

ESTUDIO DEL ETIQUETADO DE JETS
PROVENIENTES DE QUARKS B CON
TÉCNICAS DE APRENDIZAJE AUTOMÁTICO
EN EL EXPERIMENTO CMS DEL LHC

Eloy Ordiz Lera



Universidad de Oviedo

FACULTAD DE CIENCIAS

Tutores:

Javier Fernández Menéndez
Andrea Trapote Fernández

2023

Introducción

Debido a la gran cantidad de datos que se manejan en la actualidad en experimentos de física de partículas, particularmente aquellos que se realizan en grandes detectores, se han desarrollado técnicas para poder procesar los resultados de forma automática sin necesidad de hacerlo manualmente, evitando así posibles errores humanos y requiriendo mucho menos tiempo de trabajo. En este contexto surge el interés por aplicar técnicas de aprendizaje automático que permitan mejorar el rendimiento y la discriminación del proceso.

En el presente trabajo se realiza un estudio en profundidad de los siguientes temas:

1. El Modelo Estándar: Se realiza una breve introducción al modelo, su desarrollo y componentes, así como su confirmación experimental y las preguntas que deja aún por resolver.
2. El Gran Colisionador de Hadrones: La historia de su planteamiento como laboratorio a gran escala, su construcción y la organización de sus experimentos. Dentro de este, se hace especial hincapié en el detector CMS, su ubicación y se explican las coordenadas que se utilizan en este experimento.
3. Etiquetado de partículas: Se define el concepto de jets de partículas en el contexto de las colisiones protón-protón, y se enuncian los distintos criterios a la hora etiquetar un jet dependiendo del tipo de quark o gluón.
4. Aprendizaje Automático y Redes Neuronales: Un breve desarrollo de los fundamentos y la historia del Aprendizaje Automático y la Inteligencia Artificial, seguido de un desarrollo más preciso sobre las redes neuronales, con ejemplos y la aplicación práctica al presente trabajo.

La finalidad de este trabajo consiste en la búsqueda y el etiquetado de quarks b a partir de una desintegración de un gravitón, con una energía de alrededor de 600 GeV, que se desintegran en bosones de Higgs y posteriormente en quarks b . Se trata de una tarea complicada, puesto que los quarks b constituyen una señal muy reducida de sección eficaz

desconocida dentro de un fondo de sección eficaz mucho mayor de naturaleza QCD. Para ello, se buscará diseñar una red neuronal cuyo objetivo será el etiquetado de jets provenientes de esos quarks b , buscando la forma de optimizarla para aumentar su precisión y reducir el tiempo de trabajo.

Índice

Capítulo 1. El Modelo Estándar	5
1.1. Introducción, historia y desarrollo	5
1.2. Componentes e interacciones del Modelo Estándar	6
1.2.1. Fermiones	6
1.2.2. Bosones	8
1.2.3. Interacciones fundamentales	8
1.2.4. Hadronización	13
1.3. Confirmación experimental del Modelo Estándar	15
1.4. Limitaciones y preguntas abiertas	16
Capítulo 2. El Gran Colisionador de Hadrones	18
2.1. Diseño e historia	18
2.2. Funcionamiento del LHC	22
2.3. El detector CMS	23
2.3.1. Coordenadas	25
2.4. Probabilidad de impacto y ruido	29
Capítulo 3. Identificación de jets de partículas	31
3.1. Jets de partículas	31
3.2. Caracterización de los jets	35
3.2.1. Subjets	43
3.3. Generación de los datos: Pythia 8	43
3.4. Agrupación de jets: El algoritmo anti-kT	45
Capítulo 4. Aprendizaje Automático y Redes Neuronales	50
4.1. Fundamentos del Aprendizaje Automático y las Redes Neuronales	50
4.2. Historia y evolución de la inteligencia artificial	51
4.3. Redes neuronales	52

4.4. Desarrollo matemático de una DNN	56
4.4.1. Backpropagation	59
4.5. Ejemplo de una DNN	61
4.6. Diseño de la DNN utilizada	63
Capítulo 5. Presentación de resultados	67
5.1. Elección de variables	67
5.2. Variación de las variables	71
5.3. Variación de la estructura de la red	73
5.4. Variación de la función de activación	74
Conclusión	75
Anexo I. Los datos	77
Anexo II. Elaboración de los histogramas	80
Anexo III. Utilización del código.	83

Capítulo 1

El Modelo Estándar

1.1 Introducción, historia y desarrollo

El Modelo Estándar de partículas (SM, por sus siglas en inglés de Standard Model) es la teoría que pretende clasificar a todas las partículas fundamentales, que son las unidades elementales de materia y energía, según sus características y sus distintas interacciones [2].

Dado que la física siempre ha buscado demostrar leyes que rijan el comportamiento de la naturaleza, el desarrollo de la física de partículas, o física de altas energías, fomentó la creación de un modelo que fuese capaz de agrupar y clasificar a todos los constituyentes de la materia. Fue Demócrito, filósofo griego, el primero en postular que la materia estaba formada por lo que llamó «átomos», que en griego antiguo significa «indivisibles» [3]. Con el desarrollo científico de finales del siglo XIX y principios del siglo XX, se descubrió que en realidad existían otros tipos de partículas. J.J. Thompson descubrió en 1897 el electrón mediante su famoso experimento de los rayos catódicos y postuló posteriormente, en 1904, el Modelo Atómico conocido como el «pastel de pasas» en el que los electrones estaban incrustados en una nube de carga positiva (ver Figura 1). [4].

Este descubrimiento abrió la puerta a intentar buscar nuevos «constituyentes de la materia», descubriéndose así nuevas partículas como el protón [5] o el neutrón, que se consideraron elementales en un principio, pero posteriormente se demostró que estaban a su vez constituidas por otras partículas.

En este contexto, se empieza a entender que además las partículas tienen distintas formas de interactuar, y se postula la existencia de 4 interacciones: fuerte, débil, electromagnética y gravitatoria. En la década de 1950, comienza a desarrollarse la Electrodinámica Cuántica

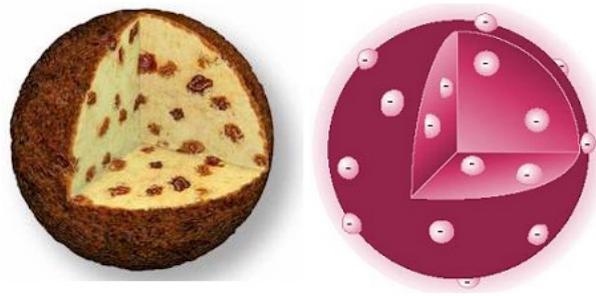


Figura 1: *Modelo Atómico de Thomson asemejado a un «pastel de pasas». Los electrones con carga negativa están incrustados en una bola de materia con carga positiva.*

(QED, por sus siglas en inglés *Quantum ElectroDynamics*), que unifica las interacciones electromagnética y débil. Posteriormente, en la década de 1970, se desarrolla la Cromodinámica Cuántica (QCD, por sus siglas en inglés *Quantum ChromoDynamics*), que explica la interacción fuerte. Estas dos teorías dieron lugar al SM, como explicación unificada de estas interacciones [6]. No obstante, aún queda fuera de este Modelo la interacción gravitatoria, la cual se trata de forma separada debido a la complejidad de unificarla en el SM.

En los próximos capítulos veremos que sucesivos resultados experimentales confirmaron este modelo, pero que a su vez existen limitaciones y preguntas aún sin responder.

1.2 Componentes e interacciones del Modelo Estándar

El SM divide las partículas en dos tipos: bosones, de espín entero, y fermiones, de espín semientero (ver Figura 2) [7].

1.2.1. Fermiones

Son las partículas que componen la materia. Como se comentó, tienen espín semientero, por lo que se rigen por la estadística de Fermi-Dirac. Dentro de estos podemos distinguir dos tipos:

- Leptones: encontramos 3 generaciones de partículas en función de su energía: La primera generación, formada por el electrón e^- (carga -1 y masa de $0.511 \text{ MeV}/c^2$) y el neutrino electrón ν_{e^-} (sin carga y masa aproximada de $10^{-6} \text{ MeV}/c^2$), fue la primera en descubrirse. Posteriormente, se descubrieron la segunda generación, formada por el muón μ^- , con una masa de alrededor de 200 veces la del electrón, con su correspondiente neutrino muón ν_{μ^-} y la tercera generación, formada por el leptón tau τ^- , de masa aproximadamente 3500 veces la del electrón, y su neutrino tau ν_{τ^-} . A pesar de las diferencias en masa, las tres generaciones interactúan de la misma forma: sufren la fuerza electrodébil [7].

El caso de los tres neutrinos es considerablemente especial, puesto que al no tener carga eléctrica, sólo sufren la fuerza débil. Además, al tener una masa tan pequeña, aún no ha sido posible determinarla con precisión, por lo que es una partícula que está en constante estudio.

- Quarks: encontramos también 3 generaciones: una primera generación formada por los quarks up (u) y down (d) con masas en el rango de 1.5 a 3.3 MeV/c^2 para el u y de 4.5 a 5.3 MeV/c^2 para el d¹. Posteriormente, una segunda generación con los quarks charm (c) (masa de 1275 MeV/c^2) y strange (s) (masa de 95 MeV/c^2) y una tercera formada por los quarks top (t) (masa de 174 GeV/c^2) y bottom (b) (masa de 4.18 GeV/c^2). Al igual que ocurría con los leptones, los quarks también sufren la fuerza electrodébil, pero además sufren también la fuerza fuerte, por lo que, en definitiva, se pueden concretar las siguientes diferencias:
 - Su carga eléctrica es fraccionaria, siendo $+2/3$ para los quarks u, c y t y $-1/3$ para los quarks d, s y b.
 - Poseen carga de color, lo que quiere decir que los quarks pueden interactuar con los gluones a través de la fuerza fuerte. En concreto, los quarks tienen un factor de color, asociado a la emisión de un gluón, de $4/3$ [7].

¹Las masas de las partículas son cantidades medidas experimentalmente. Como los quarks no se pueden aislar por el confinamiento QCD, la cantidad dada aquí se calcula indirectamente.

Además de todas estas características, cada fermión tiene su correspondiente antipartícula, idéntica pero con carga de signo opuesto. Así, por ejemplo, igual que encontramos electrones con carga -1 , podemos encontrar su antipartícula, el positrón, con carga $+1$.

1.2.2. Bosones

Son las partículas de spin entero, por lo que siguen la estadística de Bose-Einstein. Podemos también distinguirlas en dos tipos, según si son o no intermediarios de alguna interacción, también llamados mediadores de fuerza.

- Mediadores de fuerza: Tienen spin -1 . Se les denomina como mediadores de fuerza porque permiten explicar las diferentes interacciones como intercambio de bosones. Así, tenemos el fotón γ para la interacción electromagnética (sin masa y sin carga eléctrica), el gluón g para la interacción fuerte (también sin masa y sin carga eléctrica, pero con carga de color, lo que le permite interactuar con los quarks y con otros gluones. Esta es la principal diferencia de la fuerza fuerte que la hace tan particular: el propio mediador tiene carga y le permite interactuar con otros gluones) y los bosones W^+ , W^- (masa de $80.4 \text{ GeV}/c^2$ y carga eléctrica $+1$ y -1 , respectivamente) y Z^0 (masa de $91.2 \text{ GeV}/c^2$ y carga 0), que participan en la interacción débil [7].
- Bosón de Higgs H : Con masa de $125.1 \text{ GeV}/c^2$, sin carga eléctrica ni de color, es el único bosón que no actúa como mediador de fuerza. Tiene spin 0 , y su campo cuántico asociado es el responsable de que las otras partículas fundamentales tengan masa. Fue la última partícula en descubrirse, en el año 2012, lo que supuso, como veremos más adelante, una confirmación del Modelo Estándar (ver apartado 1.3) [8] [9].

1.2.3. Interacciones fundamentales

Como se comentó previamente, existen 4 interacciones fundamentales. En este apartado se explicarán brevemente las principales características de cada una.

- Interacción electromagnética

Modelo estándar de física de partículas

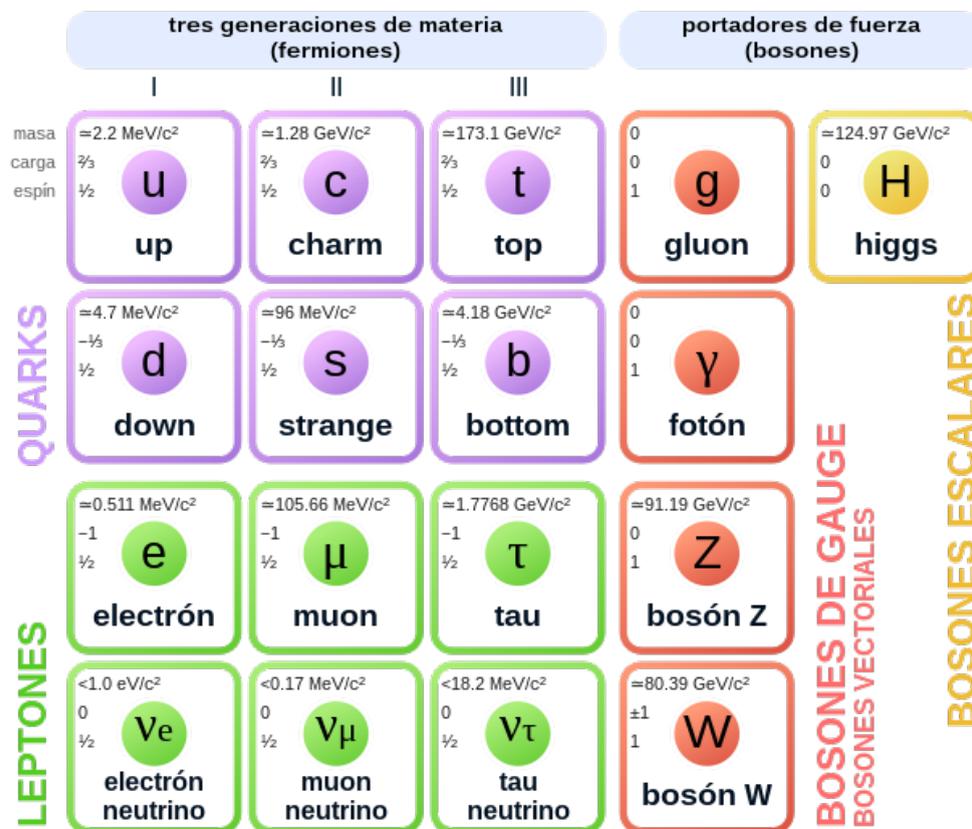


Figura 2: Estructura del Modelo Estándar para la materia. Fermiones a la izquierda y bosones a la derecha. En violeta, los quarks y en verde los leptones, ambos separados en columnas por generaciones. En rojo los bosones mediadores, también conocidos como Bosones de Gauge, y en amarillo el bosón de Higgs. Cada partícula tiene indicados, de arriba a abajo, la masa, la carga y el spin. [1]

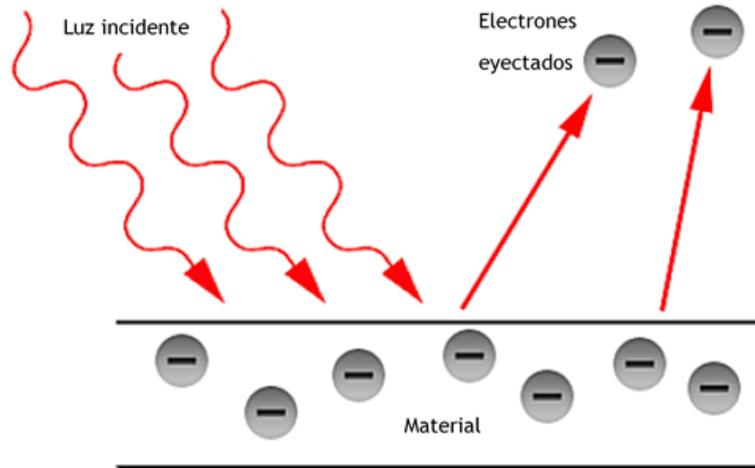


Figura 3: *Los fotones, al incidir sobre un material, produce una excitación de los electrones superficiales, los cuales salen eyectados del material. [10]*

Su fuente es la carga eléctrica, es decir, sólo se dará entre partículas que estén cargadas, por lo que sólo la sufrirán los fermiones, pero excluyendo a los neutrinos. Como se mencionó previamente, su portador es el fotón γ , que participa en interacciones como el efecto fotoeléctrico (ver Figura 3) o el efecto Compton. La carga eléctrica puede ser positiva o negativa. En caso de que dos partículas con la misma carga interactúen, esta interacción será repulsiva, mientras que si lo hacen dos partículas con carga opuesta, la interacción será atractiva. Es una fuerza proporcional al inverso de la distancia, es decir, va como $F \propto \frac{1}{r}$, por lo que su rango es infinito, y se da sea cual sea la distancia. Por último, en cuanto a la intensidad, para la interacción electromagnética esta viene dada por la constante de estructura fina $\alpha_{EM} = \frac{1}{137}$ [7].

- Interacción débil

Se da debido a los bosones mediadores, W^+ , W^- y Z^0 , que transportan la interacción débil entre partículas cargadas (interacciones mediadas por W^+ y W^-) y neutras (interacciones mediadas por Z^0). En concreto, los bosones W^+ y W^- intervienen en procesos como la desintegración β (ver Figura 4), donde hay cambios en el número de protones y neutrones en el núcleo, mientras que el bosón Z^0 interviene en procesos de scattering, como la dispersión neutrón-neutrón o neutrón-electrón. Esta interacción va

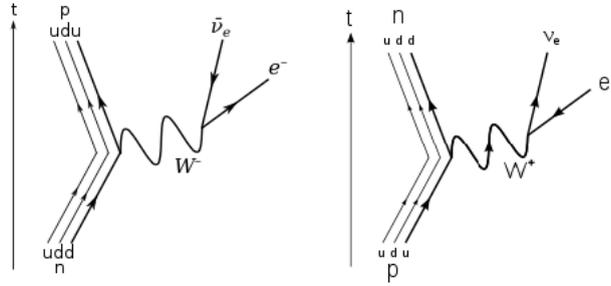


Figura 4: Procesos en los que participan bosones W^+ y W^- , conocidos como desintegraciones beta. En la figura de la izquierda, un quark d de un neutrón se desintegra en un bosón W^- y un quark u , por lo que el neutrón pasa a ser un protón. Esto hace que el átomo en el que se encuentran, para seguir siendo neutro electrónicamente, libere un electrón y un antineutrino en lo que se conoce como desintegración β^- . En la figura de la derecha tenemos el mismo proceso pero inverso: el protón se convierte en neutrón por medio de un bosón W^+ , liberando un positrón y un neutrino. Este proceso se denomina desintegración W^- .

como $F \propto \frac{e^{-m_N r}}{r}$, por lo que cae fuertemente con la distancia. El rango aproximado es de 10^{-18} m, lo que quiere decir que a distancias mayores no se produce interacción. En cuanto a la intensidad, es algo más débil que la fuerza electromagnética, pero su intensidad depende de la energía. A energías del orden de 1 TeV, la fuerza débil alcanza el rango de la fuerza electromagnética, unificándose y entendiéndose así como una única interacción, la fuerza electrodébil [7].

- Interacción fuerte

La fuente de esta interacción es la carga de color, por lo que sólo se dará entre partículas que tengan esta, es decir, quarks y gluones. Como se comentó previamente, además de tener carga de color, el gluón es también la partícula portadora de esta interacción, que actúa, por ejemplo, en procesos de intercambio de color de los quarks (ver Figura 5). La carga de color, puede ser roja, verde o azul, con sus correspondientes anticolores, antirrojo, antiverde y antiazul. Estos colores y anticolores no son más que propiedades matemáticas abstractas que nada tienen que ver con los colores cotidianos. En cuanto a su rango, podríamos pensar que es infinito, ya que viene dado por $V \propto \kappa r$. Sin embargo, debido al confinamiento cuántico, los quarks tienen una libertad asintótica, es decir,

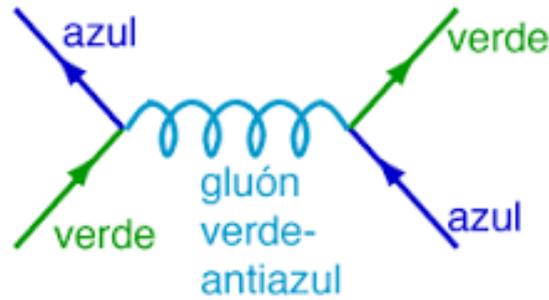


Figura 5: Diagrama Feynmann para una interacción entre quarks generada por un gluón.

[11]

siempre tenderán a interactuar con otras partículas para mantener una carga de color neta nula, que es el estado de menor energía, por lo que su rango de acción se limitará a 10^{-15} m. Por tanto, una partícula con carga de color no podrá viajar nunca como una partícula libre, sino que estará confinada en un estado con carga de color nula, como puede ser un protón, o un neutrón. Esto es lo que se conoce como hadronización, que se explicará en profundidad más adelante. En cuanto a la intensidad de interacción, se tiene un $\alpha_F = 0.1$, bastante superior a todas las demás, pero que decrece al aumentar la energía, debido al confinamiento cuántico [7].

- Interacción gravitatoria

Es la fuerza responsable de la atracción mutua entre todas las masas en el universo. Sin embargo, esta interacción es toda una incógnita a día de hoy, puesto que no se conocen del todo sus propiedades. Se sabe que la interacción va como $F \propto \frac{1}{r}$, por lo que tiene rango de alcance infinito. Sin embargo, su intensidad es muy baja, del orden de $\alpha_G = 10^{-42}$, por lo que suele ser despreciable en experimentos de partículas en los que se trabaja con energías mucho mayores. En cuanto a la partícula portadora de la interacción, se postula la existencia del gravitón como una partícula con propiedades peculiares (sería una partícula con spin 2). Sin embargo, debido a la poca evidencia científica, esta no se incluye en el SM, y generalmente se deja fuera del modelo a la interacción gravitatoria, que ya es descrita como una curvatura del espacio-tiempo causada por la masa y la energía en el marco de la relatividad general [7].

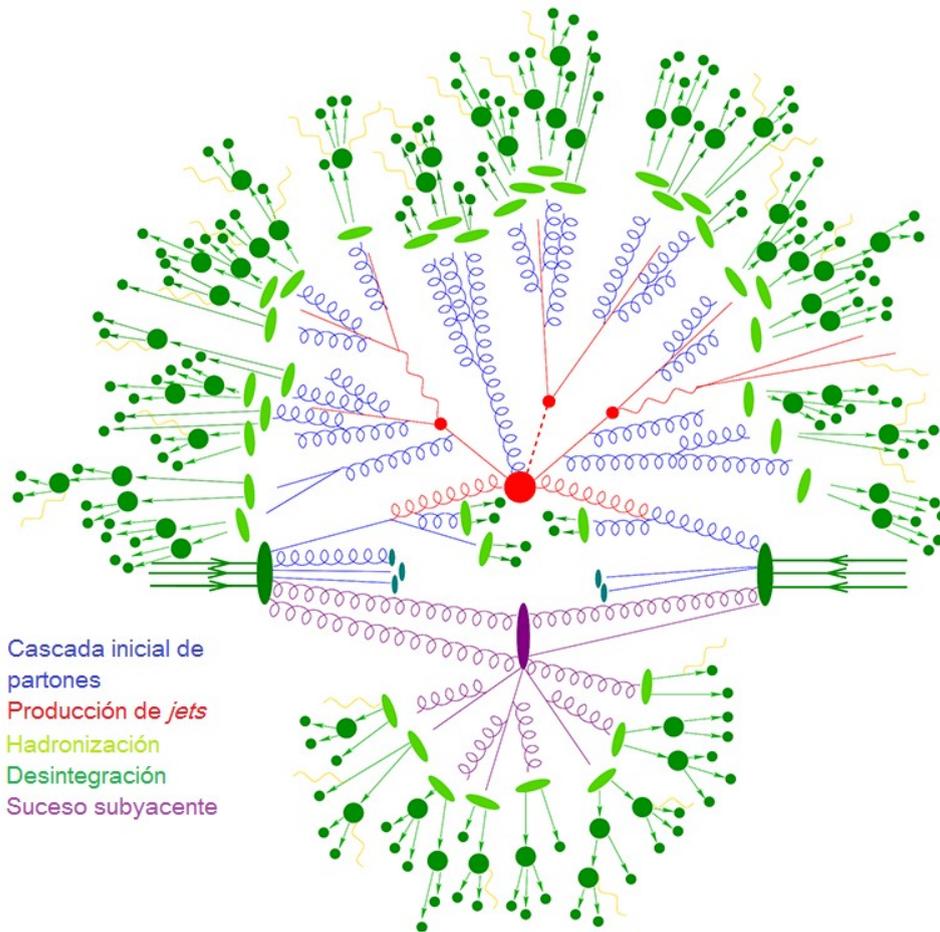


Figura 6: Esquema de una colisión protón-protón, en la que se puede ver cómo, posteriormente a la producción de los jets, las partículas resultantes hadronizan, para después desintegrarse. [12]

1.2.4. Hadronización

La hadronización es la consecuencia directa del confinamiento del color. Para que se cumpla la premisa de que una partícula con carga de color nunca puede estar aislada, estas se agrupan en partículas sin carga de color neta llamadas Hadrones. Estas partículas están formadas por 2 (en el caso de los mesones) o 3 (para el caso de los bariones) quarks de valencia, que son aquellos que caracterizan al hadrón, así como de un mar de gluones y quarks y antiquarks ligeros que se aniquilan constantemente. Este proceso se produce a continuación de una colisión protón-protón (ver Figura 6).

- Mesones: Formados por un quark y un antiquark de valencia en un mar de pares quark-

antiquark. Al estar formados por dos fermiones (de spin semientero), su spin total será entero, por lo que serán bosones. En cuanto a su carga eléctrica, el mesón puede tener carga -1, 0 o 1, dependiendo de la carga de los fermiones que lo constituyan. Algunos ejemplos de mesones incluyen:

- Piones (π^\pm, π^0): Los piones son las partículas más ligeras dentro de los mesones y desempeñan un papel importante en la interacción fuerte y en la mediación de la fuerza nuclear fuerte. Hay tres tipos de piones: π^+ , π^- , y π^0 , cada uno con su respectiva carga eléctrica. Mediante su intercambio, permite la estabilidad de los núcleos atómicos con número similar de protones y neutrones, puesto que se sabe que los neutrones aislados no son estables y decaen por la interacción débil; sin embargo, el intercambio de piones dentro del núcleo entre protones y neutrones (que ocurre de forma muy intensa y frecuente, puesto que los nucleones se encuentran muy cercanos) permite que los neutrones se conviertan en protones y los protones en neutrones, con tanta velocidad que los neutrones no tienen tiempo de sufrir desintegración [7].
 - Kaones (K^\pm, K^0): Los kaones son mesones «extraños» ya que tienen al menos un quark s. Los kaones desempeñan un papel importante en la física de partículas y en la detección de materia oscura, así como en el estudio de la simetría CP (paridad de carga) y su violación [7].
- Bariones: Formados por tres quarks de valencia en un mar de quarks, con posibilidad de tener antibariones, formados por tres antiquarks. Al estar formados por tres fermiones, su spin total será fraccionario, por lo que los bariones serán también fermiones. Su carga será, igual que para los mesones, entera, pero sin estar limitada a -1, 0 o 1, pudiendo ser, por ejemplo, 2. Algunos bariones son:
- Protón (p^+): Es el más conocido, compuesto por dos quarks u y un quark d. Tiene carga eléctrica +1 y una masa de $938 \text{ MeV}/c^2$. Son partículas estables y son esenciales para entender la constitución de los núcleos, junto con los neutrones. Además, su uso tiene aplicaciones en otros campos, como puede ser la radioterapia [7].

- Neutrón (n^0): Junto con los protones, constituyen los núcleos atómicos. Están formados por un quark u y dos quarks d. Son neutros eléctricamente y su masa es muy similar a la del protón, aunque ligeramente superior, de $939 \text{ MeV}/c^2$. Los neutrones, al contrario que los protones, no son estables fuera del núcleo, sufriendo así desintegración beta, como se comentó anteriormente [7].

Es importante destacar que el protón es estable bajo ciertas condiciones, pues su estabilidad es debida a que es la partícula más ligera de su tipo, y no puede decaer en otra más ligera sin violar las leyes de conservación fundamentales. Cuando las condiciones son más extremas, como por ejemplo en las estrellas, los protones pueden experimentar reacciones de fusión nuclear en las que se forman átomos de Helio y se libera mucha energía.

1.3 Confirmación experimental del Modelo Estándar

Son muchos los experimentos que respaldan y corroboran el Modelo Estándar, tanto previos como posteriores al planteamiento del modelo. Algunos de ellos son los siguientes:

- Descubrimiento del electrón (finales del siglo XIX): El descubrimiento del electrón por J.J. Thomson estableció las bases para la comprensión de las partículas cargadas en la estructura atómica [4].
- Modelo de Rutherford (1909): El experimento de dispersión alfa de Rutherford llevó al modelo del átomo con un núcleo cargado positivamente y electrones en órbita, lo que sentó las bases para la noción de partículas subatómicas [5].
- Descubrimiento del neutrón (1932): James Chadwick descubrió el neutrón, que es una partícula neutra presente en el núcleo atómico. Este descubrimiento ayudó a comprender la estructura nuclear [13].
- Experimento de Wilson y Anderson (1933): El experimento en una cámara de niebla llevado a cabo por C.D. Anderson y S.H. Neddermeyer reveló la existencia del positrón,

una antipartícula del electrón [14].

- Descubrimiento del pión (1937): El descubrimiento de los piones, especialmente el pión pi (π^0), proporcionó evidencia de la existencia de partículas subatómicas más allá de protones, neutrones y electrones [15].
- Experimento de Rabi (1937): Isidor Rabi y sus colaboradores desarrollaron la espectroscopia de resonancia magnética nuclear, que permitió estudiar las propiedades magnéticas de núcleos atómicos y confirmar la existencia del espín nuclear [16].
- Confirmación de la teoría electrodébil (1979): El descubrimiento de los bosones W y Z en el Gran Colisionador de Electrones y Positrones en el CERN confirmó la teoría electrodébil propuesta por Sheldon Glashow, Abdus Salam y Steven Weinberg [17].
- Descubrimiento del quark top (1995): El quark top, el sexto y último quark del Modelo Estándar, se descubrió en el Tevatrón en Fermilab [18].
- Descubrimiento del bosón de Higgs (2012): El Gran Colisionador de Hadrones (LHC) en el CERN, en concreto los experimentos ATLAS y CMS, confirmaron la existencia del bosón de Higgs, que es fundamental para proporcionar masa a otras partículas según el Modelo Estándar [8] [9].
- Experimentos en neutrinos (varias décadas): Experimentos de neutrinos, como los realizados en Super-Kamiokande y el Observatorio de Neutrinos de Sudbury, han proporcionado pruebas importantes sobre la oscilación de neutrinos y la masa de los neutrinos.

1.4 Limitaciones y preguntas abiertas

A pesar de las numerosas evidencias científicas que corroboran el Modelo Estándar, a día de hoy existen aún cuestiones sin responder sobre ciertos aspectos. Las más relevantes, y que plantean un mayor reto en la actualidad, son las siguientes:

- Gravedad y relatividad general: El Modelo Estándar no incluye la gravedad. La teoría de la relatividad general de Einstein describe la gravedad a gran escala, pero aún no se ha logrado una unificación exitosa de la relatividad general y el Modelo Estándar

en un marco común, planteándose la posible existencia de una partícula portadora de la interacción gravitatoria: el gravitón.

- **Materia y energía oscura:** El Modelo Estándar no proporciona una explicación para la materia oscura, que constituye aproximadamente el 27 % de la masa total del universo. Aunque hay evidencia observacional de la existencia de la materia oscura, su naturaleza aún es desconocida. Lo mismo ocurre con la energía oscura, que se cree que es la responsable de la expansión acelerada del universo y se cree que constituye hasta un 68 % de la densidad de energía total, pero es una gran desconocida, aún sin cabida en el Modelo Estándar [19].
- **Asimetría materia-antimateria:** Se sabe que existe una asimetría en cuanto a la cantidad de materia y antimateria en el universo, habiendo mucha más proporción de la primera, pero se desconocen las causas. Aunque el modelo predice las violaciones de la simetría CP, estas no parecen ser suficiente para demostrar esta asimetría.
- **Jerarquía de masas:** Las generaciones siempre tienen masas más altas que su generación anterior. Esto provoca que las partículas de generaciones más altas decaigan a partículas de generaciones más bajas, lo que tiene como consecuencia que, en general, sólo se observen en el universo partículas de la primera generación, quedando las de segunda y tercera generación relegadas a sucesos muy energéticos (como lo que se producen en el LHC). Actualmente se desconocen las causas de esta jerarquía, así como la respuesta a la pregunta de por qué sólo hay 3 generaciones y si podría existir una cuarta que no somos capaces de detectar por tener escalas de energía mucho más altas y tiempo de vida medios mucho más cortos.

Estos son sólo algunos ejemplos, y en realidad existen más fenómenos que el modelo no consigue explicar, como puede ser la masa de los neutrinos (que según el modelo, no deberían tener masa, pero recientes experimentos sobre las oscilaciones de neutrinos han dado como resultado que tienen masa), o ciertas inestabilidades físicas y matemáticas que ocurren a altas energías y que sugieren la existencia de una teoría más completa que incluya estos aspectos.

Capítulo 2

El Gran Colisionador de Hadrones

2.1 Diseño e historia

Para alcanzar el objetivo de estudiar más en profundidad la física de partículas y el Modelo Estándar, era necesario un salto cualitativo y cuantitativo a nivel técnico y tecnológico. Por ello, se propuso en la década de 1980 construir un colisionador de hadrones a gran escala. El proyecto fue concebido y se aprobó en 1984, comenzándose la construcción en 1988 por parte de la Organización Europea para la Investigación Nuclear (CERN, por sus siglas en francés) y con la cooperación internacional de más de 100 países. El LHC es un anillo subterráneo de 27 kilómetros de circunferencia que se encuentra a unos 100 metros bajo tierra. Gracias a este tamaño masivo, se alcanzan energías de colisión que no se pueden conseguir de otra forma. Está ubicado en la frontera entre Francia y Suiza, cerca de la ciudad de Ginebra y tiene la mayor parte del acelerador en territorio francés (ver Figura 7) [20].

Además del LHC, el CERN consta de un sistema de aceleradores encadenados, inyectando cada uno de ellos el haz de partículas en el siguiente y haciéndolo de forma que su energía va aumentando de forma progresiva, llegando este finalmente al Gran Colisionador de Hadrones. La mayor parte de estos aceleradores previos tienen sus propios experimentos asociados para los que se necesita una energía inferior que la alcanzada al final de la cadena. El procedimiento en cadena para los protones sigue el siguiente orden (ver Figura 8):

- Iones de Hidrógeno H^- parten del acelerador lineal LINAC 4, donde adquieren una energía de hasta 160 MeV. A continuación, se inyectan esos átomos al Booster, proceso durante el cual se pierden los dos electrones, por lo que a partir de aquí los protones quedan libres.
- Los protones adquieren una energía de hasta 2 GeV en el Booster.



Figura 7: Ubicación del túnel del LHC en la frontera franco-suiza. En la imagen se muestra también la ubicación del Super Sincrotrón de Protones (SPS), uno de los lugares donde se aceleran los protones antes de inyectarlos en el LHC. [22]

- A continuación pasan al Sincrotrón de Protones (PS), alcanzando energías de hasta 26 GeV. Este acelerador es también usado para acelerar iones de plomo.
- Con esta energía, se inyectan los protones al Súper Sincrotrón de Protones (SPS) donde obtienen energías de 450 GeV. En el SPS se realizan otros muchos experimentos de energías menores.
- Por último, los protones pasan al LHC, donde se dividen en dos haces que giran en sentidos opuestos. Estos haces finales alcanzan cada uno una energía de 7 TeV, lo que nos da como resultado una energía total en el centro de masas de 13.6 TeV [21].

En total, son 9 los experimentos que se llevan a cabo en el CERN, 4 de ellos de un mayor tamaño y con sus propias galerías excavadas bajo tierra (CMS, ATLAS, ALICE y LHCb) y 5 de menor tamaño, que aprovechan las colisiones de los otros experimentos (LHCf, TOTEM, MoEDAL, FASER y SND@LHC). A continuación se explicará brevemente la finalidad de cada uno de ellos, dejando de lado el detector CMS, que se verá más en detalle en el punto 2.3 de este trabajo, puesto que es una pieza fundamental para el desarrollo del mismo [23].

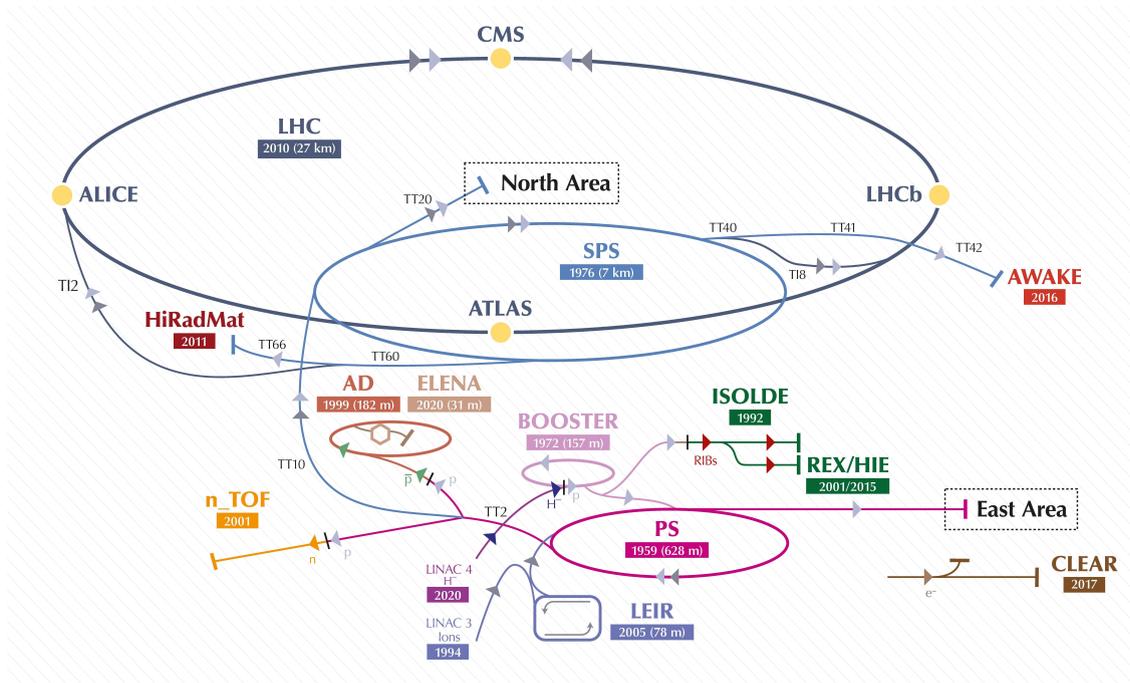


Figura 8: Conjunto de aceleradores encadenados del CERN.

- ATLAS (A Toroidal LHC ApparatuS): Es un detector de propósito general en el cual, el campo magnético se genera mediante un sistema de forma toroidal alrededor de la tubería central de unos 44 metros de largo y 25 metros de alto. Es el detector de mayor tamaño que se ha construido hasta la fecha. Entre algunos de sus experimentos, destacan la búsqueda de partículas supersimétricas, la medición de propiedades generales de partículas subatómicas y la exploración de fenómenos en la física de partículas más allá del Modelo Estándar.
- ALICE (A Large Ion Collider Experiment): Está especializado en analizar colisiones de iones pesados (generalmente de plomo) a energías muy altas, en las cuales se produce un plasma de quarks y gluones, con condiciones similares a la que se cree que existían en las primeras fracciones de segundo después del Big Bang. El detector es una estructura cilíndrica masiva de aproximadamente 16 metros de longitud, 16 metros de diámetro y un peso de alrededor de 10.000 toneladas.
- LHCb (Large Hadron Collider beauty): Se encarga del estudio de la asimetría entre materia y antimateria a través de quarks pesados, principalmente a través del quark b y su antipartícula. Las diferencias entre dicho quark y su correspondiente antiquark

proporcionan información sobre las violaciones en la simetría CP que podrían ayudar a comprender mejor el universo. El detector LHCb tiene un diseño en forma de espectrógrafo de un solo brazo con una longitud de unos 20 metros, que se enfoca en el plano de producción de partículas.

- LHCf (Large Hadron Collider forward): Su objetivo principal es estudiar las partículas producidas en las colisiones de alta energía que ocurren en el LHC, particularmente en la región frontal. Estas partículas incluyen fotones y neutrones, y su estudio es importante para entender la interacción de partículas de alta energía con la atmósfera de la Tierra y para mejorar los modelos de simulación de partículas en colisiones de alta energía. Se encuentra situado cerca del detector ATLAS y está formado por dos pequeños detectores de tipo Cherenkov. Además, el LHCf también ha participado en estudios relacionados con rayos cósmicos con el objetivo de comprender la formación y propagación de estos en la atmósfera.
- MoEDAL (Monopole and Exotics Detector At the LHC): Busca partículas altamente ionizantes y exóticas, como los monopolos magnéticos, que son partículas teóricas con una única carga magnética.
- TOTEM (TOTal cross section, Elastic scattering and diffraction dissociation Measurement): Estudia las propiedades de las colisiones de protones a energías extremadamente altas, especialmente en lo que se refiere a la sección eficaz total, la dispersión elástica y la disociación por difracción. Se ubica en torno al detector CMS, extendiéndose varios cientos de metros alrededor de él.
- FASER (ForwArd Search ExpeRiment): Se centra en la búsqueda de partículas exóticas y materia oscura. Su diseño permite detectar partículas que interactúan muy débilmente, como los neutrinos. Este tipo de partículas han sido propuestas por alguna teoría más allá del Modelo Estándar para explicar nociones como la materia oscura (aunque actualmente sólo se baraja la posibilidad de que esto ocurra en alguna forma de condensación [24]), la asimetría entre materia y antimateria. Se encuentra junto al experimento ATLAS.
- SND@LHC (Scattering and Neutrino Detector at the LHC): Se basa en el estudio de

los neutrinos producidos en las colisiones. Recientemente, en agosto de 2023, detectó por primera vez neutrinos, en conjunto con el experimento FASER. También se localiza en las inmediaciones de ATLAS.

2.2 Funcionamiento del LHC

Para poder acelerar los protones hasta las elevadas energías con las que se trabaja, se utilizan una serie de técnicas experimentales basadas en generar una diferencia de potencial que los acelere. Esta aceleración no puede ser arbitraria ni aleatoria, pues esto haría que los protones adquiriesen velocidades mayores o menores de las deseadas, o que algunos protones de los haces se desfasaran con respecto a los demás. Para ello, se aplica el potencial con una frecuencia adecuada para que una vez alcanzada la energía deseada, los protones que estén correctamente situados no sufrirán más aceleración, mientras que aquellos que se encuentren desfasados serán acelerados o frenados para que se mantengan unidos al grupo.

Además, el acelerador consta de electroimanes superconductores a lo largo de su superficie, capaces de generar campos de hasta 8T, que tienen como objetivo curvar la trayectoria de los haces y mantenerlos colimados, de forma que aumente la probabilidad de colisión, así como inyectar los haces en los detectores, y frenarlos y conducirlos fuera del LHC una vez ha concluido el experimento. También encontramos potentes imanes en los propios detectores, que tienen como finalidad medir el momento y el signo de la carga de las partículas cargadas una vez se han producido las colisiones. Esto se verá en mayor detalle en el punto 2.3, de forma específica para el detector CMS [21].

Para asegurar una utilización correcta del costoso aparataje experimental, se ha diseñado un calendario de uso (ver Figura 9) basado en dos tipos de períodos, conocidos como Run (periodos de funcionamiento) y Long Shutdown (periodos largos de parada). En los periodos de Run, además de la toma de datos, se incluyen la puesta a punto y el ajuste de los diversos parámetros, así como posibles paradas para resolver fallos técnicos. En los periodos de Long Shutdown, se busca realizar mejoras técnicas en todos los sistemas, con el fin de mejorar la calidad de los detectores o aumentar la energía que pueden adquirir los haces. En la fecha

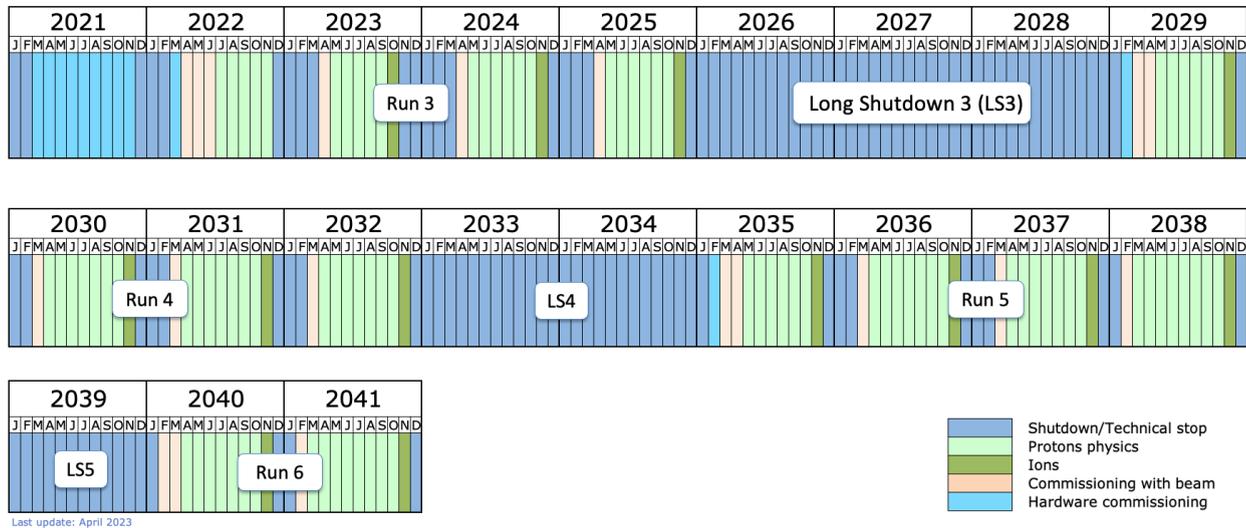


Figura 9: *Calendario de actuación a largo plazo del LHC. [25]*

de desarrollo de este trabajo, el LHC se encuentra en el ecuador del Run 3.

La planificación actual del LHC, plantea su funcionamiento hasta el año 2041. Se espera que al terminar el año 2025, den comienzo las operaciones de Acelerador de Hadrones de Alta Luminosidad (HL-LHC, por sus siglas en inglés), que tendrían como objetivo aumentar la luminosidad del acelerador, esto es, el número de colisiones por unidad de tiempo. Para ello, se requiere de la mejora de algunos de los componentes principales, como los sistemas de enfriamiento de los electroimanes. Esta mejora aumentaría la probabilidad de observar sucesos que ocurren con muy poca frecuencia a partir del año 2029.

2.3 El detector CMS

El Compact Muon Solenoid (CMS) es el detector del LHC en cuyos datos se basan este trabajo. Se trata, al igual que ATLAS, de un detector de carácter general, destacando a nivel técnico por tener en su interior el solenoide más grande de su tipo jamás construido, capaz de crear un campo magnético muy potente, así como por su compacidad, teniendo una longitud de unos 28 metros y una altura de unos 15 metros, pero con un peso de 14 millones de kilogramos [26].

Este detector cuenta con unas características que le permiten identificar desde fotones hasta muones, pasando por muchas otras partículas que podrán diferenciarse en función de sus características. Para ello, el detector consta de una serie de capas concéntricas (ver Figura 14):

- Detector de trazas (Tracker): Fabricado a base de silicio, se encarga de reconstruir las trayectorias de las partículas tras la colisión. La elección del material no es aleatoria, puesto que, debido a la cercanía que este detector tiene con el punto de colisión, a muy altas energías, debe ser muy resistente. Este detector está construido en varias capas, la primera de ellas hecha a base de píxeles de entre 3 y 16 cm de radio, y las siguientes a base de tiras de mayor longitud. Al interaccionar las partículas con los detectores, se producen señales eléctricas que posteriormente pueden ser reconstruidas en la trayectoria de la partícula.
- Calorímetro electromagnético (ECAL): Se utiliza para medir la energía de electrones y fotones, además de provocar que estas partículas se detengan. Las partículas impactan contra cristales de $PbWO_4$, que tienen una alta resistencia a la radiación, y el destello que se produce en el impacto es medido por fotodetectores situados en los cristales que lo transforman en una señal eléctrica que permite identificar la partícula según su intensidad.
- Calorímetro hadrónico (HCAL): Permite detectar los hadrones de una manera similar al ECAL, aunque con capas alternadas de latón (que absorbe los impactos) y de plástico centellador (que emite luz)
- Solenoide superconductor: Permite medir el momento de las partículas cargadas, así como el signo de su carga. Es capaz de generar un campo electromagnético de 3.8T.
- Detectores de muones: Se ubican en el exterior el detector, intercalados con láminas de hierro, de forma que aíslan el campo magnético generado por el solenoide en el interior del detector. El detector CMS cuenta con cuatro tipos de detectores: tubos de deriva (DT), cámaras de tiras catódicas (CSC), cámaras de placas resistivas (RPC) y multiplicadores gaseosos de electrones (GEM). Su funcionamiento se basa en la ioni-

zación del gas que contienen por el paso de los muones, de manera que los electrones arrancados se recogen en un ánodo dando lugar a una corriente eléctrica.

Cabe destacar que los neutrinos son invisibles a todos estos detectores, debido a sus características (no tienen carga y su masa es muy pequeña), pero se pueden detectar gracias a la conservación de la energía transversa (ver 2.3.1 Coordenadas).

2.3.1. Coordenadas

Para caracterizar un jet, necesitaremos 3 coordenadas espaciales, puesto que trabajamos con colisiones reconstruidas y medidas en 3 dimensiones. Uno podría pensar en utilizar coordenadas cilíndricas, ya que el detector CMS tiene forma de cilindro. Este es un buen punto de partida, pero es necesario hacer algunas transformaciones que faciliten el trabajo.

Siendo los ejes xyz los representados en la Figura 12, con el x apuntando al centro del acelerador, el y hacia el cielo y el z tangente al acelerador, tendremos el ángulo azimutal ϕ como el ángulo formado en el plano xy medido desde el eje x y que puede tomar valores $\in [-\pi, \pi)$.

Tendremos también el ángulo polar θ , medido desde el eje z hasta el y , tomando valores $\in [-\pi/2, \pi/2]$ y que en este caso, se transformará en la coordenada definida como $\eta = -\ln \tan \theta/2$, llamada pseudorrapidez. La dependencia de esta coordenada con respecto al ángulo polar θ se muestra en las Figuras 10 y 11. Así, la pseudorrapidez es invariante de Lorentz en los rangos relativistas de energía en los que se trabaja en el LHC, por lo que se facilita el trabajo bajo transformaciones. Además, la producción de partículas es aproximadamente constante respecto a la pseudorrapidez [35], lo cual también facilita los cálculos en los datos experimentales. Este observable, toma valores entre $-\infty$ y ∞ , siendo creciente cuando θ decrece. No obstante, la pseudorrapidez $|\eta| < 2.4$, debido a que los subdetectores de CMS no pueden cubrir ángulos menores de 10° ni mayores de 170° , ya que esas son las direcciones que toman inicialmente los haces protones. Además, se debe tener en cuenta que para partículas cargadas en CMS, el

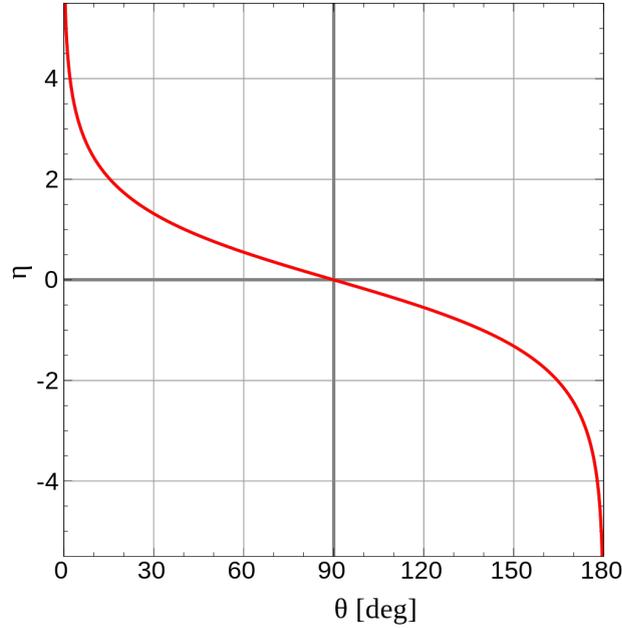


Figura 10: Dependencia de la pseudorrapidez η en función de la coordenada θ .

HCAL mide valores $|\eta| < 3.1$, por las dimensiones del Tracker.

La tercera coordenada vendrá definida a partir del momento. El momento \vec{p} de una partícula se define como $\vec{p} = m\vec{v}$, siendo \vec{v} la velocidad de la partícula y $m = m_0\gamma$ la masa relativista, con m_0 la masa en reposo de la partícula y $\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$. El módulo del momento viene dado por $|\vec{p}| = \sqrt{p_x^2 + p_y^2 + p_z^2}$. Podemos derivar de aquí el momento transverso, definido como $p_T = \sqrt{p_x^2 + p_y^2}$, coordenada utilizada en el CMS por ser también invariante de Lorentz.

Además, la energía que los calorímetros detectan es la energía transversa $E_T = E \sin \theta$, dependiente del momento transverso (Figura 13). Usando la conservación de la energía, como en el eje y la energía inicial es 0 antes de la colisión, pues el haz está girando en el LHC en el plano xz , también deberá serlo tras la colisión. Así, si la suma de las energías en el eje y después de la colisión es muy distinta de 0, querrá decir que hay algunas partículas que no se han detectado. Calculando estas diferencias, se tiene que estas partículas, por sus características energéticas, son los neutrinos.

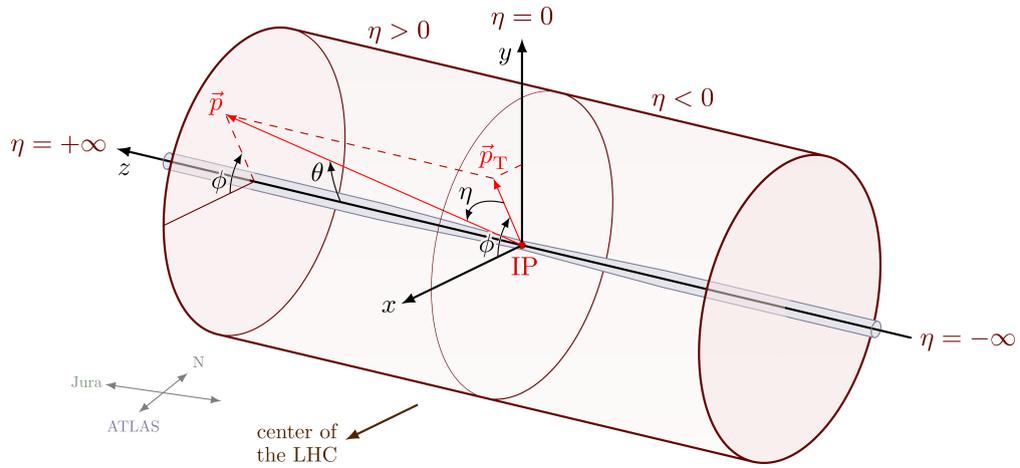


Figura 11: Representación de los distintos valores que puede tomar la coordenada η en una vista externa. [27]

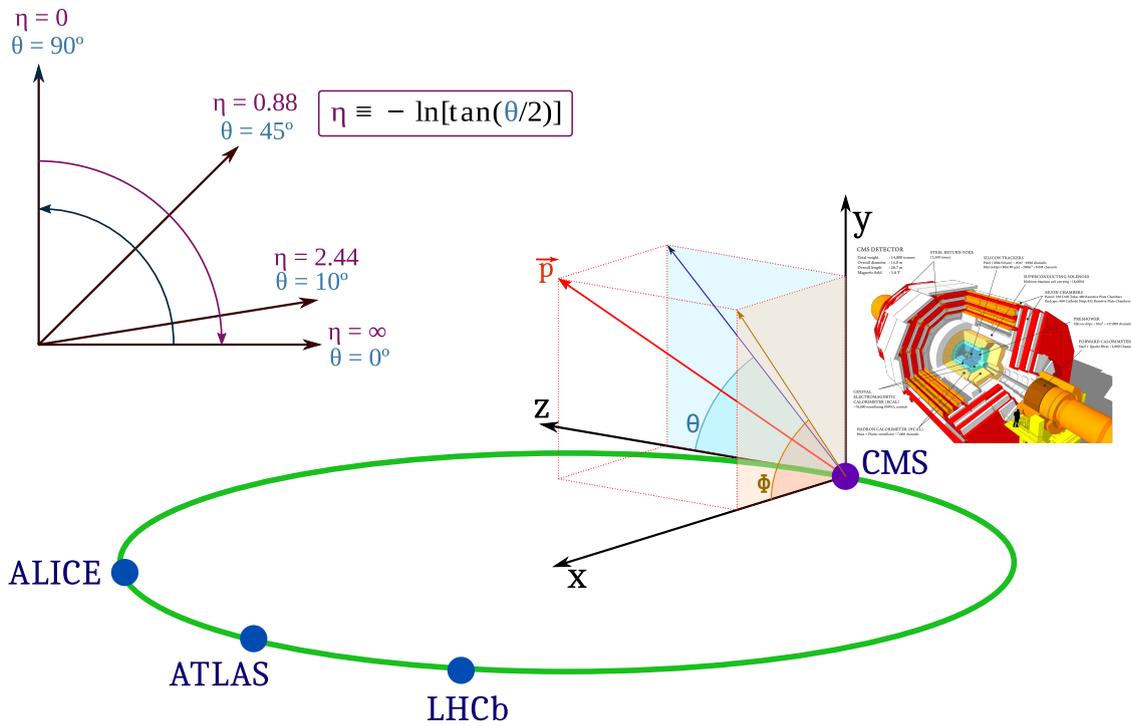


Figura 12: Coordenadas utilizadas en el detector CMS. [36]

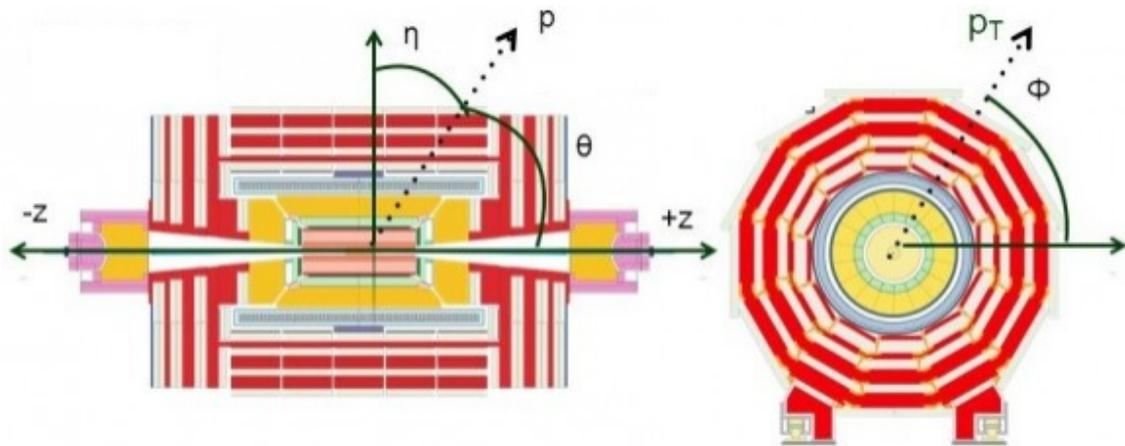


Figura 13: Orientación de las distintas coordenadas en el detector CMS, en sus vistas lateral y frontal. [37]

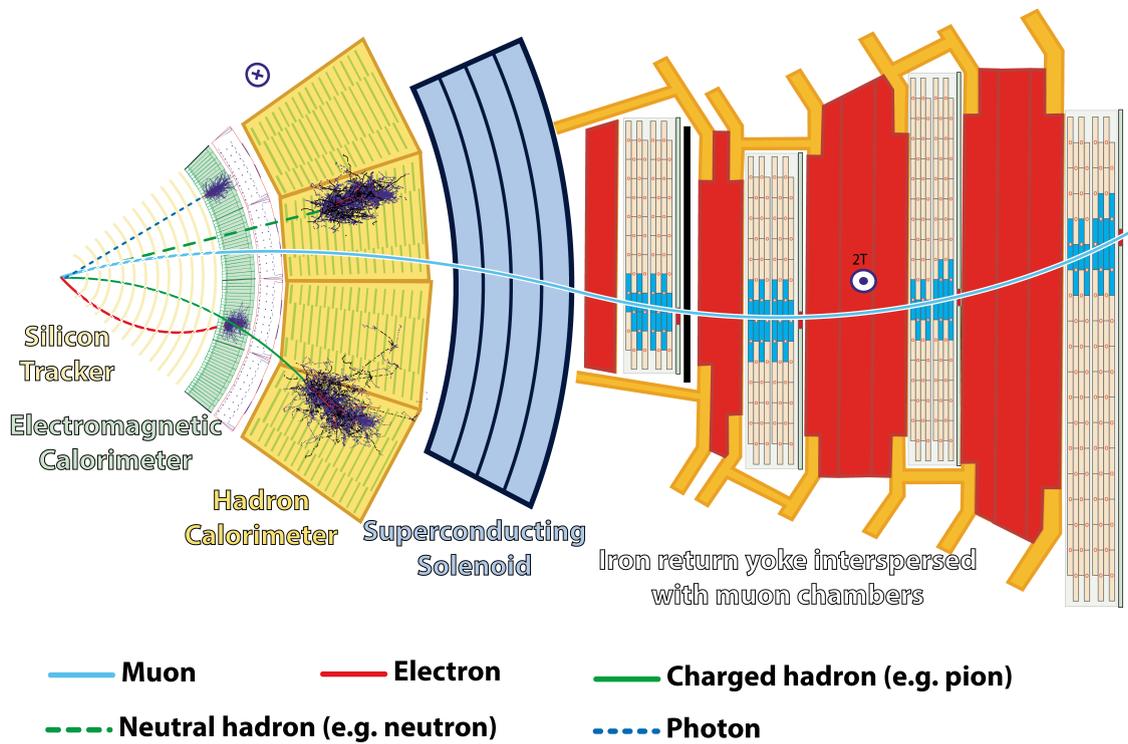


Figura 14: Trayectoria de las distintas partículas a través del detector CMS [28].

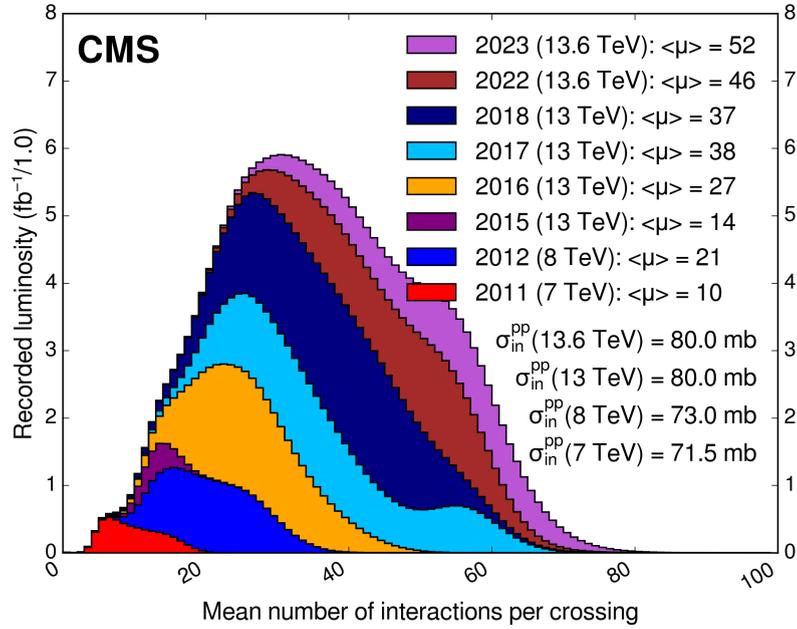


Figura 15: *Número de vértices en la interacción para los años 2011-2023. Podemos ver que a menor luminosidad instantánea en las interacciones, menor número de vértices se tiene.*

[29]

2.4 Probabilidad de impacto y ruido

Debido al alto coste energético que supone acelerar los protones hasta esas velocidades, conviene asegurar una alta probabilidad de colisión. Para esto, en lugar de colisionar un único protón contra otro (lo cual sería extremadamente difícil), se aceleran varios paquetes de protones a la vez. Esto tiene como consecuencia directa el apilamiento (pile up), que no son más que las colisiones protón-protón que ocurren de forma prácticamente simultánea en el mismo punto. Este apilamiento depende de la luminosidad instantánea ², de forma que a mayor luminosidad, mayor es el número promedio de colisiones simultáneas (ver Figura 15).

Estos impactos no son del todo simultáneos, sino que tienen diferencias del orden de 10^{-10} s. Por su parte, el detector CMS es capaz de tomar datos con una frecuencia de 40MHz,

²La luminosidad instantánea de un haz se define como el número de partículas por unidad de superficie y de tiempo de dicho haz. De esta forma, la luminosidad es el inverso de la sección eficaz, que se define como una probabilidad (en unidades de superficie) de que ocurra una interacción entre dos partículas.

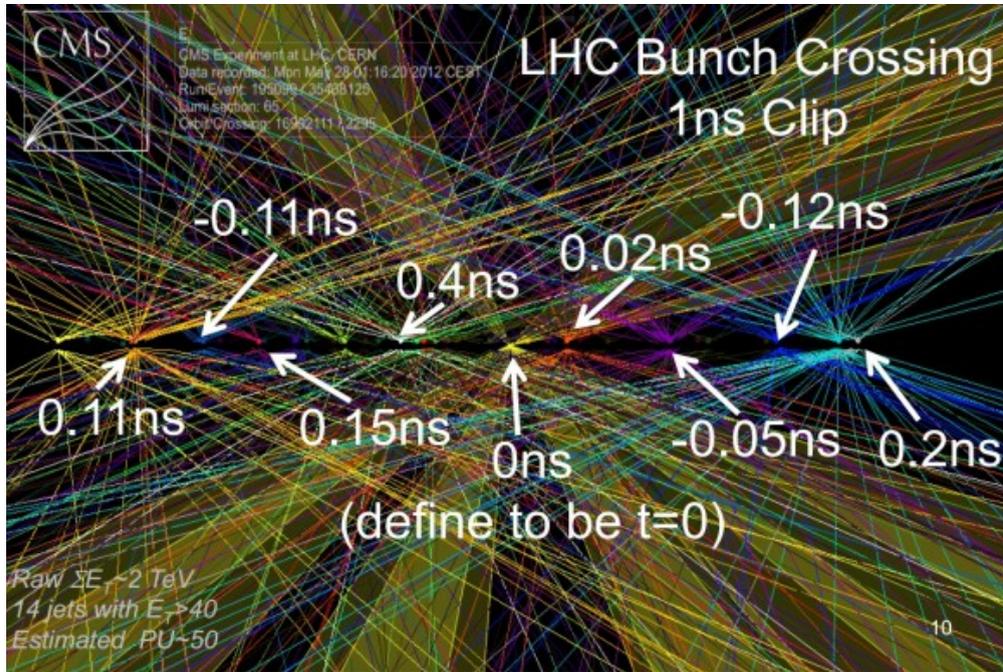


Figura 16: Apilado de colisiones en un ancho de unos 10 centímetros, en el cual se ha tomado como sistema de referencia (0ns) el vértice central. El resto de vértices difieren en un margen del orden de 0.1 ns. [30]

ya que el cruce de haces se produce cada 25ns, por lo que muchos de estos impactos no pueden resolverse por separado (ver Figura 16).

Sin embargo, la colisión de varios paquetes de protones entre sí tiene otra desventaja, que es la enorme salida de datos, del orden de $4 \cdot 10^7$ colisiones por segundo, como se comentó previamente. La mayoría de estos datos no tienen interés o son redundantes, por lo que para reducirlo y seleccionarlos se utiliza un sistema de trigger. Este sistema consiste en el descarte de aquellas colisiones que no cumplen ciertas características, como pueden ser un umbral mínimo de energía o un número de trazas determinado. Para llevarlo a cabo, se hace uso tanto del hardware del detector, como del software.

Capítulo 3

Identificación de jets de partículas

3.1 Jets de partículas

Se conoce como jet de partículas al «chorro» de partículas que se origina, normalmente, tras la hadronización de un quark (ver Figura 17). Se trata de grupos de partículas con una larga vida media que se producen cuando un quark o un gluón se hadronizan. Estas partículas constituyentes son hadrones cargados (en torno al 65 % del total), fotones (alrededor del 25 %) y hadrones neutros (aproximadamente un 10 % del total). Las propiedades de un jet suelen venir determinadas por sus constituyentes, como veremos más adelante. Además, debido a que los hadrones que forman los jets decaen en partículas más estables, se forman dentro del jet subestructuras conocidas como «subjets» (ver apartado 3.2.1).

Dada la corta vida media del quark top, este no puede ser observado directamente, por lo que se podrán detectar sólo los productos de su desintegración. Además, su corta vida media es también menor que la escala de tiempo en la que actúa la fuerza fuerte, lo que tiene como consecuencia que este quark no sufra hadronización. Por lo tanto, nunca se podrán detectar en el LHC jets de quarks top (jets t), pero estas partículas podrán identificarse a partir de los productos de su desintegración.

Se sabe que el 96 % de las veces, el quark top se desintegra en un quark b y un bosón W por la interacción débil. Siendo posible identificar los productos de esta desintegración, en especial el quark b, será posible caracterizar al quark t. Esto es de gran relevancia, pues en el LHC se producen una gran cantidad de quarks top, y este, al ser tan pesado, es el quark que más se acopla al bosón de Higgs, por lo que proporcionará también información relevante sobre el mismo.

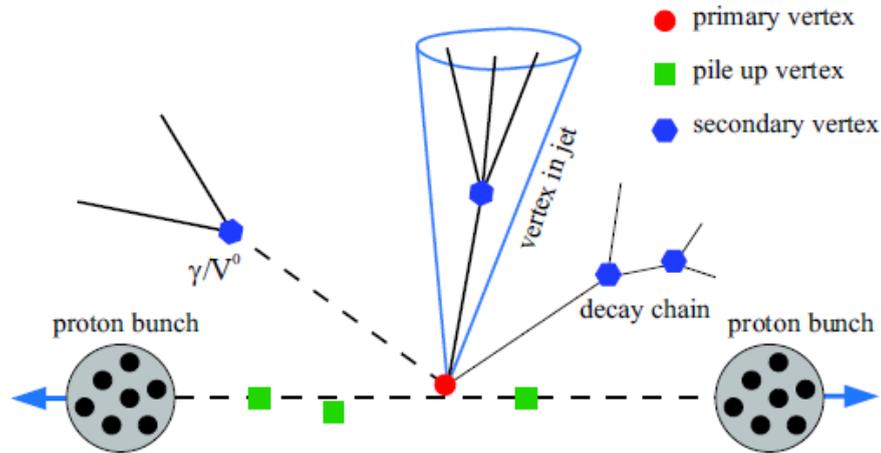


Figura 17: Diagrama que representa la formación de los jets en las colisiones protón-protón.

[31]

En cuanto al resto de quarks, sí sufren hadronización, pero dependiendo del caso tendremos que los hadrones se desintegrarán antes o más tarde, según su vida media. En el caso de los quarks b y c, esto ocurre antes en el detector, dando lugar a un vértice secundario, como se aprecia en la Figura 17, y por tanto, tendrán propiedades que permitirán identificarlos. En estos casos, se formarán jets de partículas con unas características muy concretas que se explicarán en el punto 3.2. Además, tenemos cuatro partículas que hadronizan, pero la vida media de sus hadrones es mayor, por lo que llegarán a los calorímetros antes de haber decaído. Estos son los quarks up, down y strange, así como el gluón, que como comentamos previamente, también interactúa a partir de la fuerza fuerte. Al tener estas partículas una mayor vida media, serán necesarios otros procedimientos para intentar caracterizarlos.

En concreto, el proceso que se estudia en este trabajo consiste en la desintegración de un gravitón de Kaluza-Klein³, creado tras la colisión pp, en dos bosones de Higgs (ver Figura 18). A continuación, los bosones se desintegran en dos quarks b cada uno, dando lugar a una producción final total de 4 quarks b. Esta producción de quarks puede ocurrir de forma que los quarks estén resueltos, acelerados o semi-resueltos, como se puede ver en la Figura

³La teoría de Kaluza-Klein es una propuesta teórica formulada en 1920, que pretende unificar la gravedad y el electromagnetismo en una sola teoría. Para ello se proponen dimensiones adicionales, además de las 3 espaciales y la temporal.

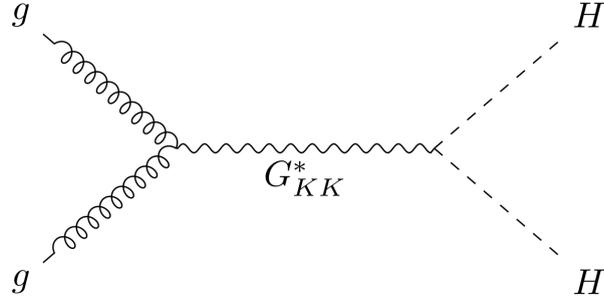


Figura 18: *Diagrama de Feynmann de la creación de un gravitón de Kaluza-Klein y su posterior desintegración en dos bosones de Higgs [32].*

19. Los quarks estarán resueltos en caso de que la energía de producción no sea muy alta, mientras que no estarán resueltos si la energía es alta.

La energía mínima para que se formen los dos bosones de Higgs es $2M_H = 250\text{GeV}$. Por lo tanto, si la energía de producción no es mucho mayor, los bosones producidos serán poco energéticos y por tanto, pronto decaerán en quarks b muy poco colimados, que como consecuencia podrán ser resueltos fácilmente. En este caso, se detectarán 4 jets AK4 (ver definición en apartado 3.4), uno para cada quark.

Por el contrario, si la energía es mucho mayor que $2M_H$, los productos de sus desintegraciones estarán muy colimados, dando lugar a dos jets AK8 en lugar de los 4 jets AK4 del caso anterior. Estos sucesos se conocen como sucesos acelerados.

El último caso es el caso intermedio entre los dos mencionados previamente. Sólo uno de los bosones de Higgs tiene una alta energía, por lo que sólo ese bosón se desintegrará formando un jet AK8, mientras que el otro dará lugar a dos jets AK4.

Además, tendremos una señal de fondo que serán los procesos QCD, es decir, aquellos en los que no interviene ningún bosón mediador, y que tienen como productos finales quarks y gluones. Estos procesos tienen una sección eficaz muy alta, por lo que no es sencillo diferenciar la señal que queremos estudiar (producción y desintegración de los bosones de Higgs)

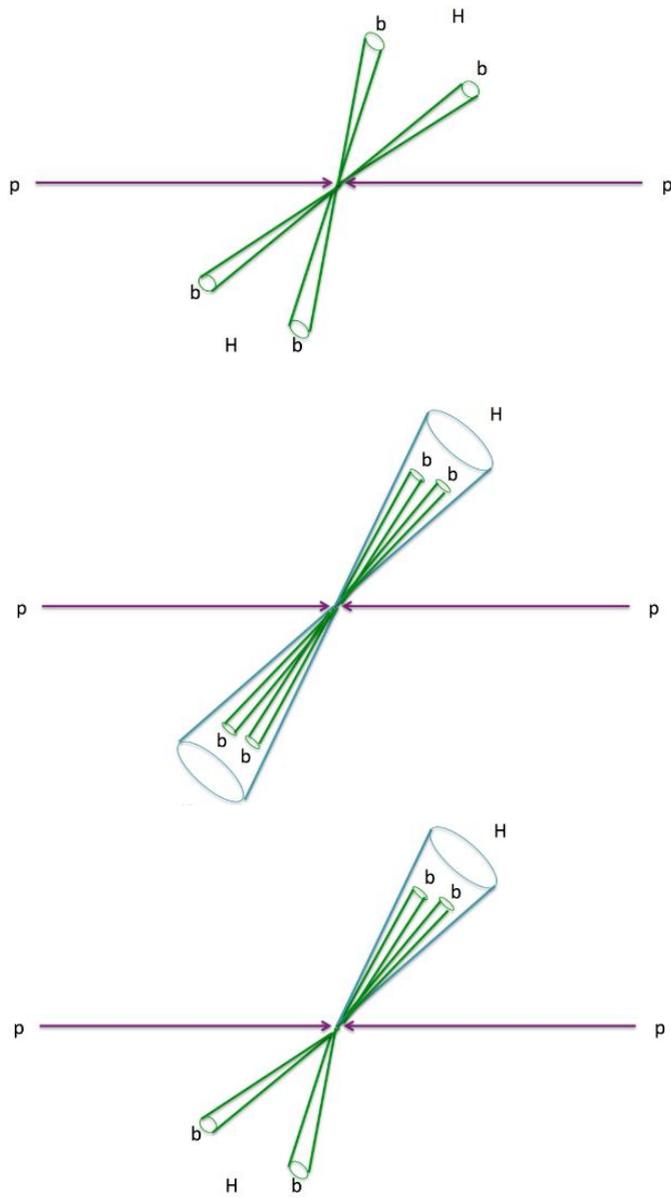


Figura 19: *Distintos diagramas de producción de quarks b . Los quarks b producidos en la colisión pueden ser resueltos (imagen superior), acelerados (imagen intermedia) o semi-resueltos (imagen inferior). [33]*

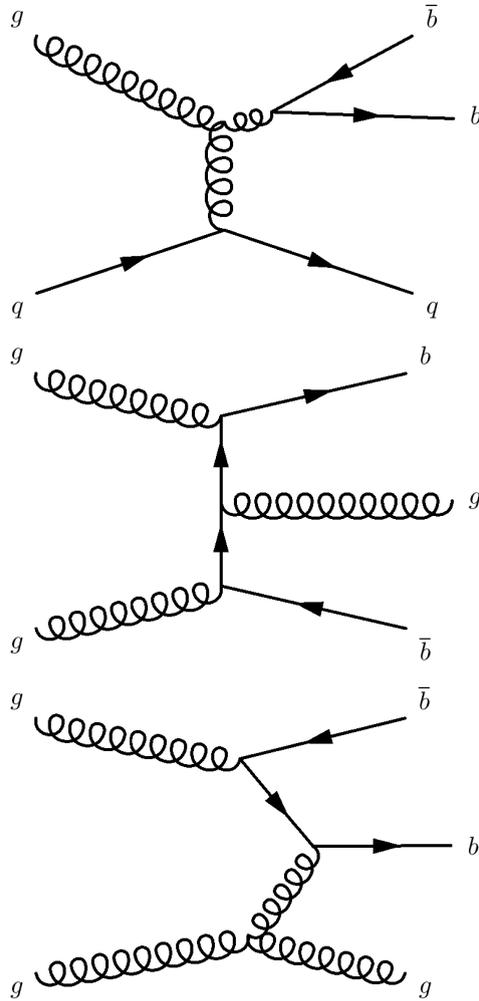


Figura 20: *Diagramas de Feynmann para distintos procesos QCD de la señal de fondo [34]*

de la señal de fondo (QCD). Algunos de estos procesos se pueden ver en la Figura 20.

En los próximos apartados se explicarán los observables utilizados para estudiar los jets, y la forma de caracterizar un jet de cada tipo.

3.2 Caracterización de los jets

Las técnicas de identificación de jets b hacen uso de los observables relacionados con las propiedades de los hadrones presentes en dichos jets. Aquí estudiaremos cuales son estas propiedades que diferencian a los jets b del resto.

Como hemos visto, al contrario que los quarks t , los quarks b sí hadronizan en un vértice primario, pero debido a la vida media de los hadrones que contienen un quark b (alrededor de 1.5 ps, dependiendo del hadrón), estos recorrerán una cierta distancia hasta desintegrarse en lo que se conoce como vértice secundario, que no es más que el punto donde se agrupan las trazas surgidas de la desintegración. Esto es algo que sólo ocurre con jets pesados (sea de quarks b o c) por lo que, pudiendo detectar estos vértices secundarios sabremos que nos encontramos ante un jet b o c (ver Figura 22).

El vértice secundario es caracterizado por lo que se conoce como parámetro de impacto, el cual se define como la distancia entre el vértice primario y las trazas desplazadas en el punto de acercamiento máximo. Este parámetro de impacto, con su correspondiente vector que apunta desde el vértice primario hasta el punto más cercano a las trazas, puede venir dado en 3 dimensiones, como su proyección en el plano transversal xy (IP_{xy}) o como su proyección longitudinal en el eje z (IP_z), el cual puede ser positivo o negativo, dependiendo de si las trazas provienen de un SV (predominantemente positivo) o no (simétricamente positivo o negativo) [38]. Este signo viene determinado por el producto escalar de la dirección del eje del jet y el vector resultante entre el PV y el punto de máximo acercamiento (DCA) (ver Figura 21). Así, para las trazas que provienen de un SV, el ángulo θ será relativamente pequeño (por lo menos, menor de $\pi/2$), por lo que IP será positivo, pero para trazas que no provienen de un SV será aleatoriamente positivo o negativo.

Otro observable útil para caracterizar un jet b , es su momento transversal, puesto que, debido a su alta masa, tendrá un p_T mayor que los jets de otros sabores. En este caso, un criterio que se puede utilizar para detectar jets es seleccionar sólo trazas que cumplan $p_T > 1$ GeV, con una eficiencia del 98 %.

Además, aproximadamente el 20 % de desintegraciones de hadrones que contienen un quark b , tienen como producto un muón o un electrón, por lo que la presencia de estos leptones poco energéticos (dentro de un cierto radio ΔR) también será un criterio para seleccionar posibles candidatos [42].

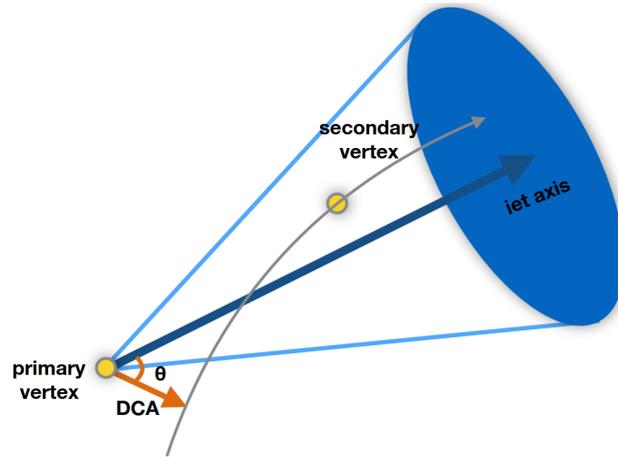


Figura 21: Representación esquemática de los parámetros que influyen en el signo de IP . Para $\theta < \pi/2$, el producto escalar tendrá signo positivo y por tanto IP será positivo. Por el contrario, si $\theta > \pi/2$, será negativo.

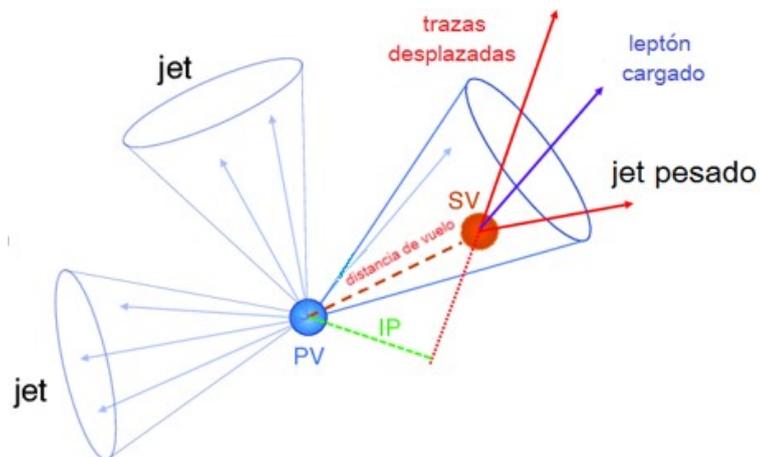


Figura 22: Caracterización de los parámetros relevantes del jet: Vértice Principal (PV), Parámetro de Impacto (IP), Vértice Secundario (SV) y distancia de vuelo. [39]

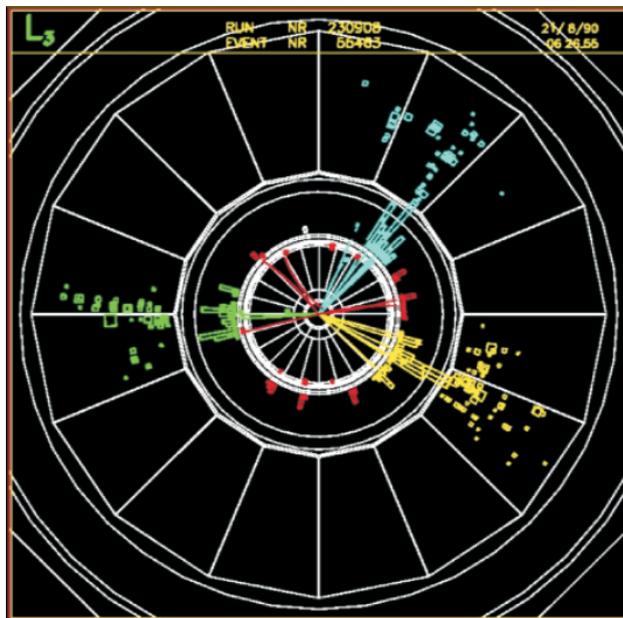


Figura 23: Suceso de 3 jets, registrado en el detector L3 en el año 1990. Confirma la existencia del gluon, la partícula mediadora de la fuerza de color. Un electrón y un positrón chocan a muy alta energía, creando un par quark-antiquark, en colores azul y amarillo. Uno de ellos radia un gluon, que es la partícula que da lugar al tercer jet, en color verde. [52]

Por su parte, la detección experimental inicial que respalda la existencia del gluón se basa en la observación de sucesos que implican tres chorros de partículas, como el representado en la Figura 23. Estos sucesos se registraron durante las colisiones entre electrones y positrones en el acelerador PETRA de DESY en Hamburgo en 1979. La interpretación principal de estos sucesos sugiere la generación de un gluón altamente energético que se manifiesta junto al par quark-antiquark resultante de la aniquilación de los electrones y positrones originales que colisionan [40]. La detección de estos pudo llevarse a cabo debido a que los jets de gluones tienen algunas propiedades que permiten diferenciarlos de los quarks ligeros (u, d y s).

Una de estas propiedades es la multiplicidad [41]. Se define la multiplicidad como el número de partículas (cargadas o neutras) candidatas a estar presentes en el jet [42]. Se ha observado una relación entre las multiplicidades de los jets de gluones y los jets de quarks ligeros de $\frac{n_{gluon}}{n_{quark}} = 1.25 \pm 0.5$ cuando se usa el «algoritmo k_t » [43] y de $\frac{n_{gluon}}{n_{quark}} = 1.10 \pm 0.4$

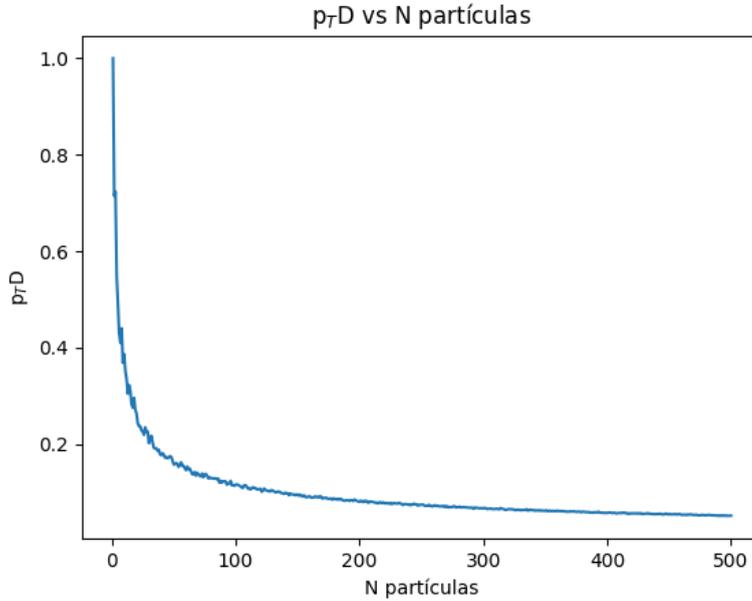


Figura 24: Gráfico que representa el valor de la función de fragmentación con respecto al número de partículas en las que se fragmenta un jet, con $1\text{GeV} < p_T < 100\text{GeV}$.

cuando se usa el algoritmo de «cono iterativo» [44], con una apertura de 30° [45].

Otra propiedad relevante a la hora de clasificar jets de gluones y quarks ligeros es su función de fragmentación. Esta es una función que describe la producción de partículas en el estado final del jet. Se trata de una relación (normalizada) que nos dice cómo de fragmentado está el jet. Se define como:

$$p_T D = \frac{\sqrt{\sum_i p_{T,i}^2}}{\sum_i p_{T,i}} \quad (1)$$

Esta función está definida de forma que para jets que se fragmentan en una sola partícula, $p_T D \rightarrow 1$, y dicha partícula llevaría todo el momento transversal del jet, y, a su vez, $p_T D \rightarrow 0$ para jets formados por infinitas partículas, como se puede ver en la Figura 24. La observación experimental nos dice que la función de fragmentación de los gluones es más suave, teniendo así otra forma de clasificar este tipo de jets [42].

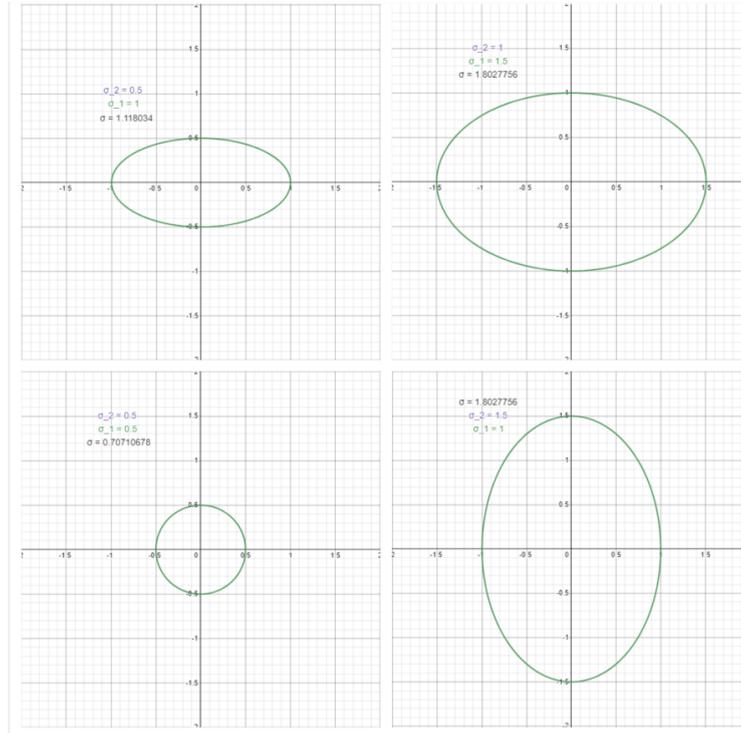


Figura 25: Representaciones de posibles cortes transversales según la forma del jet. Se puede ver que cuanto más colimado está el jet, menor es el valor de σ .

Además es importante tener en consideración el tamaño del jet. Aunque siempre podemos considerar un jet como una estructura cónica, no se trata de un cono perfecto, por lo que al realizar un corte transversal, en lugar de un círculo tendremos una elipse. Esta elipse tendrá sus dos ejes, mayor (σ_1) y menor (σ_2), a partir de los cuales se puede obtener un eje promedio, tomando la media cuadrática de ambos $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$. Experimentalmente se ha observado que los jets provenientes de quarks están más colimados, es decir, su amplitud cónica σ será más estrecha, mientras que los jets gluónicos serán más anchos (ver Figura 25). Por tanto, comparando estos parámetros será más sencillo determinar si se trata de un jet originado por un quark ligero o un gluón [42].

También es importante definir la «N-subjettiness», otra magnitud relativa a la forma del jet. Este observable se define, bajo la suposición de que un jet tiene N subjets (ver apartado 3.2.1), como:

$$\tau_N = \frac{\sum_k p_{T,k} \min(\Delta R_{1,k}, \dots, \Delta R_{N,k})}{\sum_k p_{T,k} R} \quad (2)$$

donde ΔR es la distancia entre cada partícula del jet y el eje de su subjet más cercano y R es un parámetro prefijado del «algoritmo *anti-k_t*» [39] (ver apartado 3.4). Este observable es una medida para valorar si se han definido todos los subjets que ese jet tiene. Si τ_N tiene un valor bajo, querrá decir que todas las partículas producto del jet tienen asignado un subjet lo suficientemente cerca. En caso de que τ_N tenga un valor alto, tendremos que algunas partículas están «demasiado lejos» de su subjet más cercano y por tanto quedarán por definir uno o más subjets que no se habían definido previamente (ver Figura 26).

Por último, se define el observable «z ratio». Este observable se define como:

$$z = \Delta R(SV_0, SV_1) \frac{p_T(SV_1)}{m(SV_0, SV_1)} \quad (3)$$

donde SV_0 y SV_1 son los vértices secundarios con menor incertidumbre en la distancia de vuelo, asociados a los dos ejes τ (0 y 1), $\Delta R(SV_0, SV_1)$ es la distancia entre los vértices, y $m(SV_0, SV_1)$ es la masa invariante correspondiente al cuadrimomento total de los dos vértices sumados. Este observable nos da información relativa a un posible sistema de dos vértices secundarios [39].

No obstante, se ha observado que tanto la primera relación como el resto de propiedades dependen del rango de energía en el que se trabaje [46], por lo que es importante tener en cuenta este factor a la hora de realizar un estudio.

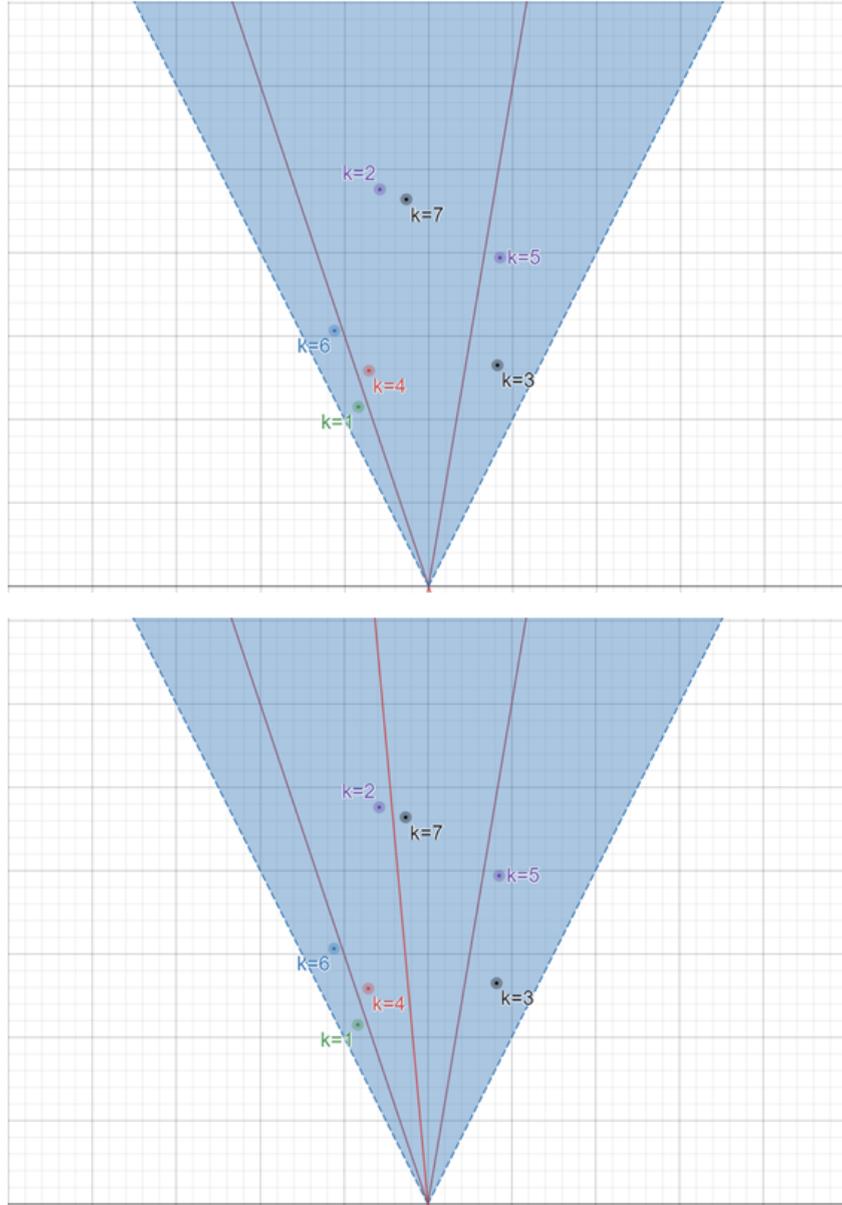


Figura 26: Esquema simplificado de la posible estructura de un jet simple (región azul) formado por 7 partículas. En un primer lugar, se definen dos sujetos, representados por sus ejes (líneas rojas). Sin embargo, esta configuración da un τ_N muy elevado, ya que las partículas $k=2$ y $k=7$ no tienen un sujet realmente cerca, por lo que el valor de $\min(\Delta R_{1,k}, \dots, \Delta R_{N,k})$ será alto para esas partículas. Sin embargo, si se define un posible tercer sujet, como se muestra en la figura de debajo, el valor de τ_N es mucho menor, ya que las partículas $k=2$, $k=7$ tendrán ahora un sujet mucho más cerca. Esto es un ejemplo muy simplificado de lo que ocurre en el jet pero sirve para entender la importancia del observable descrito.

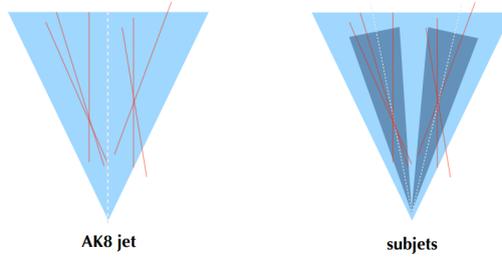


Figura 27: Representación de un jet AK8 (ver apartado 3.4) y sus posibles subjets.

3.2.1. Subjets

Es necesario además definir qué es un subjet. Se utiliza la palabra subjet para definir a las subestructuras del jet formadas tras el decaimiento de algunos de los productos de dicho jet principal. Es decir, algunas partículas pesadas que se han producido tras la formación del jet, decaen en otras partículas más ligeras, formándose subjets (ver Figura 27).

Para resolver estos subjets se han desarrollado métodos como el algoritmo de «soft-drop». El funcionamiento de estos algoritmos suele implicar reconstruir los jets usando los algoritmos generales (expuestos en el apartado 3.4) y deshacer el último paso del proceso. De esta forma se define un eje para el subjet alrededor del cual se estructura su forma cónica.

3.3 Generación de los datos: Pythia 8

Trabajar con simulaciones permite prever el comportamiento de los datos a la vez que se reproducen de forma fiable los datos experimentales, considerando la teoría física y modelización correctas. Por ello, es necesario tener la capacidad de generar y simular datos de la misma naturaleza que los que allí se dan. Para ello, se utilizan métodos como el método de Monte Carlo, basado en algoritmos computacionales de muestreo numérico aleatorio. Mientras que, como se comentó previamente, los datos recogidos en el detector deben pasar un proceso de filtrado (apartado 2.4), para posteriormente reconstruir los sucesos de colisión en los archivos para que puedan ser analizados con el software necesario, en el caso de los datos generados este proceso es algo más complicado. Esto se puede ver en el diagrama de

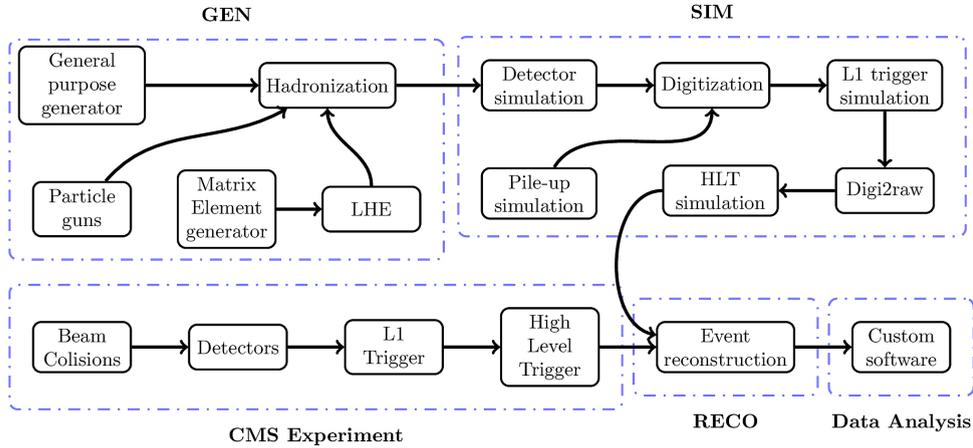


Figura 28: Diagrama que representa la generación de datos frente a la obtención directa de datos del detector [47].

la Figura 28.

Podremos distinguir dos etapas, una de generación (GEN) y otra de simulación (SIM). Dentro de la etapa de generación, se requiere del uso de un generador de elementos de matriz y otro de carácter general sucesivamente, el primero para los partones y el segundo para la hadronización. Además, se puede incluir el uso de un simulador más específico, como un simulador de lanzamiento de partículas.

A continuación viene la etapa de simulación, donde se simula el paso de las partículas por los detectores, simulando también los efectos que se dan en los detectores y en los elementos electrónicos, además de los triggers o apilamientos que puedan darse. Esta simulación se realiza con GEANT4 [48]. Con todo esto, se pasa a la etapa de reconstrucción, al igual que se hace con los datos experimentales, para poder ser posteriormente tratados y analizados con el software específico.

Los datos en este caso han sido producidos a partir del programa generador de sucesos «Pythia 8» [49]. Este programa simula colisiones de alta energía y genera las posibles partículas que se podrían producir como producto de las mismas usando técnicas de Monte Carlo. Este programa se divide en tres niveles:

- El nivel del proceso: representa el proceso de dispersión entre partículas, incluyendo las posibles resonancias que se puedan producir. Normalmente se describe como una perturbación del estado inicial.
- El nivel de partones: incluye los estados final e inicial de radiación de partículas, teniendo en cuenta los productos de los jets y la posibilidad de que ocurra un cambio en la carga de color de los quarks.
- El nivel de hadrones: representa el proceso de hadronización por medio del confinamiento QCD (apartado 1.2.4). Por lo general se trata de un modelo no perturbativo, por lo que requiere el ajuste de ciertos parámetros.

Este proceso permite la generación de un suceso simulado de forma realista como el que se puede obtener de un detector.

Además de estos 3 niveles, existen una serie de parámetros iniciales que deben ser configurados, así como elementos que permiten a los distintos niveles transferirse la información de uno a otro, así como almacenar información y poder acceder a ella. Además, también consta de una interfaz (LHE) para la lectura y escritura de sucesos. Toda esta red se representa en la Figura 29.

3.4 Agrupación de jets: El algoritmo anti-kT

Existen numerosos algoritmos diseñados con el fin de agrupar jets, como pueden ser los algoritmos de «Seeded Cone», el «algoritmo k_t » (mencionados previamente) el «algoritmo de Cambridge-Aachen» o el «algoritmo anti- k_t ». Este último es el que utilizaremos en el presente trabajo.

En general, se define la distancia entre dos partículas (i,j) como:

$$d_{ij} = \min(p_{Ti}^{2k}, p_{Tj}^{2k}) \frac{\Delta_{ij}^2}{R^2} \quad (4)$$

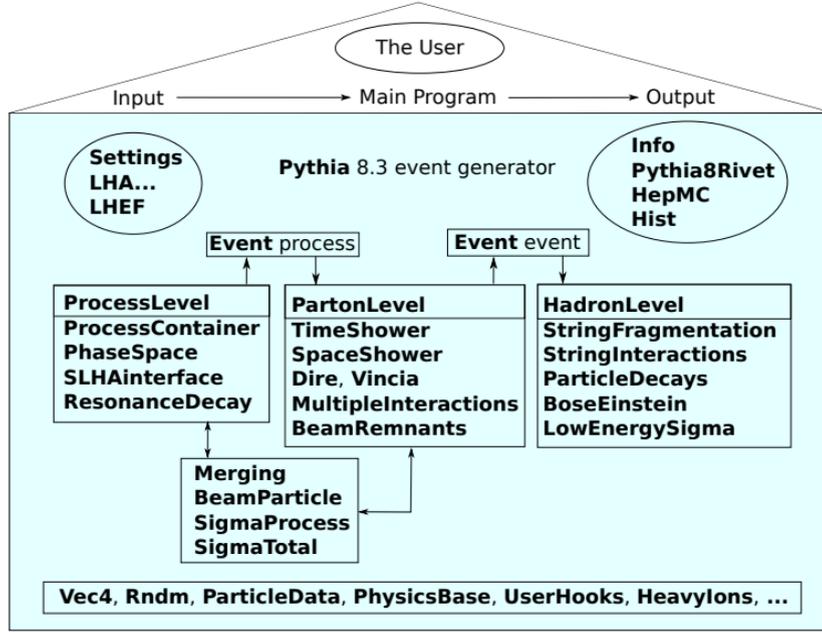


Figura 29: Esquema simplificado de la estructura del generador de sucesos Pythia 8, diferenciando los tres niveles, la información transferida entre ellos y algunos de los parámetros relevantes. [49]

donde $\Delta_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$, p_{Ti} es el momento transversal, η_i la pseudorapidez y ϕ_i el ángulo azimutal de la partícula. Además, se tienen dos parámetros: R es un parámetro equivalente a un «radio de cercanía» entre partículas, y k es un parámetro, cuya variación nos da los distintos algoritmos de agrupación de jets [50]. Además, también se define la distancia entre el haz (B) y la partícula como:

$$d_{Bi} = p_{Ti}^{2p} \quad (5)$$

Según el valor que se le otorgue al parámetro k , tendremos un método u otro. Para $k = 1$, se tiene el algoritmo k_t , y en general, para $k > 1$, el comportamiento obtenido es similar. Para $k = 0$ se tiene un caso especial, que se corresponde con el algoritmo de Cambridge-Aachen. El caso del algoritmo anti- k_t se da para $k = -1$, que a su vez, dará resultados similares para cualquier $k < 0$, pero conviene trabajar con $k = -1$ por simplicidad [50]. Una comparación entre estos algoritmos puede verse en la Figura 30

Para explicar el funcionamiento del algoritmo, se definen los conceptos de partícula «suave» (soft particle en inglés) y partícula «fuerte» (hard particle) como aquellas que tienen, respectivamente, un bajo o un alto momento transverso. Así, dependiendo del algoritmo que se esté utilizando, estas partículas suaves y fuertes se agruparán de distintas formas. Supondremos por tanto un suceso que tenga unas pocas partículas fuertes bien separadas, con momentos transversos $p_{t,1}, p_{t,2}, \dots$ y varias partículas suaves con momentos transversos $p_{t,i}, p_{t,j}, \dots$, como se ve en el ejemplo de la Figura 31

Según el algoritmo anti- k_t ($k=-1$), la ecuación 4 para la distancia entre una partícula fuerte y otra suave queda:

$$d_{1i} = \min\left(\frac{1}{p_{T1}^2}, \frac{1}{p_{Ti}^2}\right) \frac{\Delta_{1i}^2}{R^2} \quad (6)$$

Así, la distancia vendrá únicamente determinada por el momento transverso de la partícula fuerte y el correspondiente Δ_{1i} . En cambio, la distancia entre dos partículas suaves será mucho mayor, puesto que sus momentos transversos son menores y por tanto sus inversos serán mucho mayores. De esta forma, las partículas suaves se agrupan en torno a las partículas fuertes que tengan cerca, en lugar de agruparse con otras partículas suaves. Esta agrupación, además, no hace variar la forma del jet en caso de que una partícula suave se agrupe con otra fuerte.

En caso de que haya dos partículas fuertes en un radio de $R < \Delta_{12} < 2R$, entonces se generan dos jets, que no podrán ser perfectamente cónicos. Si el momento transverso de la partícula 1 es mucho mayor que el de la partícula 2, entonces el jet de 1 será perfectamente cónico, mientras que el de 2 será parcialmente cónico, perdiendo la parte que se solapa con el otro jet. Si por el contrario tienen momentos transversos muy similares, entonces ambos serán parcialmente cónicos.

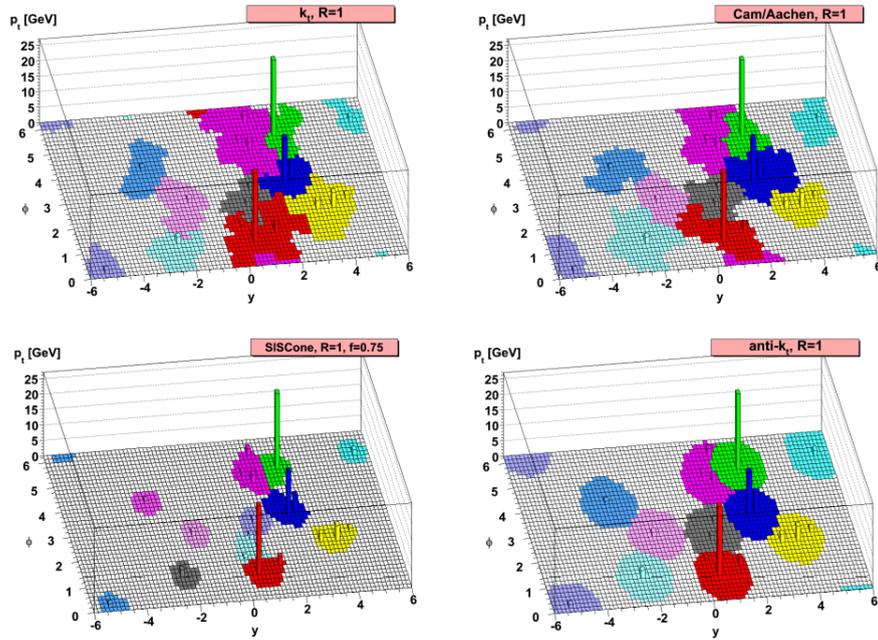


Figura 30: Comparación entre los resultados de algunos de los diferentes algoritmos de agrupación de jets para un mismo suceso, donde se muestran las distintas áreas de cada jet [50].

En el caso de que las partículas fuertes se encuentren en $\Delta_{12} < R$, ambas partículas fuertes formarán un único jet, centrado en el partícula que tenga un momento transverso mucho mayor, en caso de que la haya, o con una forma un poco más compleja (unión de conos) en el caso de que tengan momentos transversos parecidos.

Por tanto, la principal característica de este algoritmo es que permite la variación de forma del jet únicamente cuando se involucran dos o más partículas fuertes, pero esta forma no se modifica al incluir partículas suaves.

En el presente trabajo se utiliza el algoritmo anti- k_t con $R = 0.8$, produciendo jets conocidos como jets AK8.

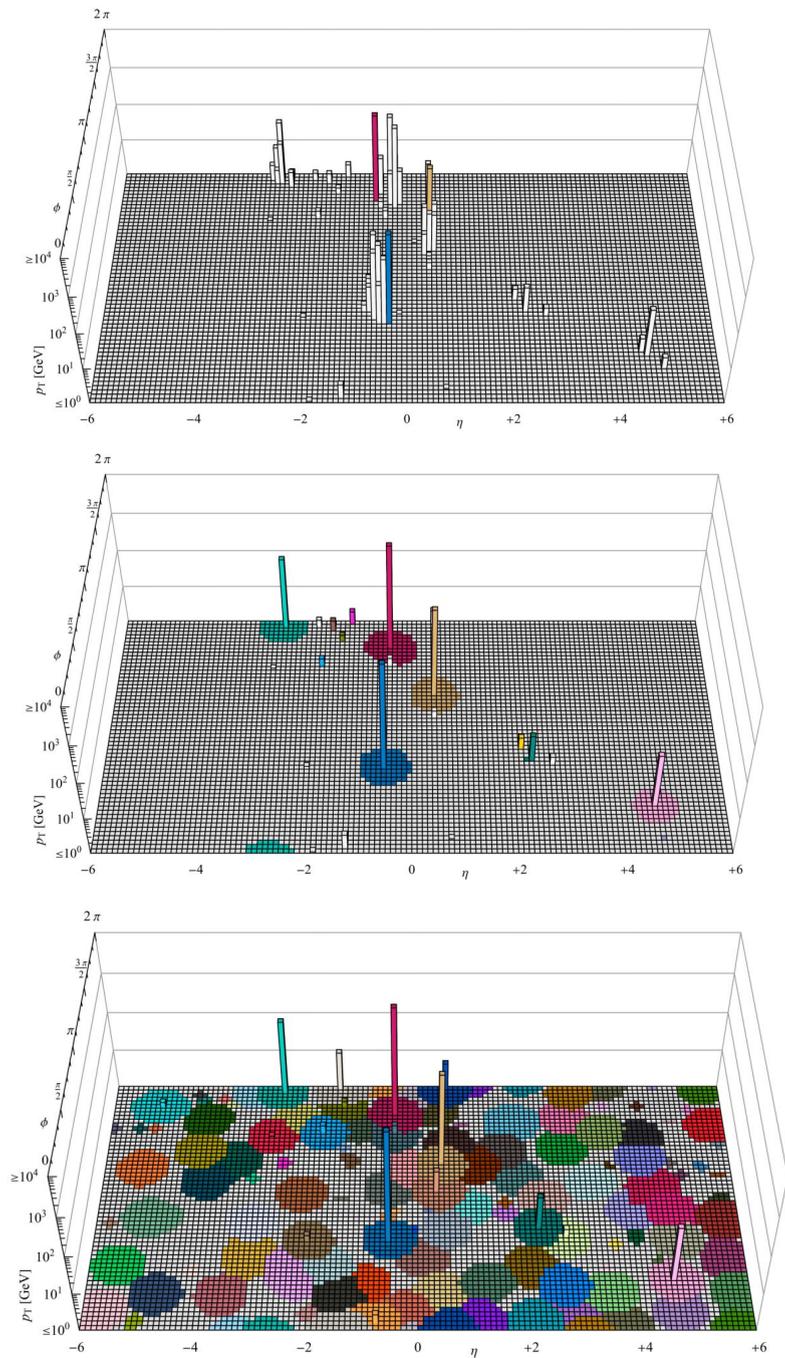


Figura 31: Evolución del agrupamiento de partículas a través del algoritmo anti- k_t en 3 pasos para $R = 0.5$. Inicialmente se detectan las partículas más fuertes. A continuación, se agrupan las partículas suaves dentro del entorno de las fuertes, y por último, se agrupan las áreas con partículas más suaves que no tienen cerca ninguna partícula fuerte. En la tercera figura se puede apreciar la formación de conos totales y parciales [51].

Capítulo 4

Aprendizaje Automático y Redes Neuronales

4.1 Fundamentos del Aprendizaje Automático y las Redes Neuronales

Para poder explicar en qué consiste el Aprendizaje Automático (o en inglés, Machine Learning (ML)) es necesario definir previamente la inteligencia artificial (IA). Esta consiste en la simulación de procesos de inteligencia humana a partir de la programación. Así, la IA busca crear máquinas y sistemas capaces de realizar tareas que normalmente requieren inteligencia humana, como el razonamiento, la resolución de problemas, el aprendizaje, la percepción, la comprensión del lenguaje natural y la toma de decisiones. Este es un campo interdisciplinario en constante evolución que combina la informática, las matemáticas, la estadística y la investigación cognitiva para crear sistemas que imitan aspectos de la inteligencia humana.

El aprendizaje automático se define como el subcampo de la inteligencia artificial que se centra en el desarrollo de algoritmos y modelos que permiten a las computadoras aprender y mejorar su desempeño en tareas específicas a través de la experiencia y los datos, en lugar de depender de programación explícita. Es decir, el aprendizaje automático permite a las máquinas aprender a realizar y optimizar ciertos procesos, por ejemplo, identificando patrones o estimando los resultados de sus propias predicciones.

Estos sistemas requieren de lo que se conoce como un entrenamiento previo. Este entrenamiento consiste en ejemplos históricos que la red analiza de la forma adecuada con el fin

de aprender de sus resultados, ya revelados, o bien interpretando anomalías como desviaciones del promedio. Así, por ejemplo, en lugar de crear un programa que resuelva ecuaciones algebraicas con un algoritmo fijo, se puede pensar en entrenar una inteligencia artificial con ecuaciones de las cuales ya se conocen las soluciones. El modelo utiliza estos datos para aprender relaciones y patrones subyacentes, ajustando sus parámetros internos de manera que pueda hacer predicciones precisas sobre nuevos datos no vistos.

El aprendizaje automático ha revolucionado la forma en que abordamos problemas complejos en la actualidad, permitiendo la automatización de tareas, la toma de decisiones basada en datos y el descubrimiento de conocimientos a partir de grandes conjuntos de información. Esto es de gran importancia en la ciencia, y en concreto, en la física, por las grandes cantidades de datos que se suelen manejar, pues estos métodos ayudan a acortar los tiempos y los gastos energéticos de algunos procesos, así como a mejorar sus resultados.

4.2 Historia y evolución de la inteligencia artificial

Los inicios de la inteligencia artificial se remontan al año 1950, cuando el matemático Alan Turing se preguntó si serían las máquinas capaces de pensar. Poco antes de eso, en el año 1943, ya se había publicado un artículo titulado «*A Logical Calculus of Ideas Immanent in Nervous Activity*» (Warren McCullough, Walter Pitts), en el que se presentaba el primer modelo matemático para la creación de una red neuronal, pero no fue hasta 1950 que los estudiantes Marvin Minsky y Dean Edmonds crearon el primer ordenador de red neuronal, conocido como *Snarc*, al tiempo que Alan Turing publicó el Test de Turing (ver Figura 32). Este test, que se sigue utilizando a día de hoy, consiste en colocar a un entrevistador (entidad C) a interrogar simultáneamente a una persona (entidad B) y a una máquina (entidad A). Si el entrevistador no era capaz de distinguir entre la persona y la máquina a partir de sus preguntas, la máquina había pasado el test.

La creación en el año 1952 de una IA capaz de aprender a jugar al ajedrez de forma

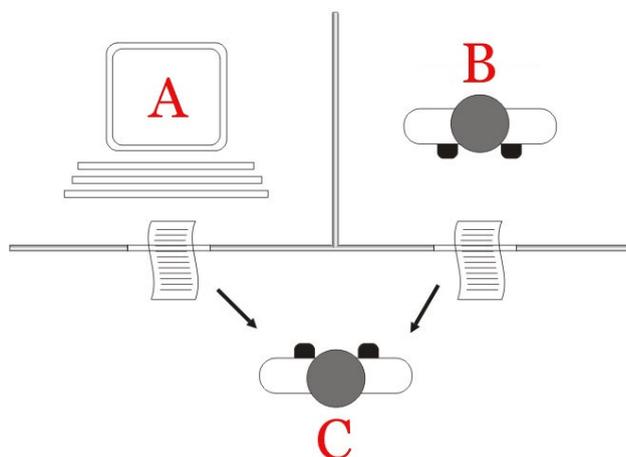


Figura 32: *Dibujo esquema del funcionamiento del Test de Turing.*

autónoma, fue un hito en la historia. Sin embargo, la falta de avances en la investigación de la traducción automática simultánea utilizada para traducir del ruso al inglés en el contexto de la Guerra Fría, conllevaron a un parón en la investigación, que comenzó a remontar tras la primera victoria de una máquina sobre el campeón mundial de ajedrez Gary Kasparov en el año 1996. Posteriormente, numerosos avances tecnológicos propiciaron el lanzamiento por parte de Google de un sistema de reconocimiento de voz en el año 2008, lo que dio inicio a una nueva etapa de pleno desarrollo.

En la última década, se ha producido un desarrollo exponencial de este campo, destacando la creación en 2012 de una red neuronal que fue entrenada con 10 millones de vídeos de YouTube y logró aprender a reconocer un gato sin que se le enseñara lo que es un gato, así como la introducción de estos sistemas en videojuegos o en la propia vida cotidiana, con el lanzamiento en el año 2022 de ChatGPT.

4.3 Redes neuronales

Por su parte, el inicio del desarrollo de las redes neuronales tuvo lugar en el año 1958, con la invención de Frank Rosenblatt del Perceptron, inspirado en el trabajo previamente mencionado de Warren McCulloch y Walter Pitts. El Perceptron es la primera «neurona artificial» (ver Figura 33). Su funcionamiento es sencillo: tiene dos entradas (puede tener

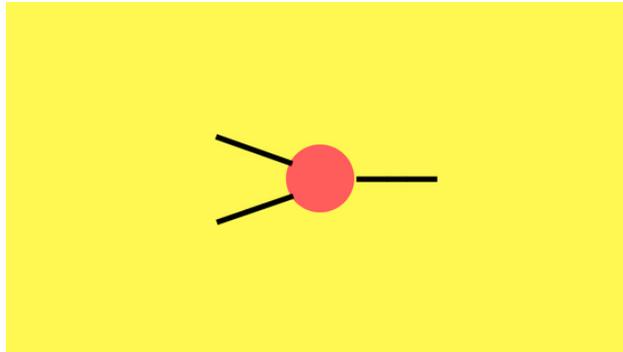


Figura 33: *Esquema del Perceptrón, la primera neurona artificial. [53]*

más, pero en su versión más simple, hablamos únicamente de dos) de tipo binario, es decir, pudiendo ser únicamente 0 o 1 (las entradas serán x_1, x_2).

Cada entrada viene acompañada de su respectivo «peso» (w_1, w_2), un número que expresa la importancia de la respectiva entrada con la salida. Si la suma de las multiplicaciones de los pesos por las entradas ($x_1w_1 + x_2w_2$) es mayor que cierto umbral, la salida de la neurona será 1. En caso contrario, la salida será 0. Esta neurona primigenia se utiliza para operaciones binarias sencillas o para crear funciones lógicas (AND, OR), pero su importancia radica en que sirvió como punto de partida para el desarrollo de redes neuronales más complejas.

El siguiente paso se dio en 1965, con la creación de un Perceptrón de Multicapas (Multi-layer Perceptron en inglés) (ver Figura 34). Esto no es más que una ampliación de la neurona primigenia, teniendo así un sistema de varias capas formadas por distintas neuronas. Se introducen así los conceptos de capas de entrada, capas ocultas y capas de salida.

Este sistema permitía realizar operaciones más complejas, pero seguía teniendo ciertas limitaciones. Entre ellas, sólo se podía trabajar con valores de entrada y salida binarios y, lo más importante, los pesos y umbrales debían ser asignados manualmente para cada neurona, lo que suponía que a mayor número de capas, fuera más difícil conseguir los valores para los resultados deseados.

Es en la década de 1980 cuando podemos considerar que comienza el desarrollo del apren-

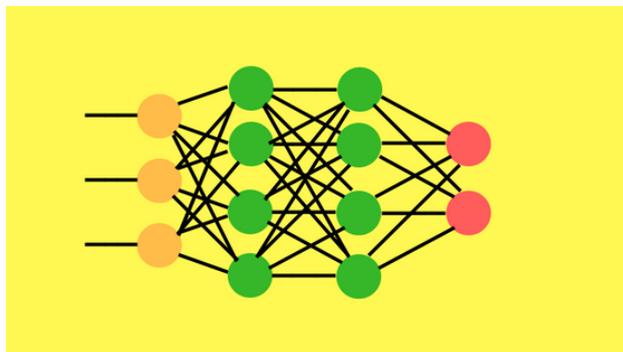


Figura 34: *Perceptrón de Multicapas, donde se distinguen las capas de entrada (amarillo), capas ocultas (verde) y capas de salida (rosa). [53]*

dizaje automático. Se crearon primeramente las neuronas *Sigmoideas*, similares al Perceptrón pero con la posibilidad de trabajar con entradas de números reales. La salida vendrá determinada por lo que se conoce como función de activación [57], que para el caso de la neurona Sigmoide es $d(z) = \frac{1}{1+e^{-z}}$ (ver Figura 35).

Por tanto, la salida dejará de ser necesariamente 0 o 1, y podrá tomar cualquier valor entre estos dos, calculado como $d(wx + b)$, donde b es el "bias" (en castellano, sesgo), que controla qué tan predispuesta está la neurona a dar como resultado un 1 o un 0, sin importar los pesos. Así un bias alto hará que la neurona sea más propensa a sacar un 1 como respuesta.

También se introduce el concepto de redes Feedforward, que son aquellas redes en las que la salida de una capa se utiliza como entrada para la siguiente capa, como ocurría con el Perceptrón de Multicapas. Esta definición es relevante ya que en el año 1986 nace el algoritmo Backpropagation, que dio inicio a las redes neuronales supervisadas. El funcionamiento es sencillo: comparando los resultados de la red con los resultados reales ya conocidos, la propia red calcula el error obtenido y propaga este hacia las capas anteriores, de forma que estas realizan pequeños ajustes y se logra de esta forma que la red aprenda de sus resultados y los optimice, lo que se trata propiamente del Aprendizaje Automático.

Gracias a estos avances, nacieron en el año 1989 las Convolutional Neural Networks, lo que le otorga al ordenador la capacidad de «ver». Este sistema consta de varias capas que

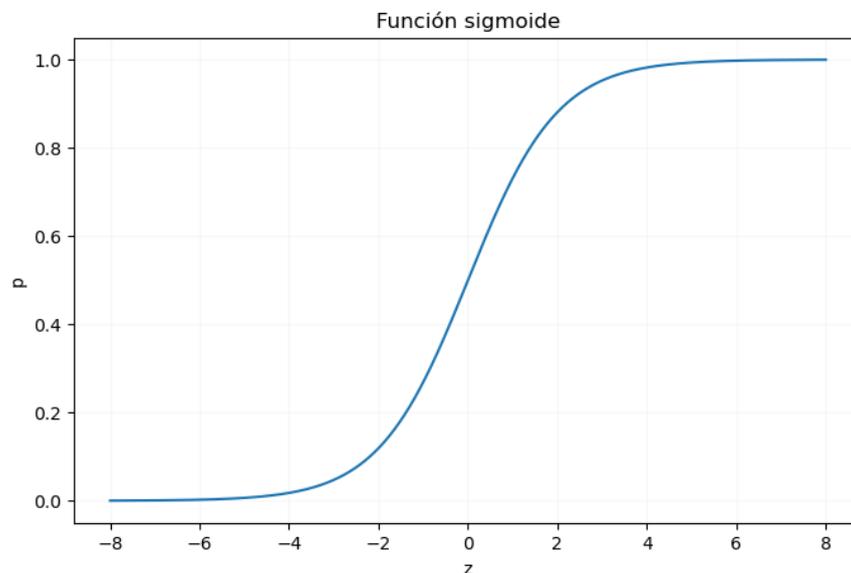


Figura 35: *Gráfica de la función sigmoide, teniendo por primera vez resultados que pueden ser distintos de 0 y 1.*

permiten la extracción de características de una imagen y su posterior clasificación. Gracias a esto es posible clasificar imágenes, detectar diversos tipos de tumores automáticamente, enseñar a conducir a los coches autónomos así como otras muchas aplicaciones.

Posteriormente, en 1997, se desarrolló un nuevo tipo de red llamado Long / Short Term Memory (LSTM) o Red Neural Recurrente, que consiste en unas celdas de memoria que permiten a la red recordar valores históricos. Esto ha tenido aplicaciones como el reconocimiento de voz o de escritura. Sin embargo, este tipo de redes tenía un inconveniente, y es que en aquellos modelos que tenían un gran número de capas eran demasiado difícil asignar los pesos adecuados antes de empezar a entrenarlos, por lo que el desarrollo quedó estancado.

Sin embargo, en el año 2006 nace el Deep Learning, con el desarrollo de las Deep Believe Networks (DBN), cuyo funcionamiento se basa en pre-entrenar las distintas capas antes de empezar a utilizar la red, de manera que la propia red se autoasigna los pesos iniciales, en lugar de ser determinados de forma manual como se hacía hasta entonces. Después de este pre-entrenamiento, la red debe ser entrenada como se hacía hasta entonces, pero la buena elección de los pesos iniciales hace que sea mucho más eficiente.

Así, tenemos toda una variedad de redes neuronales en función de sus tipos de capas, su estructura y sus parámetros de entrada. En este trabajo utilizaremos una Red Neuronal Profunda (DNN, por sus siglas en inglés) por su sencillez y eficacia.

En función del entrenamiento requerido, podemos distinguir tres tipos de redes:

- Supervisada: La propia red y el modelo se construyen en función del input (datos de entrada) y el output (datos de salida). En este caso, todos los datos de entrenamiento vienen etiquetados.
- No supervisada: Se introduce el input, y el modelo ofrece un output, que será considerado por una persona para su uso o no, dependiendo de si este es o no satisfactorio.
- Reforzada: Se introduce el input y el modelo da un output. En función de si los resultados son buenos o malos, el modelo obtiene un feedback positivo o negativo, en función del cual hace variar sus parámetros. Esto se puede repetir de forma cíclica hasta que los resultados se acerquen lo suficiente a lo esperado.

En este trabajo, se empleará un modelo de red supervisada, ya que todos los datos de entrada vienen etiquetados para entrenar correctamente al modelo, distinguiendo entre muestras de Hbb y de QCD.

4.4 Desarrollo matemático de una DNN

Como se comentó el apartado anterior, los pilares básicos para el funcionamiento de una DNN son su estructura y su función de activación. Una neurona artificial simple, suma todos los inputs multiplicados por sus respectivos pesos y al resultado le aplica una función no lineal, que será su función de activación. Además, se le podrá aplicar un bias (b), como se explicó previamente. De esta forma, una neurona que tenga 3 inputs, x_1, x_2 y x_3 , con sus respectivos pesos w_1, w_2, w_3 , y una función de activación $\sigma(x) = \frac{1}{1+e^{-x}}$, devolverá como resultado $\sigma(x_1w_1 + x_2w_2 + x_3w_3 + b)$.

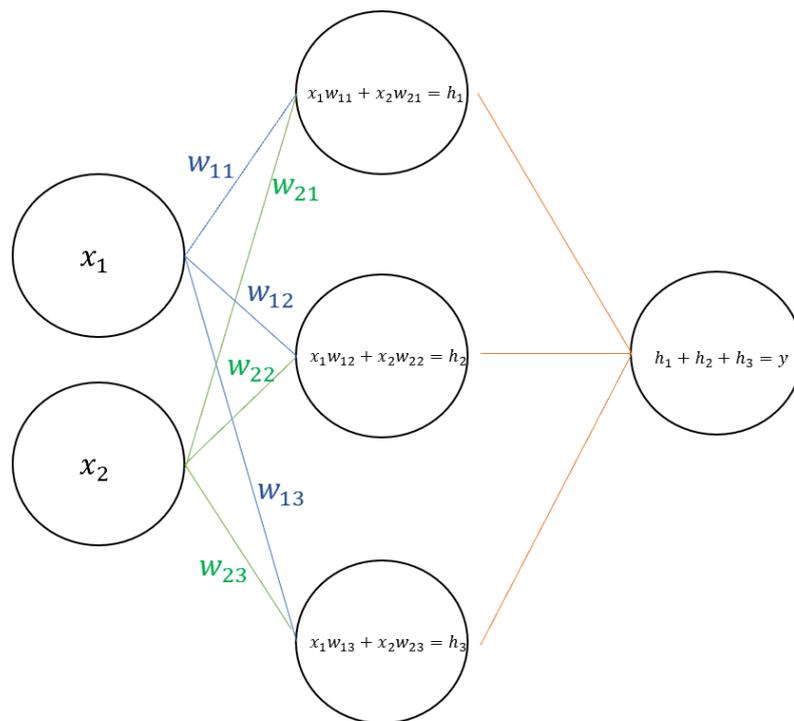


Figura 36: Esquema de una red neuronal de ejemplo.

Cuando disponemos de más de una capa, es necesario hacer un desarrollo algo más extenso. Supongamos que tenemos una red como la de la Figura 36. En este caso tenemos 2 inputs (x_1, x_2) , una capa formada por 3 neuronas y un output. La complicación radica en que, cada input tiene su respectivo peso asociado a cada neurona, por lo que en este caso, tenemos 6 pesos (w_{11}, w_{12}, w_{13} para el input x_1 y w_{21}, w_{22}, w_{23} para el input x_2). Así, tendremos para la primera neurona $a_1^{(1)} = \sigma(x_1w_{11} + x_2w_{21} + b_1)$, para la segunda $a_2^{(1)} = \sigma(x_1w_{12} + x_2w_{22} + b_2)$ y para la tercera $a_3^{(1)} = \sigma(x_1w_{13} + x_2w_{23} + b_3)$, donde el superíndice representa el número de capa. Esto puede escribirse de forma matricial de la siguiente manera:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \mathbf{W} = \begin{pmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \\ w_{13} & w_{23} \end{pmatrix}, \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

Así, por tanto:

$$\mathbf{a}^{(1)} = \begin{pmatrix} a_1^{(1)} \\ a_2^{(1)} \\ a_3^{(1)} \end{pmatrix} = \sigma(\mathbf{W} \cdot \mathbf{x} + \mathbf{b})$$

A continuación tenemos una segunda capa, compuesta por una única neurona. Esta neurona recibe los outputs de las 3 neuronas de la capa anterior, cada una con su respectivo peso. En este caso, sólo tendremos w_1, w_2 y w_3 y un único bias b . Por lo tanto, en esta segunda capa tenemos que $a^{(2)} = \sigma(w_1a_1^{(1)} + w_2a_2^{(1)} + w_3a_3^{(1)} + b)$, que de forma matricial puede escribirse como:

$$a^{(2)} = \sigma(\mathbf{W}^{(2)}\mathbf{a}^{(1)} + b); \mathbf{W}^{(2)} = \begin{pmatrix} w_1 & w_2 \end{pmatrix}$$

Así, podemos expresar esto como la aplicación de dos funciones sobre el input \mathbf{x} , tal que $f(\mathbf{x}) = f^{(2)}(f^{(1)}(\mathbf{x}))$, siendo $f^{(1)}$ la primera capa, oculta, y $f^{(2)}$ la capa de salida.

Este ejemplo se puede extender a una red con más entradas y neuronas en la primera capa. Así, para una red con n inputs y m neuronas en la primera capa, tendríamos las

matrices:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}; W = \begin{pmatrix} w_{11} & w_{21} & \dots & w_{n1} \\ w_{12} & w_{22} & \dots & w_{n2} \\ \dots & \dots & \dots & \dots \\ w_{1m} & w_{2m} & \dots & w_{nm} \end{pmatrix}; b = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{pmatrix}$$

Y, por tanto:

$$\mathbf{a}^{(1)} = \begin{pmatrix} a_1^{(1)} \\ a_2^{(1)} \\ \dots \\ a_m^{(1)} \end{pmatrix} = \sigma(\mathbf{W} \cdot \mathbf{x} + \mathbf{b})$$

Si además tuviéramos un total de p capas, se debería repetir el proceso de forma similar, teniendo en cuenta el número de neuronas de la capa, así como las entradas de la capa anterior con sus respectivos pesos. Se tendría así:

$$f(\mathbf{x}) = f^{(p)}(\dots f^{(2)}(f^{(1)}(\mathbf{x})))$$

4.4.1. Backpropagation

El backpropagation consiste en la modificación de los pesos de la red una vez se ha obtenido un resultado final, en función de lo que se parezca este al resultado deseado, al que llamaremos y .

Para verlo, supongamos un ejemplo como el de la Figura 37. Una vez realizado el procedimiento anterior, el resultado será $a_2 = \sigma(z_2)$, donde $z_2 = w_2 a_1$, y a su vez se habrá calculado $a_1 = \sigma(z_1)$, donde $z_1 = w_1 x$.

Podremos calcular el error cuadrático como: $c = (y - a_2)^2$. Como lo que queremos es minimizar este error, calcularemos las derivadas, obteniendo:

$$\frac{\partial c}{\partial w_2} = \frac{\partial c}{\partial a_2} \frac{\partial a_2}{\partial z_2} \frac{\partial z_2}{\partial w_2} \tag{7}$$

Estas derivadas se pueden calcular para nuestra red:

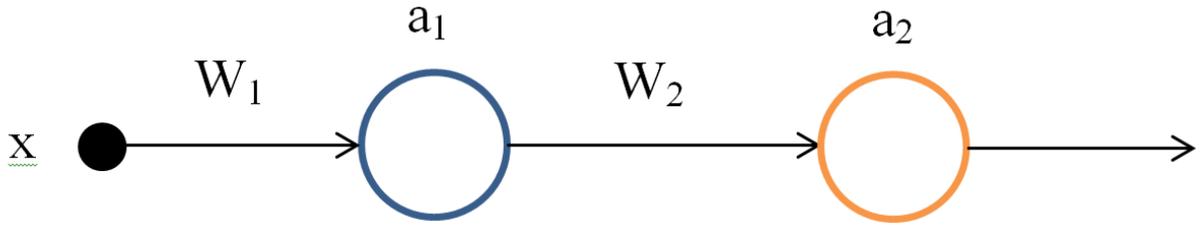


Figura 37: Esquema de una red neuronal formada por dos capas de una única neurona.

$$\begin{aligned}\frac{\partial c}{\partial a_2} &= 2(y - a_2) \\ \frac{\partial a_2}{\partial z_2} &= \sigma'(z_2) \\ \frac{\partial z_2}{\partial w_2} &= a_1\end{aligned}$$

De esta forma, se tendrá que:

$$\frac{\partial c}{\partial w_2} = 2(y - a_2)\sigma'(z_2)a_1 \quad (8)$$

Por tanto, para la que la función c sea mínima, deberá cumplirse la condición $2(y - a_2)\sigma'(z_2)a_1 = 0$.

A su vez, podremos optimizar el valor de w_1 . Para ello, deberemos tomar esta vez el camino de w_1 de la Figura 38:

$$\frac{\partial c}{\partial w_1} = \frac{\partial c}{\partial a_2} \frac{\partial a_2}{\partial z_2} \frac{\partial z_2}{\partial a_1} \frac{\partial a_1}{\partial z_1} \frac{\partial z_1}{\partial w_1} \quad (9)$$

Sin embargo, ya conocemos las dos primeras derivadas, por lo que estas se pueden reutilizar. Sólo es necesario calcular:

$$\begin{aligned}\frac{\partial z_2}{\partial a_1} &= w_2 \\ \frac{\partial a_1}{\partial z_1} &= \sigma'(z_1)\end{aligned}$$

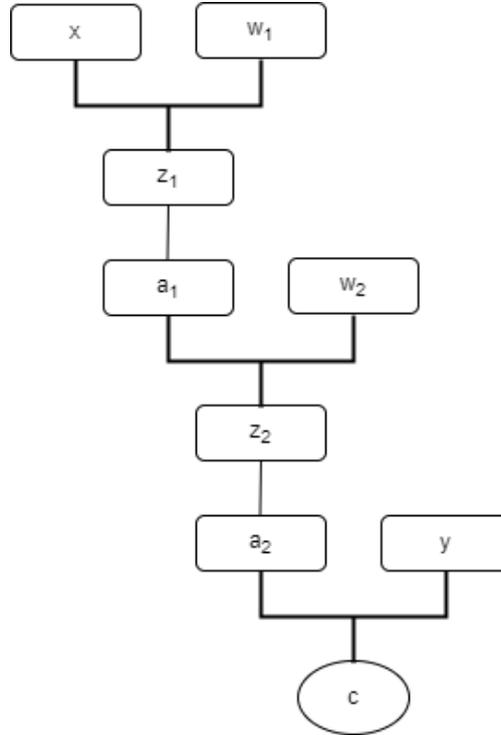


Figura 38: Esquema de las dependencias de cada variable en la red de la Figura 37.

$$\frac{\partial z_1}{\partial w_1} = x$$

Tendremos, por tanto:

$$\frac{\partial c}{\partial w_1} = 2(y - a_2)\sigma'(z_2)w_2\sigma'(z_1)x \quad (10)$$

Y la condición que deberá cumplirse será: $2(y - a_2)\sigma'(z_2)w_2\sigma'(z_1)x = 0$. Es posible extender este ejemplo a una red más compleja siguiendo el mismo procedimiento, pero teniendo en cuenta todos los parámetros.

4.5 Ejemplo de una DNN

Veamos un ejemplo del funcionamiento de una DNN similar a la utilizada en este trabajo pero aplicada a un campo distinto, como puede ser la estimación de la orientación política de una persona en función de su edad y sus ingresos [55].

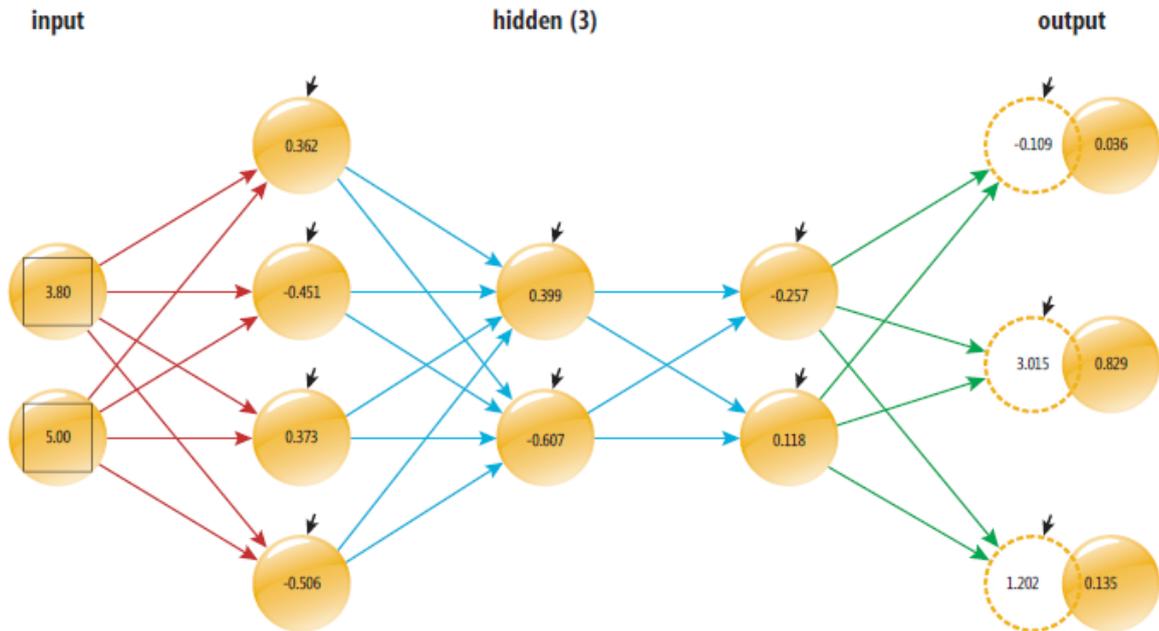


Figura 39: Ejemplo de red neuronal profunda 2-(4,2,2)-3. [55]

Así, tendremos tan sólo dos inputs, que en el caso del ejemplo son 3.80 (que hace referencia a la edad normalizada, 38 años) y 5.00 (que hace referencia a los ingresos, 50000€ al año).

La red utilizada consta de 8 nodos «ocultos» distribuidos en 3 capas, la primera con 4 nodos y las segunda y tercera con 2 (ver Figura 39). La función de activación utilizada en este caso es la función tangente hiperbólica. Las flechas que unen unos nodos con otros representan los pesos y las flechas negras que señalan estos nodos, son las bias de cada uno. Así, aplicando la función $\tanh(xw + b)$, se calculan los valores que toma cada nodo, todos ellos con un valor entre -1 y +1, ya que son los valores que puede tomar la función $\tanh()$.

Por último, tenemos 3 nodos de salida, cuyos valores finales se normalizan para poder ser interpretados como una probabilidad. Así, cada nodo de salida se puede corresponder con una orientación política, por ejemplo, conservador, moderado y liberal. La persona del ejemplo, según sus características, tendrá un 82.9% de probabilidad de ser moderado, un 3.6% de ser conservador y un 13.5% de ser liberal.

Se trata de una red neuronal profunda de tipo 2-(4,2,2)-3, ya que tiene 2 inputs, 3 capas

de 4, 2 y 2 nodos respectivamente, y 3 outputs. Tan sólo falta por determinar los valores de los pesos (las flechas de colores) y las bias (las flechas negras que apuntan a los nodos). Para esto es para lo que usamos el pre-entrenamiento, usando una gran cantidad de datos de aprendizaje con valores de entrada y salida conocidos y a continuación, utilizando un algoritmo de optimización para encontrar los valores de los pesos y las bias que ofrezcan un resultado con el mínimo error posible.

4.6 Diseño de la DNN utilizada

Para el presente trabajo se ha optado por utilizar un programa de ejemplo en Python proporcionado en la web de OpenData del CERN [58], que utiliza la dependencia «keras» para generar la DNN. Se trata de una red de tipo DFF representada en la Figura 40. En cuanto a los datos, estos son generados de forma simulada con Pythia 8, de modo que reproducen colisiones protón-protón con una energía de centro de masas de 13 TeV. Hay varias series de datos, por lo que una de ellas se utilizará para entrenar al modelo (es decir, determinar los valores de los pesos y las bias) y otra distinta para comparar la eficacia de este entrenamiento, una vez estén calculados los valores de los pesos y las bias.

El programa consta de unas variables de entrada preseleccionadas, que son posteriormente normalizadas. A continuación tenemos 3 capas, ambas con función de activación «relu» (Rectified Lineal Unit), que convierte los valores negativos en 0 y mantiene los positivos como están. La primera capa está formada por 64 nodos, mientras que las dos siguientes están formadas por 32 nodos cada una. Por último tenemos la salida, que tan sólo tendrá 2 nodos, correspondiendo a si es positivo (en este caso, si se trata de un jet b) o negativo (es un jet de tipo udsg o c). Estos resultados pueden ser un positivo verdadero, un falso positivo, un negativo verdadero o un falso negativo. Así, se definen las tasas de positivo verdadero o «True positive rate» y de falso positivo o «False positive rate»

$$True\ positive\ rate = \frac{Verdaderos\ positivos}{Verdaderos\ positivos + Falsos\ negativos} \quad (11)$$

$$False\ positive\ rate = \frac{Falsos\ positivos}{Falsos\ positivos + Verdaderos\ negativos} \quad (12)$$

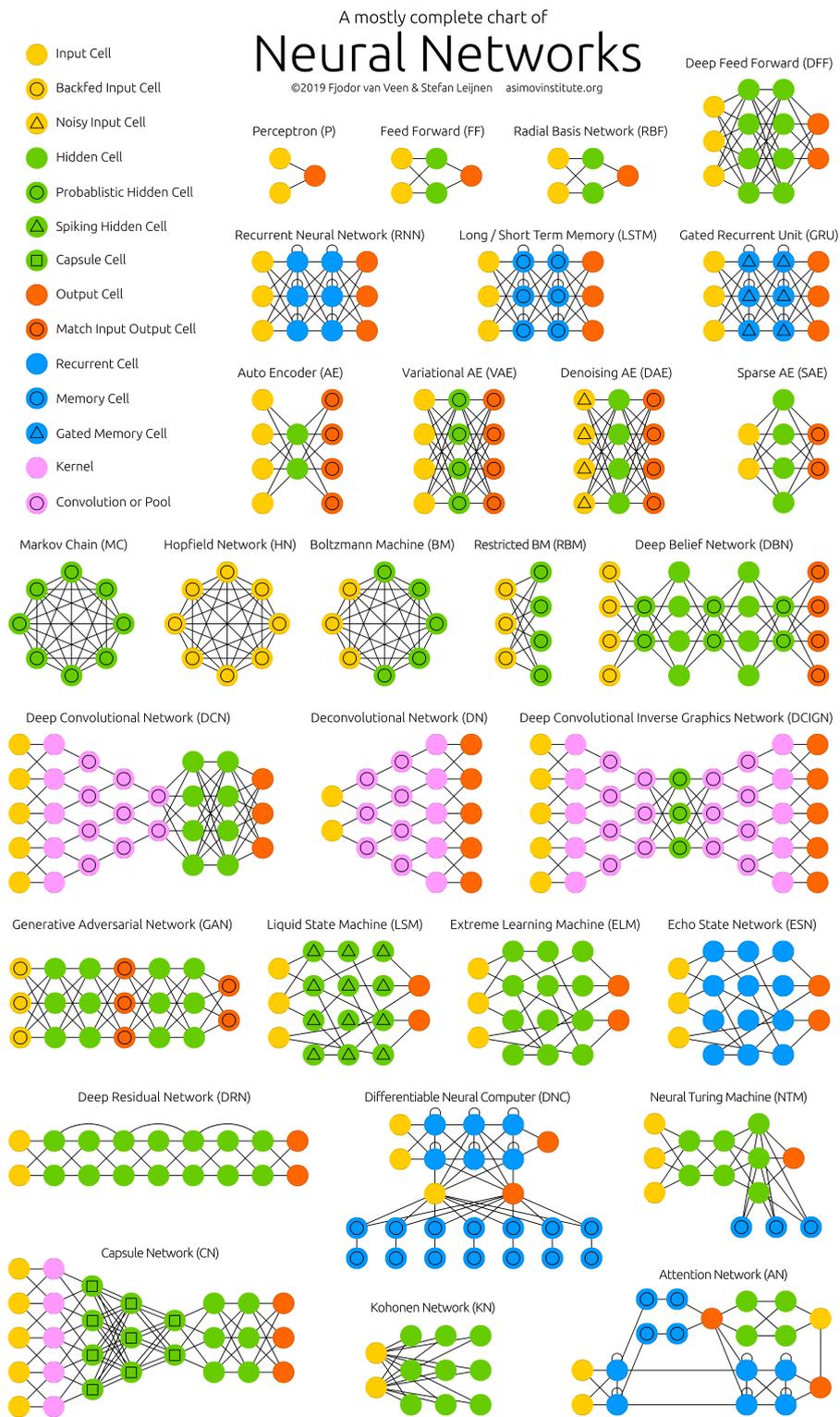


Figura 40: Gráfico de los distintos tipos de redes neuronales. [54]

Layer (type)	Output Shape	Param #
input (InputLayer)	[(None, 27)]	0
bn_1 (BatchNormalization)	(None, 27)	108
dense_1 (Dense)	(None, 64)	1792
dense_2 (Dense)	(None, 32)	2080
dense_3 (Dense)	(None, 32)	1056
output (Dense)	(None, 2)	66

=====
 Total params: 5,102
 Trainable params: 5,048
 Non-trainable params: 54
 =====

Figura 41: *Tabla generada por el programa que muestra cada una de las capas, con su número de nodos y el número de parámetros que maneja.*

Con estos dos parámetros, se puede construir la curva ROC (Receiver Operating Characteristic), que muestra la relación entre la tasa de verdaderos positivos y la tasa de falsos positivos. El área bajo la curva nos da el porcentaje de precisión del modelos.

El entrenamiento del modelo consiste en introducirle una serie de datos con sus respectivas variables y los resultados correspondientes. El modelo ajusta los pesos y las bias mediante un proceso llamado *callback*, basado en el sistema del *backpropagation* definido en el apartado 4.4.1, que finaliza cuando la mejora de precisión que tiene el modelo respecto al anterior no es considerable.

Por último, se introduce la serie de datos que quiere ser analizada (similar a la de entrenamiento) y se comparan los resultados ofrecidos por la red con los reales. El resultado se

ve reflejado en una curva ROC como la de la Figura 42, que representa los falsos positivos frente a los positivos verdaderos y cuyo área bajo la curva nos da el porcentaje de acierto. A partir del estudio de las variables, las capas de modelo y su función de activación, se intentará buscar la forma de optimizar estos resultados.

Capítulo 5

Presentación de resultados

5.1 Elección de variables

Antes de comenzar a buscar la forma de optimizar la curva ROC, es imprescindible reconocer aquellas variables, asociadas a sus respectivos observables, cuyo análisis será más determinante dependiendo de las diferencias que presenten los jets de cada tipo al respecto. Todas las variables utilizadas están definidas en el Anexo I. Para comenzar, se hará una representación de cada una de las variables diferenciando aquellos jets que tienen quarks b (procedentes de un bosón de Higgs) del ruido (procesos QCD, en este caso). Un ejemplo de variable cuyo análisis no aporta información interesante, sería, por ejemplo, la variable 'fj_phi', referida al valor del ángulo ϕ , representada en la Figura 43, puesto que los valores que se presentan para Higgs y para QCD son similares dada la simetría que existe en el plano transversal. Por el contrario, en la Figura 43 también se ve representada la variable 'fj_tau21', correspondiente al observable τ_{21} que representa la relación entre jets de 1 y 2 «ramas», la cual es muy diferente para una señal de Higgs respecto de la señal de fondo.

Para realizar este estudio, se hace un muestreo general de las 75 variables a grandes rasgos, y se seleccionan las que pueden ser más discriminadoras. En la Figura 44 se muestra un ejemplo de esta representación. De este primer filtrado, podemos descartar algunas variables, como puede ser 'fj_z_ratio', correspondiente al observable «z ratio» descrito en el apartado 3.2, así como detectar otros que pueden ser determinantes a la hora de entrenar a la red, como por ejemplo 'fj_ptDR', definida como el momento transversal p_T por el ΔR entre dos jets formados a partir del jet principal.

En algunos casos, es preciso realizar un proceso de representación más exhaustivo para poder determinar correctamente si se trata de una variable útil o no. Este proceso se explica

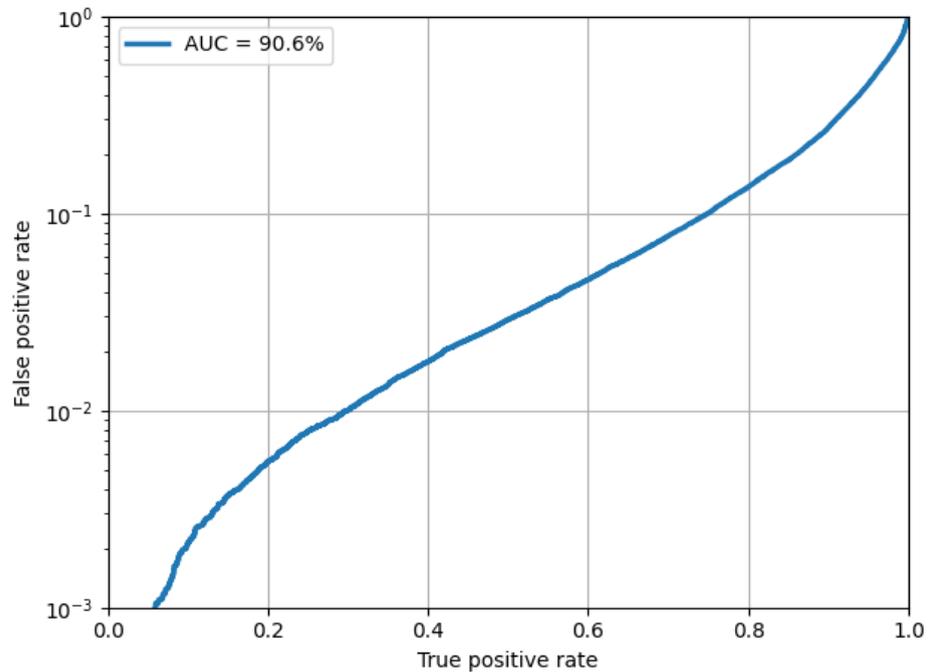


Figura 42: Curva ROC generada por el programa para los valores por defecto.

en detalle en el Anexo II.

A partir de este proceso, se han seleccionado 35 variables, clasificadas según su relevancia aparente.

A continuación se realiza un segundo filtrado, comprobando la importancia real de cada variable en el programa. Para ello, se ejecuta el programa con todas las variables, y se suprimen de una en una, siguiendo el siguiente criterio: si al suprimir una cierta variable la precisión baja, esa variable es necesaria y se mantiene. Si por el contrario, la precisión aumenta o se mantiene, quiere decir que esa variable no está aportando una información relevante, por lo que se puede suprimir.

Aplicando esto a cada una de las 35 variables, se tiene que de esas, 25 son necesarias, y por tanto se han suprimido 10.

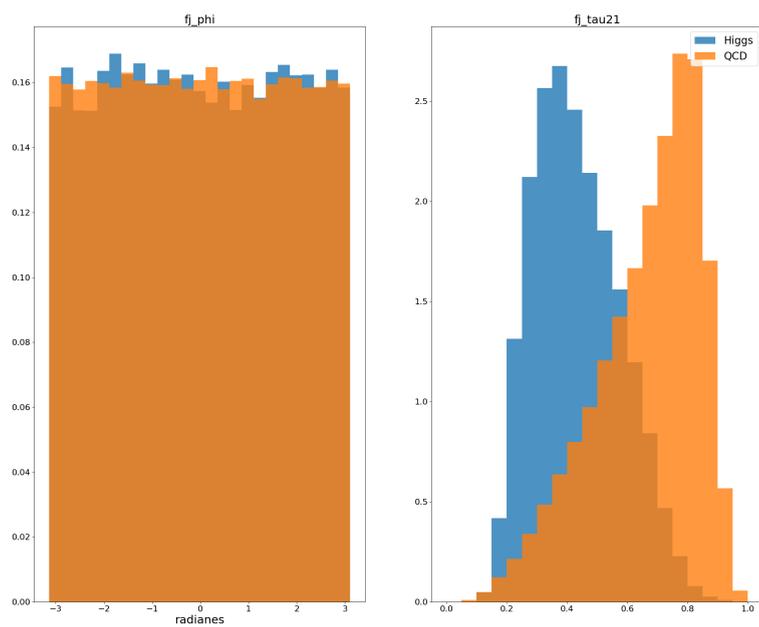


Figura 43: A la izquierda, distribución de densidad de los valores del ángulo ϕ para los distintos jets. Se puede ver que estas densidades son similares tanto para QCD como para Higgs, por lo que no será una buena variable a tener en cuenta a la hora de utilizar la red neuronal. A la derecha, distribución correspondiente al observable τ_{21} , con valores muy distintos para Higgs y QCD, por lo que fj_tau21 será una buena variable a tener en cuenta.

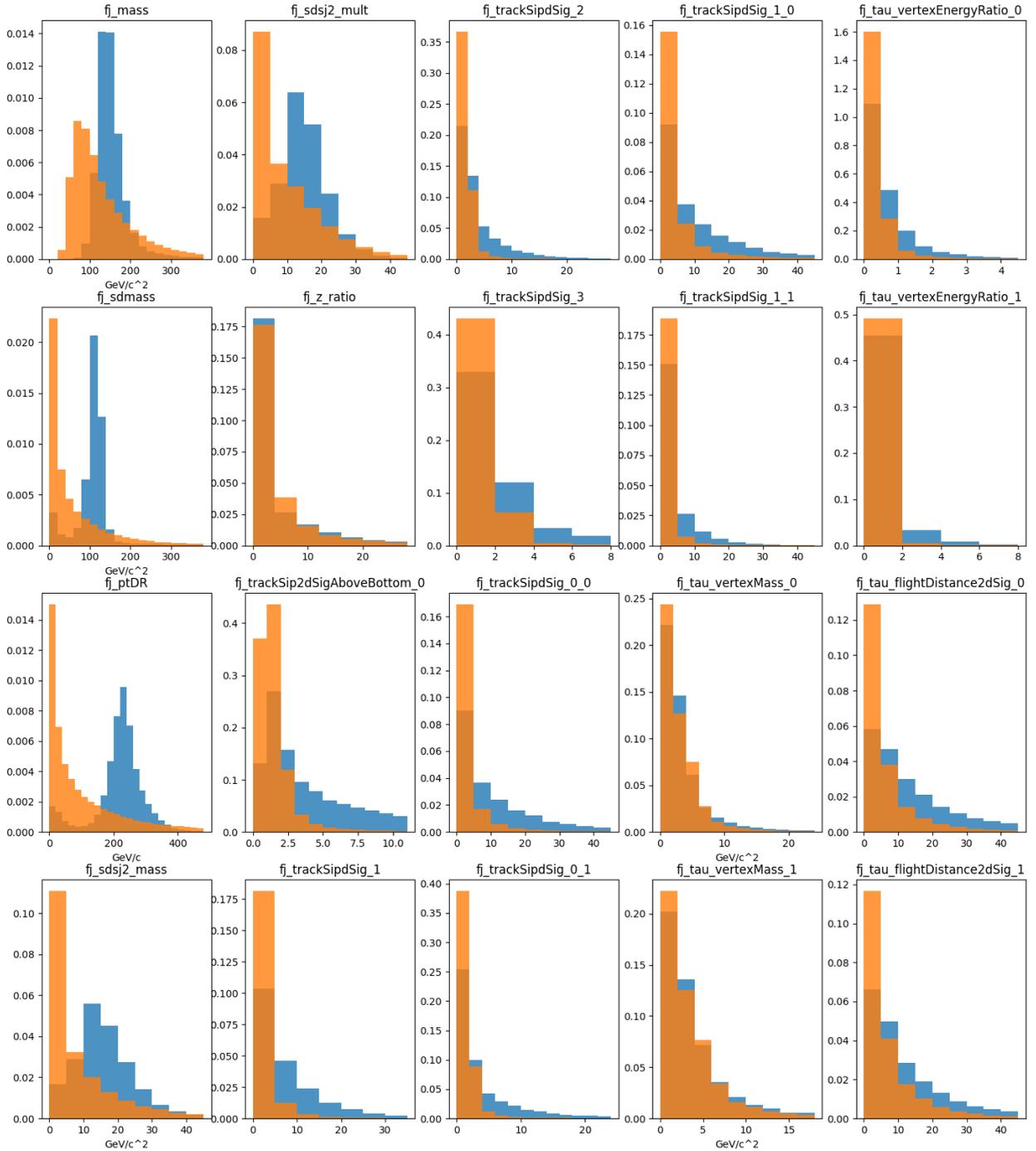


Figura 44: Muestreo de 20 de las 75 variables para realizar un primer filtrado.

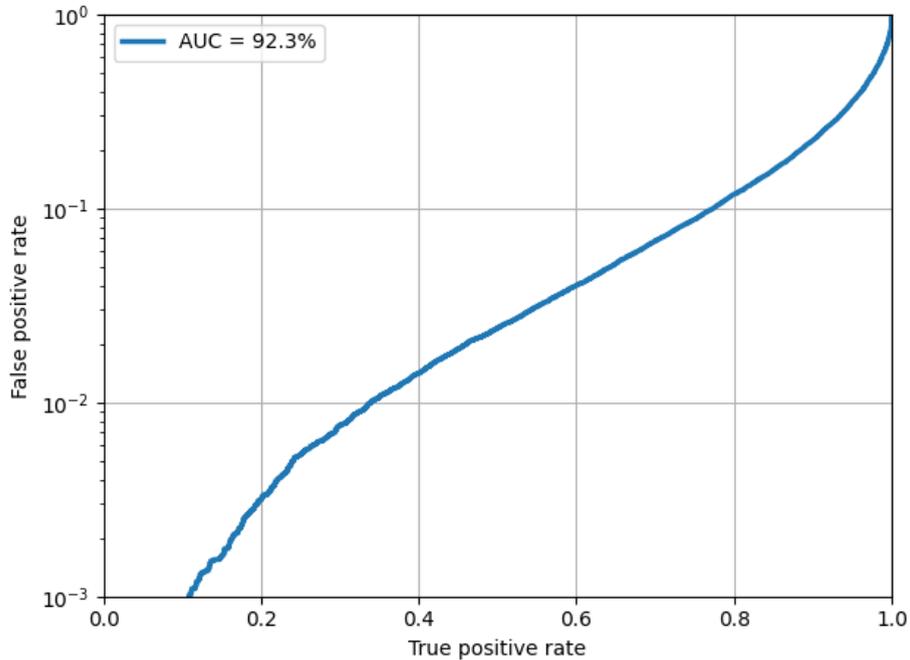


Figura 45: *Curva ROC tras el entrenamiento de la DNN con 10 observables.*

5.2 Variación de las variables

Una vez hemos seleccionado las 25 variables, se realizan pruebas con 10, 20 y 25 variables, para ver si utilizando un menor número de variables es posible obtener resultados más precisos con un menor coste. La primera prueba se hará seleccionando únicamente los 10 que presentan mayores diferencias entre un tipo de jet y otro. El resultado puede verse en la Figura 45. Con una reducida selección de variables, se ha obtenido una precisión del 92.3%, bastante alta para requerir únicamente 10 variables. Además, se ha medido el tiempo de trabajo del programa, siendo en este caso 1:13.04 min. Para la segunda prueba se seleccionan, además de estas 10 variables, las siguientes 10 que presentan mayores diferencias, obteniéndose como resultado la Figura 46. En este caso, la precisión es considerablemente superior, de un 95.7%, y el tiempo de espera es de 56.71s. Por último, se han añadido otras 5 variables, las cuales tenían diferencias menos perceptibles (Figura 47). En este caso la precisión ha sido de un 96.4%, considerablemente mayor, con un tiempo de espera de 53.21s, siendo este el menor de todos. Los motivos de estas diferencias se explican con mayor detalle en el apartado de conclusiones.

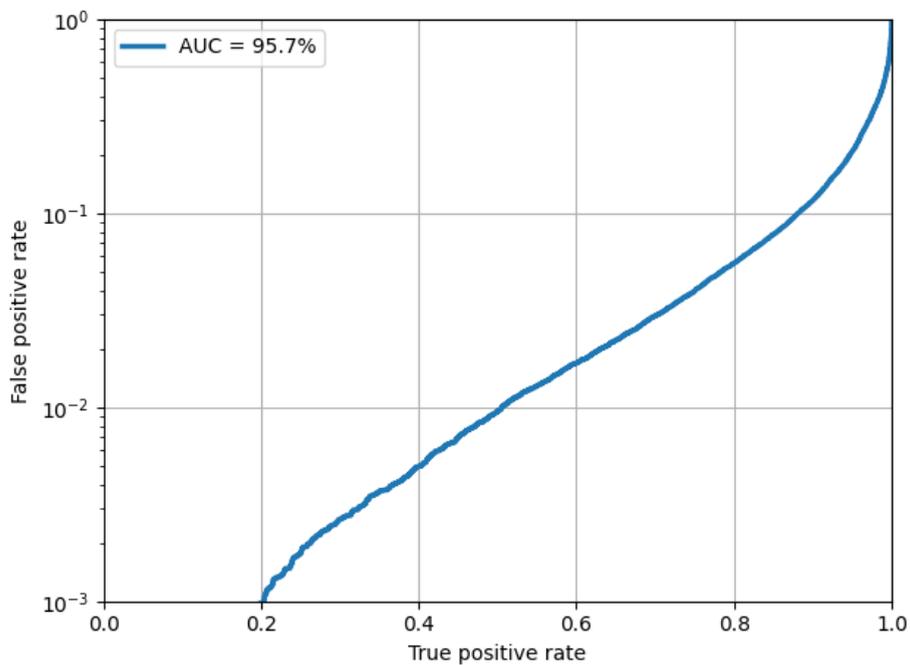


Figura 46: Curva ROC tras el entrenamiento de la DNN con 20 observables.

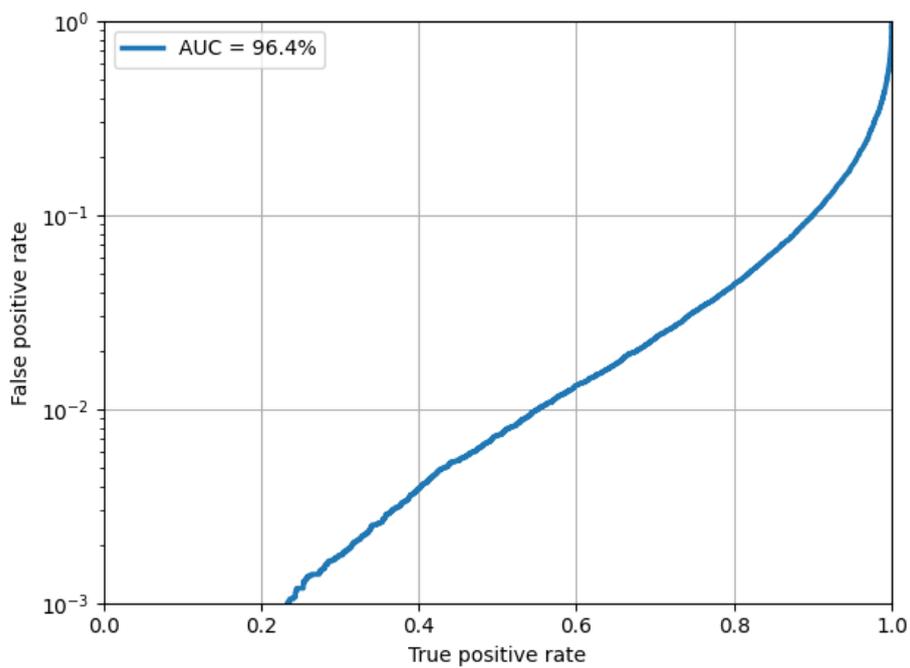


Figura 47: Curva ROC tras el entrenamiento de la DNN con 25 observables.

Nº capas	Estructura	Precisión	Tiempo
1	32	96.4	4:25.40
	64	96.4	3:42.20
2	32-32	96.3	1:51.46
	64-32	96.4	1:27.09
	64-64	96.4	1:14.94
3	32-32-32	96.3	1:08.22
	64-32-32	96.4	0:53.21
5	64-64-32-32-32	96.3	0:48.41
	64-64-64-64-64	96.2	0:44.65
8	64 x 8	96.1	0:49.40
	64 x 4 + 32 x 4	96.2	1:00.68
12	64 x 12	96.1	1:20.65

Figura 48: *Tabla resumen de los resultados al variar el número de capas*

5.3 Variación de la estructura de la red

Otro de los factores que afectan al rendimiento del programa es la estructura de la red. Por defecto, se ha trabajado con la red de la Figura 41. Sin embargo, en este apartado se probará a cambiar la estructura de esta, añadiendo y quitando capas y modificando el número de neuronas de cada una, con el objetivo de aumentar la precisión y reducir la carga de trabajo, como se hizo en el apartado 5.2.

El resultado se muestra en la tabla de la Figura 48. Teniendo en cuenta tanto la precisión como el tiempo de espera, el mejor modelo posible es el formado por una red de 3 capas, de tipo 64-32-32. Redes con capas de únicamente 32 neuronas presentan una menor precisión, mientras que las redes con al menos una capa de 64 neuronas, pero con 1 o 2 capas, presentan tiempos de trabajo mayores. No obstante, al aumentar el número de capas a más de 3 no se logra mejorar la precisión, sino que se consigue el efecto contrario, debido a que se pierde parte de la información por el backpropagation a capas tan alejadas entre sí. Además, se puede ver que cuando todas las capas son iguales, la precisión también disminuye.

Función de activación	Precisión (%)
relu	96.4
sigmoid	96.6
softmax	96.3
softplus	96.5
softsign	96.4
tanh	96.5
selu	96.4
elu	96.5
leaky_relu	96.4
relu6	96.3
silu	96.4
gelu	96.4
hard_sigmoid	96.4
linear	86.0
log_softmax	95.1

Figura 49: Tabla resumen de los resultados al variar la función de activación

5.4 Variación de la función de activación

El tercer parámetro que nos puede permitir mejorar el rendimiento del modelo es la elección de la función de activación, ya que una función de activación adecuada puede hacer que el proceso tenga una precisión mayor. Por ello, se ha probado con 14 funciones de activación disponibles en la documentación de *keras* [56], manteniendo las 25 variables comentadas en el apartado 5.2 y la estructura de red más eficaz del apartado 5.3, ya que es la configuración que mejores resultados ha dado para esos parámetros.

Los resultados se muestran en la Figura 49. La función que ha dado un mejor resultado ha sido la función 'sigmoid', con una precisión de 96.6% seguida de las funciones 'softplus', 'tanh' y 'elu'. Por otra parte, las funciones 'relu' (por defecto), 'softsign', 'selu', 'leaky_relu', 'silu', 'gelu' y 'hard_sigmoid' han dado resultados aceptables, aunque menos precisos que las 4 anteriores. Por el contrario, hay funciones que por su definición no han dado buenos resultados, como es el caso de la función 'linear'.

Conclusión

En este trabajo se exploró el ámbito del etiquetado de quarks b en el experimento CMS del CERN, combinando una revisión teórica del Modelo Estándar, los experimentos que tienen lugar en el LHC del CERN, los criterios para el etiquetado de quarks y gluones y una breve introducción a las redes neuronales, con su aplicación práctica a través de una red neuronal de aprendizaje automático. El objetivo principal del trabajo era la optimización de esta red neuronal de aprendizaje automático de forma que presente una mayor eficacia, aumentando la precisión y reduciendo el tiempo de espera.

Los resultados han sido bastante positivos en cuanto a optimización. Gracias al primer filtrado, se han suprimido numerosos observables que no aportaban información relevante, y eso ha permitido reducir considerablemente los tiempos de ejecución. Con el segundo filtrado se ha determinado cuales son las variables indispensables y cuales son superfluas, pudiendo reducir así también los tiempos de ejecución. Con estas 25 variables se han realizado varias pruebas, eligiendo en primer lugar únicamente las 10 más relevantes, a continuación las 20 más relevantes y por último el total de 25 variables, que daba como resultado la elección más óptima. Cabe destacar que, contrario a lo que se pudiera intuir en una primera aproximación, los tiempos son menores cuanto mayor es el número de variables. Esto es debido a que la red tiene un sistema de autodetención (callback) que entra en juego únicamente cuando no es capaz de mejorar más, independientemente del número de pasos que lleve. Por esto, al tener menos variables, la red tarda más tiempo en llegar a este punto en el que se detiene sola, ya que debe realizar un mayor número de pasos. Al aumentar el número de variables, aumenta el tiempo que se emplea en cada paso, pero se tarda mucho menos en llegar al punto de autodetención, por lo que el tiempo total se reduce.

Algo similar ocurre al variar las capas de la red. Cuando tenemos una única capa, la red requiere mucho más tiempo para llegar al punto de autodetención, ya que los pasos son muy rápidos, pero se requiere un gran número de pasos. Conforme aumenta el número de capas, los pasos que se dan son más lentos, ya que en lugar de pasar por una única capa debe pasar

por más, pero el número de pasos requerido para alcanzar el punto de no mejora es menor, por lo que el tiempo disminuye. La precisión, aunque es similar para pocas capas, disminuye al aumentar el número de capas. Dentro de un mismo número de capas, la precisión es mayor si en lugar de ser todas las capas iguales, se combinan capas de 64 y 32 neuronas.

Respecto a la función de activación, se ha visto que presenta un impacto reducido en la precisión, siempre y cuando se trate de una función distinta de la función «linear», aunque la elección de una buena función de activación permite potenciar los buenos resultados. En este caso, la función más óptima es la función 'sigmoid'.

Aunque estos resultados parecen considerablemente buenos, se pueden reconocer algunas limitaciones. La elección de los datos se ha hecho de manera aleatoria a partir de los generados por Pythia 8 [58], ya que no se dispone de los equipos adecuados para trabajar grandes cantidades de datos. Un estudio más completo podría realizarse como un estudio estadístico seleccionando varias fuentes de datos y viendo si los resultados mejoran o empeoran en función de dicha selección. Además, las funciones de activación que se han utilizado son funciones generales, que se pueden aplicar en numerosos campos. Un estudio exhaustivo sobre el campo podría ayudar a desarrollar una función de activación aplicada al etiquetado de quarks b que diese resultados más precisos.

En definitiva, el modelo final consta de la siguiente estructura:

- 25 variables, elegidas cuidadosamente por ser las que presentan una mayor discriminación entre Higgs a $b\bar{b}$ y QCD.
- Una red formada por 3 capas de 64, 32 y 32 neuronas respectivamente, por ser la que mejores resultados ha proporcionado, teniendo en cuenta su precisión y su tiempo de trabajo.
- La función 'sigmoid' como función de activación de cada una de las capas.

Con todo esto, se ha obtenido una precisión final del 96.6%, que se puede considerar como buena.

Anexos

Anexo I. Los datos

Como se comentó en el apartado 4.6, los datos se generan de forma simulada a partir de sucesos de colisión protón-protón con una energía en centro de masas de 13 TeV con Pythia 8. El conjunto de los datos consta de 182 archivos en formato Hierarchical Data Format (HDF) h5 con un espacio total de almacenamiento de 228.1 GB, de la cual se ha seleccionado un subconjunto debido a las limitaciones de memoria del entorno de trabajo en la nube, usando una parte de la muestra de entrenamiento y validación, y otra para generar los resultados. Estos archivos de datos tienen formato .h5 y contienen información de un total de 178 variables, divididas en aquellas relativas al jet, a las partículas candidatas, a las trazas y a los vértices. Para este trabajo se han utilizado únicamente aquellas que hacen referencia a los jets y sus subjets. En un primer momento, se han utilizado un total de 75 variables, reduciéndose a 35 tras un primer filtrado, y a 25 tras el segundo. La información relativa a cada una de las variables se encuentra disponible en [58] Estas variables se definen de la siguiente manera:

- jetNtracks: Número de trazas asociadas a un jet AK8.
- nSV: Número de vértices secundarios de un jet AK8.
- n_sdsbjets: Número de subjets en el jet AK8, obtenidos con el método soft drop. En este caso, puede ser como máximo 2.
- mass: Masa total del jet AK8, previa al filtrado de los datos.
- phi: Ángulo azimutal ϕ del jet AK8.
- pt: Momento transversal p_t del jet AK8.
- tau.i: Relativa al observable «N-subjettiness» τ_N dada la hipótesis de que el jet tenga i puntas. Así, la i podrá tomar los valores 1 (en caso de que no tenga subjets), 2 (si tiene un sólo subjet) o 3 (si tiene dos subjets), por lo que tendremos 3 variables: tau1,

tau2 y tau3. Además, se definen las variables tau21 y tau32 como los cocientes entre los respectivos valores de $\tau_{N,i}$ (tau21 = tau2/tau1; tau32 = tau3/tau2).

- z_ratio: Relativa al observable z_ratio, como se define en el apartado 2.3.1.
- sdmass: Masa del jet AK8 utilizando el método soft drop.
- ptDR: Momento transverso multiplicado por el ΔR entre los dos subjets.
- relptdiff: Diferencia relativa entre los momentos transversos de los dos subjets.
- axis_i: Indica la dirección del eje del subjet i. Como un jet puede tener como máximo dos subjets, la i podrá tomar los valores 1 o 2.
- eta: Pseudorrapidez η como se define en el apartado 2.3.1.
- mult: Multiplicidad del subjet.
- ptD: Función de fragmentación como se define en el apartado 2.3.1.
- Prefijo sdsj_.: Hacen referencia a «soft drop subjet», es decir, los subjets que han sido obtenidos por el método soft drop. Así, por ejemplo, la variable «sdsj2_axis1» hace referencia al primer eje del segundo subjet.
- trackSipdSig_i: Significancia del parámetro de impacto 3D en la traza i, donde la i toma valores de 0 a 3 según el tamaño de la traza, siendo 0 el índice correspondiente a la traza mayor, 1 a la segunda mayor, y sucesivos.
- trackSipdSig_j_i: Significancia del parámetro de impacto 3D en la traza i asociada al eje del observable «N-subjettiness» j, donde j toma valores 0 y 1. Así, la variable «trackSipdSig_0_1» se refiere a la segunda traza mayor (i=1) del primer eje «N-subjettiness» (j=0).
- trackEtaRel_i: Pseudorrapidez relativa al eje del jet, donde i toma valores de 0 a 2 según el valor de la pseudorrapidez, donde el 0 es el índice correspondiente a la menor pseudorrapidez.

- `tau_vertexMass_i`: Masa total del vértice secundario para el eje de «N-subjettiness» i , donde i toma valores 0 y 1. La masa total del vértice secundario se define como la masa invariante de todas las trazas asociadas a ese vértice.
- `tau_vertexEnergyRatio_i`: Razón entre la energía del vértice secundario para el eje de «N-subjettiness» i , y la energía total del jet AK8. La energía total del jet se define como la energía total de todas las trazas asociadas a jet.
- `tau_flightDistance2dSig_i`: Significancia de la distancia de vuelo 2D con la menor incertidumbre en la distancia de vuelo 3D, asociada al eje de «N-subjettiness» i , donde i toma los valores 0 y 1.

Anexo II. Elaboración de los histogramas

Para la elaboración de los histogramas, ha sido necesario tener en cuenta dos cosas:

1. El rango de datos que debemos tomar.
2. La separación entre columnas a la hora de representar, para que la información no se vea tergiversada por la forma del histograma.

En cuanto a la primera, en algunos casos ha sido necesario representar varias veces, acotando el rango de representación, ya que inicialmente, algunos conjuntos de datos tenían valores extremos que no permitían representar de forma correcta la información (ver Figura 50). Por ello, ha sido necesario ajustar los rangos mínimo y máximo para cada muestra, descartando así los valores extremos que presenta en cada caso. De esta forma se obtienen histogramas de mucha más calidad y que presentan una información mucho más precisa (ver Figura 51).

Una vez que tenemos el rango, para lograr una representación adecuada que presente correctamente toda la información, es necesario introducir el concepto de «binning», o ancho de intervalo. El binning en un histograma se refiere al proceso de dividir el rango completo de valores de una variable continua en intervalos más pequeños y discretos, llamados «bins». El proceso de binning implica agrupar los valores continuos en categorías discretas o intervalos para simplificar la visualización y el análisis de la distribución de datos. Cada bin representa un rango específico de valores, y la altura de la barra en ese bin indica la frecuencia o el recuento de datos que caen en ese rango. El tamaño de los bins puede afectar la apariencia del histograma y la interpretación de la distribución. Bins más pequeños proporcionan más detalle pero pueden hacer que el histograma sea más sensible a pequeñas variaciones en los datos. Bins más grandes suavizan la distribución, pero podrían ocultar detalles importantes.

Para elegir un binning correcto, se utilizó en este trabajo la fórmula de Sturges. Esta fórmula, de origen empírico, da el número óptimo de bins para un conjunto de datos concreto:

$$k = 1 + \log_2(N)$$

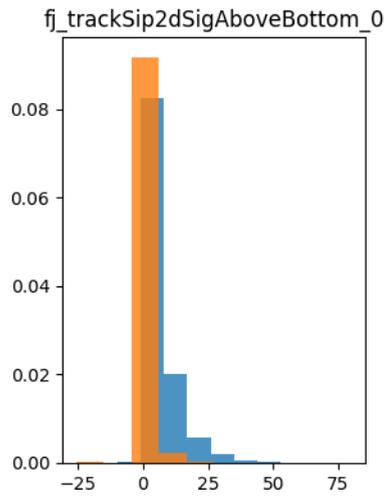


Figura 50: *Histograma mal representado para una de las variables*

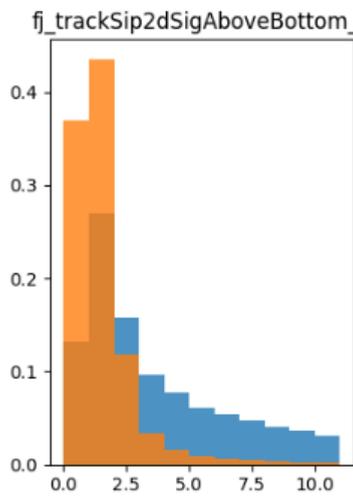


Figura 51: *Histograma bien representado para la misma variable*

siendo k el número de bins y N el número total de datos del conjunto. Para nuestro caso, el dataset contaba con un total de 187819 sucesos, con al menos 2 jets por evento, por lo que el número de bins utilizado ha de ser $k = 18.52$. Este criterio estadístico general puede ser suficiente dado el tamaño limitado de la muestra, aunque en algunos casos observables sería deseable tener en cuenta la resolución experimental del observable.

Anexo III. Utilización del código.

Para la ejecución del código, se ha hecho uso de la plataforma de trabajo de Google «Kaggle» [59], que permite un alto almacenamiento en la nube, de varios cientos de GB, grandes cantidades de CPU disponibles con 30GB de RAM, hasta 30+30 horas semanales de GPU (T4x2 y P100) para acelerar el procesamiento de DNN y hasta TPUs (Tensor Process Units) que requieren una programación especializada.

A continuación se presentan las líneas de código utilizadas:

```
# 27 features
features = [
    'fj_sdmass',
    'fj_ptDR',
    'fj_mass',
    'fj_pt',
    'fj_sdsj1_pt',
    'fj_sdsj2_mult',
    'fj_jetNTracks',
    'fj_tau1',
    'fj_relptdiff',
    'fj_sdsj1_eta'
]

# spectators to define mass/pT window
spectators = ['fj_sdmass',
              'fj_pt']

# 2 labels: QCD or Hbb
labels = ['fj_isQCD*sample_isQCD',
          'fj_isH*fj_isBB']
```

```

nfeatures = len(features)
nspectators = len(spectators)
nlabels = len(labels)

def get_features_labels(file_name , remove_mass_pt_window=True):

    # load file
    h5file = tables.open_file(file_name , 'r')
    njets = getattr(h5file.root , features [0]).shape [0]

    # allocate arrays
    feature_array = np.zeros ((njets , nfeatures))
    spec_array = np.zeros ((njets , nspectators))
    label_array = np.zeros ((njets , nlabels))

    # load feature arrays
    for (i , feat) in enumerate(features):
        feature_array[:, i] = getattr(h5file.root , feat)[:]

    # load spectator arrays
    for (i , spec) in enumerate(spectators):
        spec_array[:, i] = getattr(h5file.root , spec)[:]

    # load labels arrays
    for (i , label) in enumerate(labels):
        prods = label.split('*')
        prod0 = prods[0]
        prod1 = prods[1]
        fact0 = getattr(h5file.root , prod0)[:]

```

```

fact1 = getattr(h5file.root, prod1)[:]
label_array[:, i] = np.multiply(fact0, fact1)

# remove samples outside mass/pT window
if remove_mass_pt_window:
    feature_array = feature_array[(spec_array[:, 0] > 40) &
        (spec_array[:, 0] < 200) & (spec_array[:, 1] > 300) &
        (spec_array[:, 1] < 2000)]

    label_array = label_array[(spec_array[:, 0] > 40) &
        (spec_array[:, 0] < 200) & (spec_array[:, 1] > 300) &
        (spec_array[:, 1] < 2000)]

feature_array = feature_array[np.sum(label_array, axis=1)==1]
label_array = label_array[np.sum(label_array, axis=1)==1]
print(label_array)
print(len(feature_array[0]))

print(len(feature_array), len(label_array))

h5file.close()
return feature_array, label_array

# copy training file if it doesn't exist
import os.path
if not os.path.isfile('/kaggle/input/hbbqcd/ntuple_merged_10.h5'):
    !xrdec root://eospublic.cern.ch//eos/opensdata/cms/datascience/
    HiggsToBBNtupleProducerTool/
    HiggsToBBNtuple_HiggsToBB-QCD_RunII_13TeV_MC/
    train/ntuple_merged_10.h5 .

```

```

# load training file
feature_array , label_array =
    get_features_labels( '/kaggle/input/hbbqcd/ntuple_merged_10.h5' ,
        remove_mass_pt_window=False)

# define dense keras model
inputs = Input(shape=(nfeatures ,) , name = 'input')
x = BatchNormalization(name='bn_1')(inputs)
x = Dense(64 , name = 'dense_1' , activation='relu')(x)
x = Dense(32 , name = 'dense_2' , activation='relu')(x)
x = Dense(32 , name = 'dense_3' , activation='relu')(x)
outputs = Dense(nlabels , name = 'output' , activation='softmax')(x)
keras_model = Model(inputs=inputs , outputs=outputs)
keras_model.compile(optimizer='adam' , loss='categorical_crossentropy' ,
    metrics=['accuracy'])
print(keras_model.summary())

# define callbacks
early_stopping = EarlyStopping(monitor='val_loss' , patience=10)
model_checkpoint = ModelCheckpoint('keras_model_best.h5' ,
    monitor='val_loss' , save_best_only=True)
callbacks = [early_stopping , model_checkpoint]

# fit keras model
keras_model.fit(feature_array , label_array , batch_size=1024 ,
    epochs=1000 , validation_split=0.2 , shuffle=True ,
    callbacks = callbacks)

# copy testing file if it doesn't exist

```

```

import os.path
if not os.path.isfile('/kaggle/input/hbbqcd3/ntuple_merged_0.h5'):
    !xrdcp root://eospublic.cern.ch//eos/opendata/cms/datascience/
        HiggsToBBNTupleProducerTool/
        HiggsToBBNTuple_HiggsToBB-QCD_RunII_13TeV_MC/
        test/ntuple_merged_0.h5 .

# load testing file
feature_array_test , label_array_test =
    get_features_labels('/kaggle/input/hbbqcd3/ntuple_merged_0.h5')

# reload best weights
keras_model.load_weights('keras_model_best.h5')

# run model inference on test data set
predict_array_test = keras_model.predict(feature_array_test)

# create ROC curve
fpr , tpr , threshold = roc_curve(label_array_test[:,1] ,
    predict_array_test[:,1])

# plot ROC curve
plt.figure()
plt.plot(tpr , fpr , lw=2.5 , label="AUC= {:.1f}%" .format(auc(fpr , tpr)*100))
plt.xlabel(r 'True - positive - rate ')
plt.ylabel(r 'False - positive - rate ')
plt.semilogy()
plt.ylim(0.001,1)
plt.xlim(0,1)
plt.grid(True)

```

```
plt.legend(loc='upper-left')
plt.tight_layout()
plt.savefig('ROC.png')
plt.savefig('ROC.pdf')
```

Referencias

- [1] Particle Data Group. Review of Particle Physics, Progress of Theoretical and Experimental Physics, Volume 2022, Issue 8, August 2022, 083C01. <https://doi.org/10.1093/ptep/ptac097>. Consultada el 18/12/2023
- [2] The Standard Model. CERN. <https://home.cern/science/physics/standard-model>. Consultada el 15/12/2023
- [3] The atom. Institute of Physics. <https://www.iop.org/explore-physics/big-ideas-physics/atom>. Consultada el 15/12/2023
- [4] Hentschel, K. (2009). Atomic Models, J.J. Thomson's "Plum Pudding" Model. In: Greenberger, D., Hentschel, K., Weinert, F. (eds) Compendium of Quantum Physics. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-70626-7_9. Consultada el 15/12/2023
- [5] Rutherford, Ernest (1911). «The Scattering of α and β Particles by Matter and the Structure of the Atom». Philosophical Magazine. Series 6 21: 669-688. <https://doi.org/10.1080/14786440508637080>. Consultada el 15/12/2023.
- [6] Román, N. Sobre las interacciones fundamentales, las partículas elementales y las teorías de campos. <https://web.mat.upc.edu/narciso.roman/docs/buran1.pdf>. Consultada el 15/12/2023.
- [7] R.L. Workman et al. (Particle Data Group), Prog. Theor. Exp. Phys. 2022, 083C01 (2022). <https://doi.org/10.1093/ptep/ptac097>. Consultada el 15/12/2023.
- [8] ATLAS Collaboration (2012), Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1), págs. 1-29.
- [9] CMS Collaboration. (2012). Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Journal of High Energy Physics*, 2012(6), 81.

- [10] García, V. (2022). Estudio y dimensionamiento de instalación híbrida aislada con acumulación. Universidad Politécnica de Valencia. <https://riunet.upv.es/bitstream/handle/10251/183774/Garcia%20-%20Estudio%20y%20dimensionamiento%20de%20instalacion%20hibrida%20aislada%20con%20acumulacion.pdf?sequence=1>. Consultado el 15/11/2023.
- [11] <http://hyperphysics.phy-astr.gsu.edu/hbasees/Particles/expar.html>. Consultado el 15/11/2023.
- [12] Dreyer F., Schramm S. (2018). Introduction to Jet Substructure. https://indico.cern.ch/event/751034/contributions/3136819/attachments/1738045/2811803/VBSCan_Oct_2018_into.pdf. Consultado el 15/11/2023.
- [13] Chadwick, J. (1932). Possible Existence of a Neutron. *Nature*. <https://doi.org/10.1038/129312a0>. Consultada el 15/12/2023.
- [14] Anderson, C. D. (1933). The Positive Electron. *Phys. Rev.*, 43(6), 491-494. <https://doi.org/10.1103/PhysRev.43.491>. Consultada el 15/12/2023.
- [15] C. Lattes, G. Occhialini, H. Muirhead and C. Powell (1947). Processes Involving Charged Mesons. *Nature*. 159: 694–698. <https://doi.org/10.1007/s00016-014-0128-6>. Consultada el 15/12/2023.
- [16] Rabi, I. I. (1945). On the Process of Space Quantization. Nobel Lecture, December 12, 1944. <https://www.nobelprize.org/prizes/physics/1944/rabi/lecture/>. Consultada el 15/12/2023.
- [17] Weinberg, S. (1967). A Model of Leptons. *Physical Review Letters* 19 (21): 1264-1266. — The electroweak unification paper. <https://doi.org/10.1103/physrevlett.19.1264>. Consultada el 15/12/2023.
- [18] Liss, T.M. Tipton, P.L. (1997). The discovery of the top quark. *Scientific American* 277 (3): 54-59. <https://hep.physics.illinois.edu/home/tml/SciAmTop.pdf>. Consultada el 15/12/2023.
- [19] <https://spaceplace.nasa.gov/dark-matter/sp/>. Consultada el 15/12/2023.

- [20] Historia del CERN. <https://home.cern/about/who-we-are/our-history>. Consultada el 15/12/2023.
- [21] Estructura del LHC. <https://home.web.cern.ch/science/accelerators/large-hadron-collider>. Consultada el 15/12/2023.
- [22] Imagen de Stock. Recogida de <https://stock.adobe.com/es/images/large-hadron>. Consultada el 02/11/2023.
- [23] Experimentos en el LHC. <https://home.cern/science/experiments>. Consultada el 15/12/2023.
- [24] D.J. Buettner, P.D. Morley, Neutrinos as the Cold Dark Matter: A critical review, KATRIN and new research directions. Physics of the Dark Universe, 39, 2023. <https://doi.org/10.1016/j.dark.2022.101149>. Consultada el 10/01/2024.
- [25] LHC. <https://lhc-commissioning.web.cern.ch/schedule/LHC-long-term.htm>. Consultada el 03/11/2023.
- [26] El experimento CMS. <https://home.cern/science/experiments/cms>. Consultada el 15/12/2023.
- [27] Neutelings, I. CMS coordinate system. https://tikz.net/axis3d_cms/. Consultada el 15/11/2023.
- [28] CMS. <https://cms.cern/book/export/html/1618>. Consultada el 03/11/2023.
- [29] Public CMS Luminosity Information.
https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#Run_3_charts_of_luminosity. Consultada el 15/11/2023.
- [30] Gudacker Stefan et. al, Power-efficient high-frequency readout concepts of SiPMs for TOF-PET and HEP.
https://indico.cern.ch/event/1044975/contributions/4663809/attachments/2394614/4094100/VCI_2022_VF.pdf. Consultada el 03/11/2023.

- [31] CERN. LHC pp collisions. Taking a closer look at LHC. https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.lhc_p_collisions. Consultada el 10/01/2024.
- [32] Veatch, J. Searches for Resonant Scalar Boson Pair Production Using Run 2 LHC Proton-Proton Collision Data. *Symmetry* 2022, 14, 260. <https://doi.org/10.3390/sym14020260> Consultada el 21/01/2024.
- [33] Cocoros, A. Search for new physics in the two higgs to fourb quark final state. Johns Hopkins University. 2017. <https://jscholarship.library.jhu.edu/server/api/core/bitstreams/ff6fec15-c1e5-483a-83d2-15de4f1d6164/content>. Consultada el 22/01/2024.
- [34] List of Diagrams. Universität Zürich. <https://www.physik.uzh.ch/~che/FeynDiag/Listing.php>. Consultada el 22/01/2024.
- [35] K.J. Eskola, K. Kajantie, P.V. Ruuskanen, K. Tuominen, Rapidity dependence of particle production in ultrarelativistic nuclear collisions (2002). *Physics Letters B*, 543, págs. 208-216. <https://www.sciencedirect.com/science/article/pii/S0370269302024577>. Consultada el 03/11/2023.
- [36] Elaborada por Mario Carpintero, Universidad de Oviedo. https://www.hep.uniovi.es/jfernan/CMS_coord.pdf. Consultado el 04/11/2023.
- [37] CERN, CMS Top Summary Figures - Public Results https://twiki.cern.ch/twiki/pub/CMSPublic/PhysicsResultsTOPSummaryFigures/tt_curve_sqrts_cms.pdf. Consultada el 05/11/2023.
- [38] Minjung K. Inha University. (2017). b-jet tagging via track impact parameter. https://indico.cern.ch/event/687697/contributions/2825792/attachments/1579225/2494980/bjet_tagging_via_trackIP.pdf. Consultado el 07/02/2024.
- [39] CMS Collaboration. (2017). Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV <https://cds.cern.ch/record/2298594>. Consultada el 04/11/2023.

- [40] Barber, D. P. (1979). «Discovery of Three-Jet Events and a Test of Quantum Chromodynamics at PETRA». *Physical Review Letters* 43: 830. <https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.43.830>. Consultado el 16/11/2023.
- [41] E. D. Malaza, “Multiplicity distributions in quark and gluon jets”, *Zeitschrift für Physik C. Particles and Fields* 31 (1986) 143 – 150, <http://dx.doi.org/10.1007/BF01559605>. Consultado el 27/11/2023.
- [42] CMS Collaboration. (2013, September 11). Performance of quark/gluon discrimination using pp collision data at $\sqrt{s} = 8$ TeV. CMS Physics Analysis Summary. <https://cds.cern.ch/record/1599732/files/JME-13-002-pas.pdf>. Consultado el 27/11/2023.
- [43] Atkin, R. (2015). Review of jet reconstruction algorithms. *Journal of Physics: Conference Series*, 645, 012008. <https://iopscience.iop.org/article/10.1088/1742-6596/645/1/012008/pdf>. Consultado el 27/11/2023.
- [44] CMS Collaboration. (2023). Jet Reconstruction with the Seeded Cone algorithm in the CMS Phase-2 Level-1 Trigger, CERN. <https://cds.cern.ch/record/2859652>. Consultado el 27/11/2023.
- [45] OPAL Collaboration, “A Model independent measurement of quark and gluon jet properties and differences”, *Z.Phys. C*68 (1995) 179 – 202, <http://dx.doi.org/10.1007/BF01566667>. Consultado el 27/11/2023
- [46] DELPHI Collaboration, “Energy dependence of the differences between the quark and gluon jet fragmentation”, *Z.Phys. C*70 (1996) 179 – 196, <http://dx.doi.org/10.1007/s002880050095>. Consultado el 27/11/2023.
- [47] CERN. CMS Monte Carlo production overview. <https://opendata.cern.ch/docs/cms-mc-production-overview>. Consultado el 14/01/2024.
- [48] S. Agostinelli, et al. Geant4—a simulation toolkit, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8). Consultado el 19/01/2024.

- [49] Bierlich, C., et al. (2022). A comprehensive guide to the physics and usage of PYTHIA 8.3. SciPost Phys. Codebases, 8. <https://doi.org/10.21468/SciPostPhysCodeb.8> Consultada el 22/01/2024.
- [50] Matteo Cacciari et al. "The anti-kt jet clustering algorithm". (2008), <https://doi.org/10.48550/arXiv.0802.1189>. Consultado el 11/01/2024.
- [51] Large Hadron Collider Jet Clustering Visualization: Anti-kT. Animación realizada por Joel Walker, <https://www.youtube.com/watch?v=KUaRoWPRR2Q>. Consultado el 12/01/2024.
- [52] Casaus, J., Rodríguez, J., Sánchez, E. (2005) Cromodinámica Cuántica: el color de los quarks. Departamento de Fusión y Física de Partículas Elementales. CIEMAT. <http://fisica.ciens.ucv.ve/svf/Documentos/Feiasofi/cromodinamica.pdf>. Consultado el 16/11/2023
- [53] <https://www.aprendemachinelearning.com/breve-historia-de-las-redes>. Consultada el 04/11/2023.
- [54] <https://www.asimovinstitute.org/neural-network-zoo/>. Consultada el 05/11/2023.
- [55] <https://learn.microsoft.com/es-es/archive/msdn-magazine/2017/september/test-run-deep-neural-network-training>. Consultada el 05/11/2023.
- [56] Layer activation functions. <https://keras.io/api/layers/activations/>. Consultada el 13/12/2023.
- [57] Lederer, J. (2021) Activation Functions in Artificial Neural Networks: A Systematic Overview. <https://arxiv.org/abs/2101.09957>. Consultada el 05/11/2023
- [58] Duarte, J. (2019). Sample with jet, track and secondary vertex properties for Hbb tagging ML studies HiggsToBBNTuple_HiggsToBB_QCD_RunII_13TeV_MC. CERN Open Data Portal. Sample with jet, track and secondary vertex properties for Hbb tagging ML studies. <https://opendata.cern.ch/record/12102>. Consultada el 22/01/2024.
- [59] Herramienta Kaggle. <http://www.kaggle.com>. Consultada el 22/01/2024.