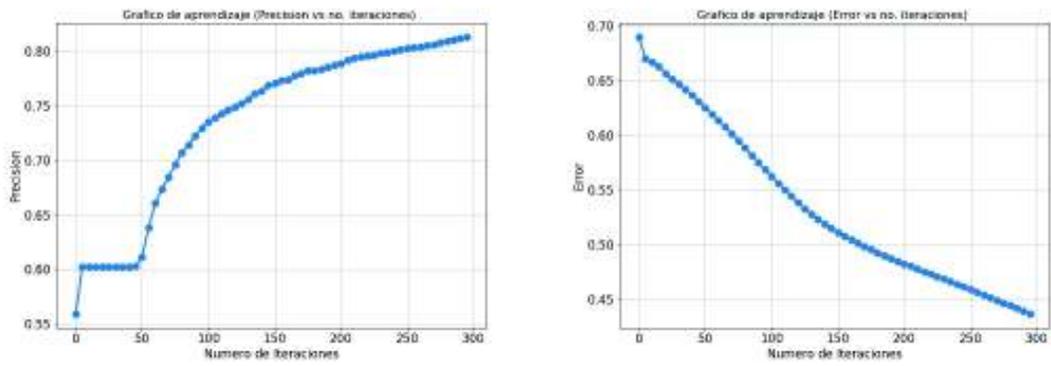


Figura 27: Gráfica de aprendizaje para la variación 4 de la red CNN

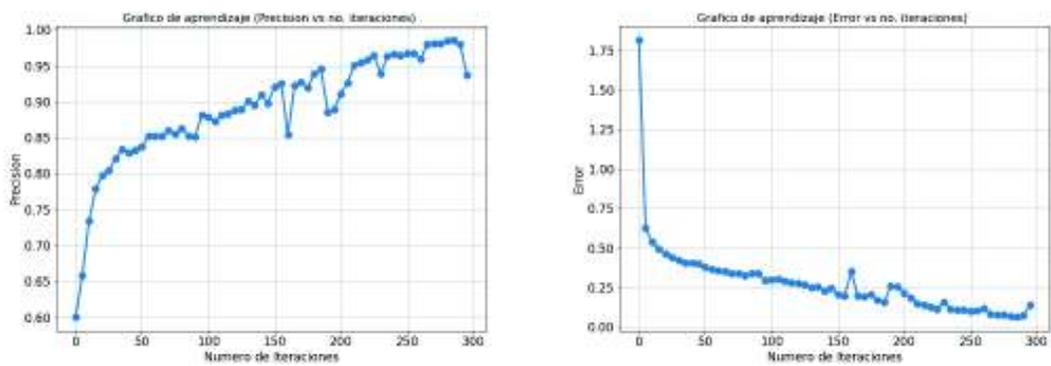


Figura 28: Gráfica de aprendizaje para la variación 5 de la red CNN

C. Algoritmo FasJet para la reconstrucción de Jets

El algoritmo usado para la reconstrucción de jets es el FastJet. El objetivo de cualquier algoritmo de reconstrucción e identificación de jets es el de reducir lo mas posible la posible contaminación (ruido) producido por deposiciones de energía en los calorímetros que no corresponden a las partículas que componen los jets originales, es decir aquellas partículas que provienen de los quarks. FastJet provee de algoritmos para la correcta identificación de jets, usualmente estos algoritmos incluyen las siguientes tareas:

- Identificación por cono: Este algoritmo identifica el flujo de energía en un determinado espacio definido por un cono de cierto radio. Para cada área definida por un cono se suma el cuadrimomento de las partículas
- Agrupamiento secuencial: Se realiza calculando la distancia entre las partículas, combina las partículas con una distancia menor para la formación del jet

Para ambas técnicas se debe realizar una substracción del ruido para evitar la posible contaminación de otros procesos que no sean jets. Fastjet combina lo mejor de los dos algoritmos mencionados anteriormente y optimiza el tiempo requerido para la identificación. La gráfica de la Figura 29 muestra el tiempo que le toma a FastJet el agrupamiento de jets

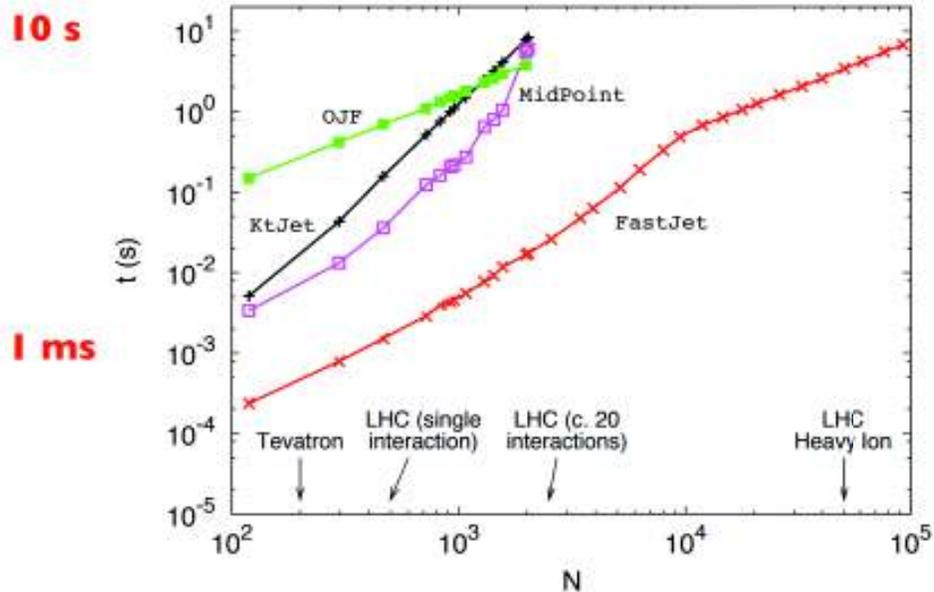
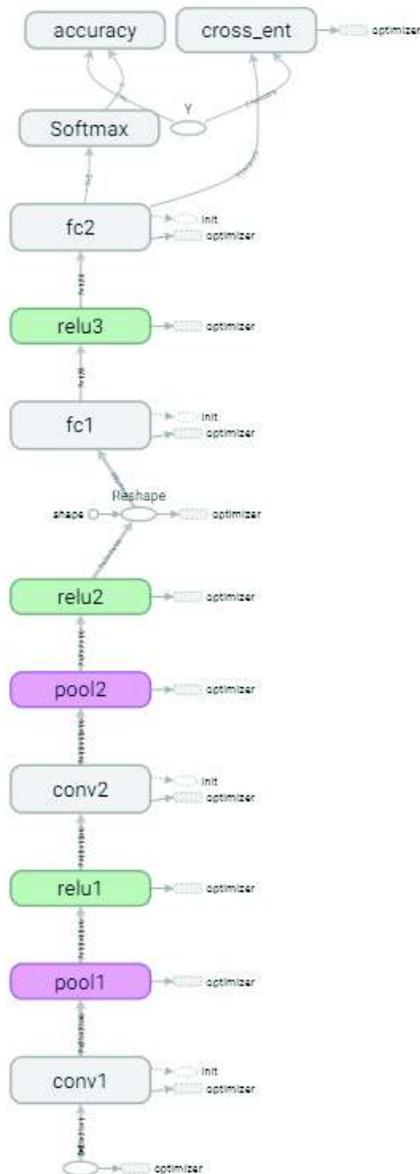


Figura 29: Tiempo que le toma al algoritmo fastjet el agrupar N numero de partículas

D. Nodos y conexiones de la red neuronal en el modelo de Tensorflow

Main Graph



Auxiliary Nodes



Figura 30: Visualización de los nodos y conexiones de la red CNN usada por tensorflow

E. Definición de momento Transversal y Pseudo-rapidez

El momento transversal (p_T) y la pseudo-rapidez son dos de las variables más importantes para caracterizar la dinámica de las partículas reconstruidas por el detector.

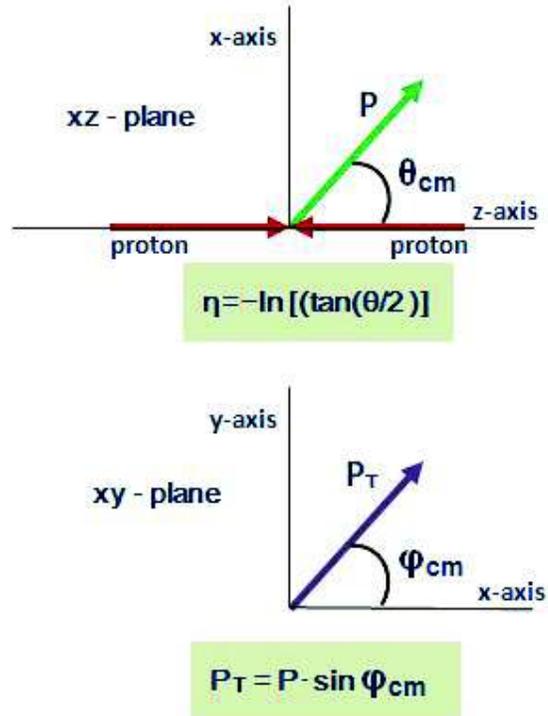


Figura 31: Visualización del momento transversal y pseudo-rapidez

Referencias

- [1] D. Griffiths. *Introduction to Elementary Particles*. Physics textbook. Wiley, 2008.
- [2] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.
- [3] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.
- [4] Sau Lan Wu and Georg Zoernig. A Method of Three Jet Analysis in $e^+ e^-$ Annihilation. *Z. Phys. C*, 2:107, 1979.
- [5] P. Soding, B. Wiik, G. Wolf, and S.L. Wu. The First evidence for three jet events in $e^+ e^-$ collisions at PETRA: First direct observation of the gluon. In *International Europhysics Conference on High-energy Physics (HEP 95)*, pages 3–14, 9 1996.
- [6] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.
- [7] S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.
- [8] A new boson with a mass of 125 geV observed with the cms experiment at the large hadron collider. *Science*, 338(6114):1569–1575, 2012.
- [9] Search for W prime to $t\bar{b}$ in the all-hadronic final state. Technical Report CMS-PAS-B2G-12-009, CERN, Geneva, 2014.
- [10] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7), Jul 2014.
- [11] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to pythia 8.2. *Computer Physics Communications*, 191:159–177, Jun 2015.
- [12] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi. Delphes 3: a modular framework for fast simulation of a generic collider experiment. *Journal of High Energy Physics*, 2014(2), Feb 2014.
- [13] Matteo Cacciari and Gavin P. Salam. Pileup subtraction using jet areas. *Phys. Lett. B*, 659:119–126, 2008.
- [14] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. *Nucl. Instrum. Meth. A*, 389:81–86, 1997.

- [15] Seyon Sivarajah. delphes-gans. <https://github.com/ss2165/delphes-gans>, 2017.
- [16] D.L. Poole, A.P.C.S.D. Poole, R.A. GOEBEL, D. Poole, P.C.S.D.L.C.I.A. Mackworth, A.K. Mackworth, A. Mackworth, R. Goebel, and P.C.S.R. Goebel. *Computational Intelligence: A Logical Approach*. Oxford University Press, 1998.
- [17] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives, 2014.
- [18] Vincent Dumoulin and Francesco Visin. A guide to convolution arithmetic for deep learning, 2018.
- [19] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85 – 117, 2015.
- [20] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [21] MadGraph5_aMC@NLO. Descarga. <https://launchpad.net/mg5amcnlo>, 2009.
- [22] Delphes-Gans. Repositorio. <https://github.com/OlgaFimbresMorales/Delphes-Gans>.
- [23] Red CNN. Repositorio. https://github.com/OlgaFimbresMorales/CNN_JetImages.