



UNIVERSIDAD NACIONAL DE LA PLATA
FACULTAD DE CIENCIAS EXACTAS
INSTITUTO DE FÍSICA LA PLATA

Trabajo de Tesis Doctoral

My thesis title

My subtitle

Autor:
Francisco SILI

Directores:

Prof. Dr. Maria Teresa DOVA
Dr. Francisco ALONSO

30 de septiembre de 2024

ACKNOWLEDGEMENTS

akwnoelasfsa

I, Francisco SILI, hereby declare that this thesis has not been and will not be,
submitted in whole or in part to another university for the award of any other degree.

La Plata,
30 de septiembre de 2024

Francisco SILI

Universidad Nacional de La Plata
Facultad de Ciencias Exactas
Instituto de Física La Plata

TESIS DE DOCTORADO

My thesis title

by Francisco SILI

ABSTRACT

some text

ÍNDICE GENERAL

Índice general	vi
Índice de figuras	ix
Índice de tablas	xi
Introduction	2
I Theory Motivation	3
1 The Standard Model and Beyond	4
II Experimental setup	5
2 El LHC y el experimento ATLAS	6
2.1. LHC	6
2.2. ATLAS	8
2.2.1. Sistema de coordenadas de ATLAS	9
2.2.2. Detector Interno	10
2.2.3. Calorímetros	12
Calorímetro Electromagnético (ECAL)	13
Calorímetro Hadronico (HCAL)	14
2.2.4. Espectrómetro de Muones (MS)	15
2.2.5. El sistema de Trigger	16
Level-1 (L1)	16
High Level Trigger (HLT)	17
2.3. Toma de datos durante el Run-2	18
3 Reconstrucción e identificación de objetos físicos	21
3.1. Reconstrucción de trazas y vértices	21
3.2. Fotones y electrones	23

3.2.1. Reconstrucción	23
3.2.2. Identificación	25
3.2.3. Aislamiento	25
3.3. Muones	27
3.4. Jets	28
3.4.1. Algoritmo de clusterización de jets anti- k_t	28
3.4.2. Jets Calorimétricos	29
3.4.3. Particle Flow (PFlow) Jets	30
3.4.4. Calibración de jets	31
3.5. Jets de sabor pesado (<i>heavy flavor</i>)	31
3.5.1. Identificación y performance de <i>b</i> -tagging	32
3.5.2. Identificación y performance de <i>c</i> -tagging	34
III Photon shower shape corrections	35
4 Photon identification and shower shapes	36
4.1. Shower shapes	36
4.2. Photon Identification	38
4.2.1. Optimisation	38
4.2.2. Efficiency measurements	39
4.3. Shower shapes variables differences between data and MC	40
4.4. Samples and event selection for the Shower Shape (SS) correction studies	41
4.4.1. Radiative Z boson decays	41
Special selection for cell-based reweighting corrections	42
4.4.2. Inclusive photons	42
5 Fudge factors	43
5.1. Calculation	43
5.1.1. Shift-only corrections	44
5.1.2. Shift+stretch corrections	45
5.2. Uncertainties	45
5.2.1. Statistical uncertainties	45
5.2.2. Systematic uncertainties	46
5.3. Results	47
6 Cell-based energy reweighting	48
IV New Physics	49
7 Analysis motivation and strategy	50
8 Signal and Background samples	51

9 Event selection	52
10 Background estimation	53
11 Background modeling	54
12 Systematic uncertainties	55
13 Statistical analysis	56
14 Results	57
Conclusions	58
Glossary	59
Bibliografía	62

ÍNDICE DE FIGURAS

2.1.	Vista general del complejo del Gran Colisionador de Hadrones (LHC) donde se muestran todos los aceleradores que conducen al LHC [3].	7
2.2.	Visión general del detector A Toroidal LHC ApparatuS (ATLAS) y de todos sus sub-detectores, incluidos los sistemas añadidos durante el Long Shut down 2 (LS2) [11].	8
2.3.	Sistema de coordenadas de ATLAS [11].	9
2.4.	Diagramas del Detector Interno (ID) que muestran los diferentes submódulos, con sus correspondientes dimensiones.	10
2.5.	Sistema de calorímetros de ATLAS, mostrando el Calorímetro Electromagnético (ECAL) y el Calorímetro Hadronico (HCAL) [20].	12
2.6.	Segmento del ECAL mostrando la disposición de las capas y celdas del calorímetro. Además, se muestran las dimensiones de las celdas en cada capa [7].	14
2.7.	Longitudes de radiación en función de $ \eta $ para cada capa del ECAL [7].	14
2.8.	Diagrama del Espectrómetro de Muones (MS) [7].	15
2.9.	Luminosity delivered by the LHC and recorded by ATLAS during the Run-2 [29] and Run-3 data taking periods. For Run-2, the fraction of data good for physics analyses is also displayed.	19
2.10.	Condiciones de pileup durante el Run-2 y Run-3.	20
3.1.	Esquema mostrando los parámetros usados para la reconstrucción de trazas [34]. .	22
3.2.	Diagrama mostrando el algoritmo de reconstrucción de electrones and fotones, extraído de Ref. [39]	24
3.3.	Diagrama mostrando el proceso del cálculo de la variable de aislamiento calorimétrico. Cuando se utiliza un cono con $R = 0.4$, se puede construir la variable E_T^{cone40} mencionada en el texto.	26
3.4.	Representación esquemática del algoritmo anti- k_t para el proceso de clusterización de jets [48].	29
3.5.	Pasos para la calibración del cuadrimomento de PFLow jets [51].	31

ÍNDICE DE TABLAS

2.1. Rendimiento deseado del detector ATLAS. Las unidades de p_T y E están en GeV. Extraído de Ref. [7]	9
3.1. Resumen de los WP de aislamiento para electrones y fotones usados a lo largo de esta tesis.	26
3.2. Medidas de eficiencias de b -tagging, y de rechazos de c -jets, l -jets y τ -jets, en los regímenes de bajo y alto p_T	34
3.3. Medidas de eficiencia de c -tagging efficiencies para c -jets, y valores de rechazos de b -jets, l -jets y τ -jets en los regímenes de bajo y alto p_T . Los valores corresponden a aquejlos luego de aplicar el veto del WP de b -tagging de 77% y de 50% de c -tagging. rejection values not correct!	34

TO DO'S AND NOTES TO KEEP IN MIND

use **orange** to highlight that there needs to be made sure that there is a discussion in previous chapters - in editing clarify where that discussion should happen!

purple: this needs a reference, have used from memory or notes

red: open question

Fixes, to dos

- test
- need to include a definition on met in the object definition part?

Thoughts to work with

- have to be consistent with times in the description - discuss with Fab about it

Might be good to answer for viva preps

- how is the reconstruction considered overall? are there different

INTRODUCTION

This thesis presents a search for new phenomena in high-mass final states with a photon and a jet in proton-proton (pp) collisions at a centre-of-mass energy of 13 TeV using data collected by the [ATLAS](#) detector. [ATLAS](#) (“A Toroidal LHC ApparatuS”) is one of the two general-purpose detectors at the Large Hadron Collider ([LHC](#)) and the biggest multi-purpose particle detector ever built. It is used to investigate a wide range of physics, from Standard Model ([SM](#)) measurements, such as precision tests of quantum chromodynamics or study of the properties of the Higgs boson, to the search of new phenomena like extra dimensions and dark matter candidates. The LHC, built by the European Organization for Nuclear Research ([CERN](#)) and installed in a 27-kilometre circular tunnel, is the world’s largest and most powerful particle collider. This machine is capable of colliding energetic beams of protons (or heavier nuclei) at rates upward of millions per second. The precision and high beam energy of the LHC allow to explore the tera-electronvolt scale, an energy range never before achieved in a particle collider.

The thesis is divided into three parts. The first part describes the theory background and motivations of the work, in which the first chapter describes the Standard Model ([SM](#)) showing the excellent agreement there is between the theory and the experimental data. It also shows the actual problems of the [SM](#), which motivates the search for New Physics.

The second part describes the experiment. [CHAPTER](#) describes the [ATLAS](#) experiment, with each sub-detector in detail, to finally describes how the samples for data analysis are produced. The reconstruction of the different physics objects is explained in [CHAPTER](#). At the end of the second part, a special chapter is dedicated to the photon identification, where a new method for correcting the shower shapes is explained and detailed.

The third part of the thesis is dedicated to the photon+jet resonance search. In [CHAPTER](#), the general strategy and motivation of the search is discussed. The samples generation is discussed in [CHAPTER](#), where the final event selection is given in [CHAPTER](#). The methods for the background estimation as well as background modeling is given in [CHAPTER](#) AND [CCHAPETER](#). The systematic uncertainties are discussed in [CHAPTER](#), and finally in [CHAPTER](#) and [CHAPTER](#), the statistical analysis with the results are shown.

Finally, the conclusions of the work are present in [CHAPTER](#).

Parte I

Theory Motivation

THE STANDARD MODEL AND BEYOND

1

*"Nothing in life is to be feared. It is
only to be understood. Now is the
time to understand more, so that we
may fear less"*

Marie Curie

another template text

Parte II

Experimental setup

EL LHC Y EL EXPERIMENTO ATLAS

2

Something.

Someone

El trabajo de esta tesis se ha realizado utilizando datos del detector [ATLAS](#), uno de los detectores de partículas que registran colisiones de protones acelerados por el Gran Colisionador de Hadrones ([LHC](#)) en el Organización Europea para la Investigación Nuclear ([CERN](#)). En el siguiente capítulo, se ofrece una introducción al [LHC](#) en la Section 2.1, seguida de una discusión del detector [ATLAS](#) en la Section 2.2. Finalmente, en la Section 2.3, se describe brevemente cuáles fueron las tomas de datos de [ATLAS](#), así como también sus propiedades. La discusión se centra en aspectos importantes para los análisis de esta tesis.

2.1. LHC

El [LHC](#) [1, 2] es el mayor acelerador de hadrones del mundo, situado en el [CERN](#), en la frontera franco-suiza. Tiene una longitud de 27 km y está situado entre 50 y 174 metros bajo tierra. El [LHC](#) está diseñado para hacer colisionar protones (e iones pesados) a una energía de centro de masa de 14 TeV. Para mantener los protones y los iones pesados en el anillo del acelerador, se utilizan un total de 9593 imanes. Estos imanes incluyen imanes superconductores dipolares y cuadrupolares, enfriados a 1,9 K (-271 °C), de los cuales los imanes dipolares generan un campo magnético de 8,3 T.

En la Figure 2.1 se muestra un esquema general de las instalaciones del acelerador [LHC](#). Los protones se obtienen de hidrógeno gaseoso eliminando sus electrones y se aceleran en un primer acelerador lineal (LINAC2) hasta 50 MeV. Posteriormente, los protones se aceleran sucesivamente en el Proton Synchrotron Booster ([PSB](#)), el Proton Synchrotron ([PS](#)) y el Super Proton Synchrotron ([SPS](#)), donde alcanzan una energía de 450 GeV antes de ser injectados en el [LHC](#). En el [LHC](#), en total, 8 cavidades de radiofrecuencia pueden impulsar la energía de los protones hasta 14 TeV. Los cuatro puntos amarillos en la imagen Figure 2.1 son cuatro puntos de interacción, que albergan los experimentos [ALICE](#) [4], [LHCb](#) [5], [CMS](#) [6], [ATLAS](#) [7], [LHCf](#) [8], [TOTEM](#) [9], [MoEDAL](#) [10], entre muchos otros.

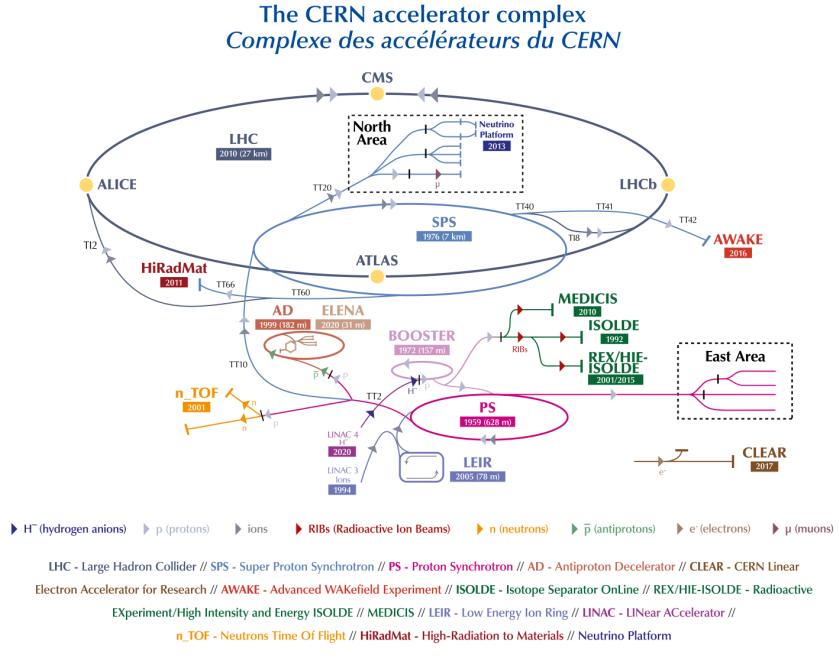


Figura 2.1: Vista general del complejo del LHC donde se muestran todos los aceleradores que conducen al LHC [3].

Los protones se inyectan en haces de $\mathcal{O}(10^{11})$ protones en el LHC con una separación de 25 ns (7,5 m). Estos haces se llevan posteriormente a colisión en los llamados cruces de haces (bunch-crossing). El esquema de llenado de la cadena del preacelerador, en combinación con los tiempos de conmutación finitos de los imanes de inyección y descarga, da lugar a patrones regulares de paquetes llenos y vacíos.

Uno de los parámetros más importantes para caracterizar el funcionamiento del acelerador es la luminosidad instantánea \mathcal{L} , definida como el número de partículas por unidad de tiempo por unidad de superficie, y puede calcularse a partir de la relación

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} F \quad (2.1)$$

donde N_b es el número de partículas por bunch, n_b el número de bunches por haz, γ_r es el factor gamma relativista, ϵ_n es la emitancia transversal normalizada del haz y β^* es la función beta en el punto de colisión que determina la dispersión transversal del haz de partículas. El término de corrección F tiene en cuenta el ángulo de cruce del haz. La frecuencia de revolución está representada por f_{rev} que es de ~ 11 kHz, y con el espaciado del haz de 25 ns, permite el cruce del haz en los cuatro puntos de interacción con una frecuencia de ~ 40 MHz.

La medida para el total de datos registrados es la luminosidad integrada a lo largo del tiempo viene dada por

$$N_{event} = L_{int} \sigma_{event} = \sigma_{event} \int \mathcal{L} dt. \quad (2.2)$$

Esta variable relaciona la luminosidad con el número de eventos. Más detalles sobre las mediciones de luminosidad en ATLAS se muestran en Section 2.3.

2.2. ATLAS

[ATLAS](#) es uno de los detectores multipropósito del [LHC](#), situado en *Point-1* en el [LHC](#). Fue diseñado y construido para estudiar las colisiones pp (y de iones pesados) a escala TeV.

La forma general del detector es la de un cilindro, como se muestra en la Figure 2.2. Tiene una longitud de 44 m y 25 m de diámetro, siendo el mayor detector de partículas construido hasta la fecha. El detector [ATLAS](#) está dividido geométricamente en dos partes: la parte central llamada *barrel*, y las tapas exteriores llamadas *end-caps*.

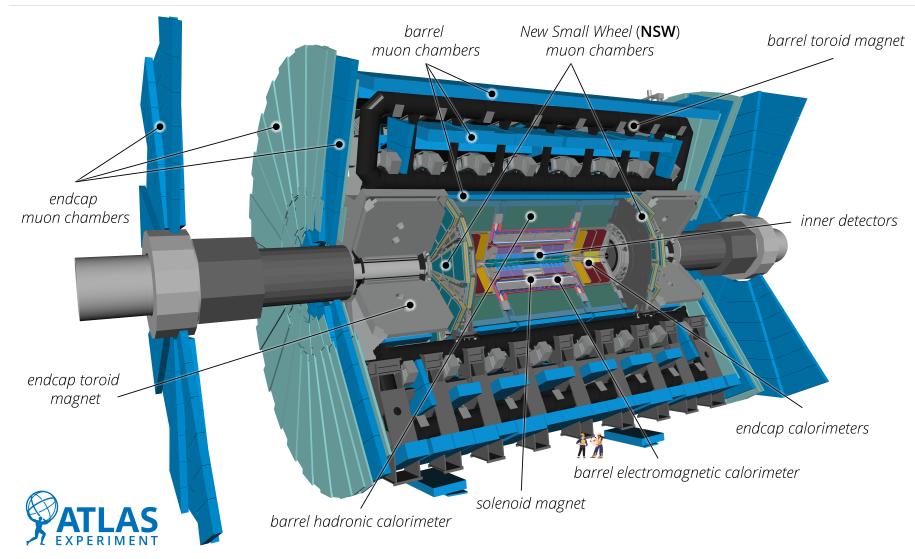


Figura 2.2: Visión general del detector [ATLAS](#) y de todos sus subdetectores, incluidos los sistemas añadidos durante el [LS2](#) [11].

[ATLAS](#) está construido en capas de subdetectores, cada uno de los cuales está diseñado para tener un papel diferente en la identificación y reconstrucción de las partículas producidas en las colisiones. [ATLAS](#) proporciona una cobertura hermética alrededor del eje del haz, permitiendo la detección de todas las partículas cargadas generadas en las colisiones en el plano ortogonal al eje del haz. Esto es particularmente importante en las búsquedas de nueva física, que se basan en análisis de balances de momento en el plano ortogonal.

Está formado por múltiples capas, empezando por el componente más interno, el Detector Interno ([ID](#)), que permite reconstruir trazas cerca del tubo del haz. Alrededor del [ID](#), hay un solenoide superconductor que crea un campo magnético axial de ~ 2 T para curvar las trazas en el [ID](#) de las partículas cargadas. Tras el primer imán, hay un sistema de dos calorímetros: el Calorímetro Electromagnético ([ECAL](#)) y el Calorímetro Hadronico ([HCAL](#)). El primero se encarga de medir la energía cinética de fotones y electrones, y el segundo mide la energía de los jets. Las partes más externas del [ATLAS](#) están construidas por el Espectrómetro de Muones ([MS](#)), que proporciona la reconstrucción del momento de los muones que atraviesan las capas internas del detector. Entrelazadas con el [MS](#), hay un total de 8 bobinas toroidales, que proporcionan un campo magnético total de 4 T (0,5 T por bobina) para medir el momento de los muones. El campo magnético de los toroides se completa con los toroides en las regiones del end-cap, que

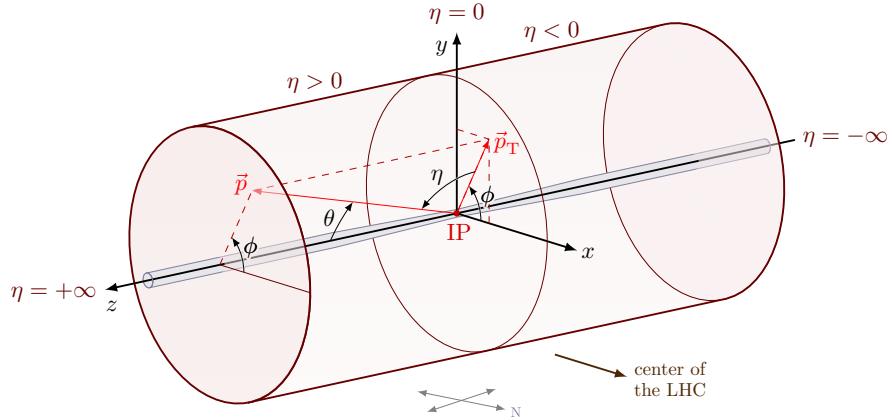
Tabla 2.1: Rendimiento deseado del detector [ATLAS](#). Las unidades de p_T y E están en GeV. Extraído de Ref. [7]

Componente del detector	Resolución requerida	Cobertura en η	
		Offline	Trigger
ID	$\sigma_{p_T}/p_T = 0.05\% p_T \oplus 1\%$	± 2.5	
ECAL	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$	± 3.2	± 2.5
HCAL (jets) barrel y end-cap dirección forward	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$ $\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$	± 3.2 $3.1 < \eta < 4.9$	± 3.2 $3.1 < \eta < 4.9$
MS	$\sigma_{p_T}/p_T = 10\% \text{ at } p_T = 1 \text{ TeV}$	± 2.7	± 2.4

también generan un campo magnético de hasta 4 T para los muones que salen en la dirección del haz.

El trabajo conjunto de todos los componentes de [ATLAS](#) permite reconstruir e identificar una gran variedad de partículas con gran precisión. En la Table 2.1, adaptado de Ref. [7], se da una visión general de las capacidades de diseño de [ATLAS](#) en términos de resolución de momento y energía. Aquí, la resolución está dada primero por un término estocástico, que mide la incertidumbre basada en la interacción de una partícula con el material, seguido de un término de ruido, que da cuenta de las incertidumbres debidas al ruido electrónico en el proceso de lectura.

2.2.1. Sistema de coordenadas de ATLAS

Figura 2.3: Sistema de coordenadas de [ATLAS](#) [11].

El sistema de coordenadas utilizado en [ATLAS](#), que se muestra en Figure 2.3, se utiliza en toda esta tesis y se describe brevemente a continuación [7]. El origen del sistema de coordenadas está en el punto de interacción nominal, con el eje x positivo apuntando hacia el centro del [LHC](#). El plano x-y es perpendicular al eje del haz, definiendo el eje z. Hacia la superficie define el eje y positivo. Alrededor del eje del haz se define un ángulo azimutal ϕ , y un ángulo polar θ es el ángulo desde el eje del haz. En lugar de θ se utiliza la rapidez y para objetos pesados tiene la forma:

$$y = \frac{1}{2} \ln[(E + p_z)/(E - p_z)]. \quad (2.3)$$

Las diferencias en la rapidez son invariantes a *boosts* a lo largo del eje del haz. Para objetos sin masa o relativistas ($m \ll p$) se utiliza en su lugar la pseudorapidez:

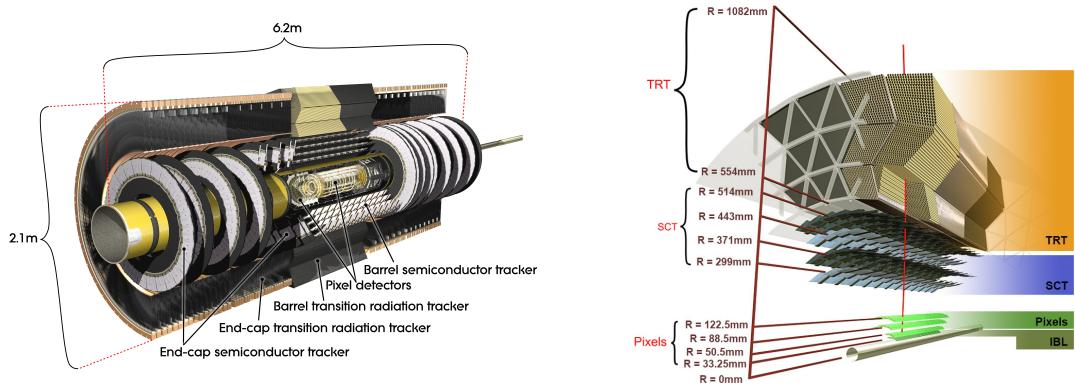
$$\eta = -\ln(\tan(\theta/2)). \quad (2.4)$$

Para cuantificar la distancia entre dos objetos, se define ΔR :

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} \quad (2.5)$$

El momento transverso y la energía se definen en el plano x-y, con el momento transverso dado como $p_T = \sqrt{p_x^2 + p_y^2}$.

2.2.2. Detector Interno



(a) El **ID** con todos sus submódulos en las regiones de barrel y end-cap. [12].

(b) Capas del **ID** mostrando su distancia al haz [13].

Figura 2.4: Diagramas del **ID** que muestran los diferentes submódulos, con sus correspondientes dimensiones.

El esquema de un corte transversal del Detector Interno (**ID**) [14] se muestra en la Figure 2.4, resaltando la distancia de cada subsistema respecto al tubo del haz. La parte más interna del **ID** se denomina *Insertable B-Layer* (**IBL**), seguido de tres capas de detectores de píxeles. A 299 mm de distancia radial del tubo del haz, cuatro capas de módulos del Semiconductor de Trazas (**SCT**) se sitúan antes del Detector de Radiación de Transición (**TRT**), que amplía el tamaño total del **ID** a un radio de 1082 mm. El **ID** permite la reconstrucción de las trazas de partículas en un rango de $|\eta| < 2.5$.

La función del **ID** es la recreación de las trazas de las partículas cargadas para determinar su carga y momento. Está inmerso en un campo magnético de 2 T generado por el sistema magnético del solenoide de **ATLAS**, que curva las trayectorias de las partículas cargadas. El radio de curvatura es proporcional al momento de la partícula y su dirección distingue las cargas positivas de las negativas. Las trazas de las partículas detectadas permiten reconstruir los vértices de colisiones primarias, lo cual es importante para distinguir las colisiones de *pile-up* (término que será descrito más adelante) de las colisiones de interés, y los vértices secundarios de decaimiento producidos por partículas de vida media larga, lo cual es crucial para la identificación de, por ejemplo, mesones B o leptones τ . A continuación, una breve descripción de cada parte del **ID** será dada.

Insertable B-Layer (IBL) Después del Run-1, durante el Long Shut down 1 ([LS1](#)) en el período de 2013-2014, el sistema detector de píxeles fue sometido a mantenimiento y actualizaciones. Dentro de este conjunto de actualizaciones, una cuarta capa de píxeles a 3,3 cm de distancia del tubo del haz fue instalada [15, 16] y ha permitido mejoras significativas en la reconstrucción de vértices de interacción y la identificación de jets iniciados por quarks b .

Detector de Píxeles La capa de píxeles más interna, el [IBL](#), está rodeada por tres capas de detectores de píxeles, dispuestas alrededor del tubo del haz [17, 18]. El método de detección de partículas cargadas es la medición de cargas inducidas depositadas en una capa de silicio, producto de la ionización. La primera capa se encuentra a una distancia de 50,5 mm del centro del tubo del haz. Como se puede ver en la [Figure 2.4a](#), en la región del end-cap los detectores de píxeles consisten en 3 discos alrededor del tubo, aumentando la longitud del detector de píxeles del [ID](#) a 1,4 m a lo largo del eje del haz. El detector de píxeles consta de un total de 1,744 módulos de píxeles con un tamaño nominal de $50\mu m \times 400\mu m$ en el plano (ϕ, z) , que comprenden más de 80 millones de canales de lectura. La parte de píxeles y [IBL](#) del detector [ATLAS](#) es crucial para la reconstrucción de trazas, ya que proporciona 4 puntos de medición (*hits*) en todo el rango de cobertura de pseudorapidez ($|\eta| < 2.5$.).

Semiconductor de Trazas (SCT) El detector de píxeles y [IBL](#) se encuentran dentro de los módulos del [SCT](#) [?]. Al igual que los módulos detectores de píxeles, los módulos del [SCT](#) están basados en semiconductores, dispuestos en capas cilíndricas alrededor del tubo del haz en la región del barrel, formando discos en los end-caps. Dado que los módulos del [SCT](#) sólo proporcionan una localización precisa a lo largo de un eje, se combinan dos módulos uno detrás de otro y rotados entre sí para obtener información espacial bidimensional. En el barrel hay cuatro capas y en los end-caps, nueve discos en cada lado (véase la [Figure 2.4a](#)). Incluyendo los discos de los end-caps, el [SCT](#) se extiende hasta $|z| < 2735\text{mm}$.

Detector de Radiación de Transición (TRT) La última parte del [ID](#) es el [TRT](#) [19], el cual en la región barrel se extiende de 554 mm a 1082 mm de distancia radial. Este detector se compone de tubos detectores de 4 mm de diámetro, dispuestos en paralelo al tubo del haz en la región barrel, y radialmente en los end-caps. En el rango de $|\eta| < 2.0$, se sitúan tres anillos en el barrel y 18 unidades en los end-caps, proporcionando típicamente 36 impactos por traza. Los tubos están entrelazadas con fibras de polipropileno, que cuando las partículas las atraviesan, crean la radiación de transición. En el interior de los tubos hay un fino cable de tungsteno que recoge las cargas. El nivel de radiación y las cargas recogidas en cada tubo pueden utilizarse para discriminar entre electrones y piones cargados. El [TRT](#) sólo ofrece información espacial en el plano $(R - \phi)$, y no se puede extraer información en la dirección z debido a la orientación de estos tubos. Hay un total de 50000 tubos en la región del barrel, mientras que en los end-caps se sitúan 320000 tubos.

2.2.3. Calorímetros

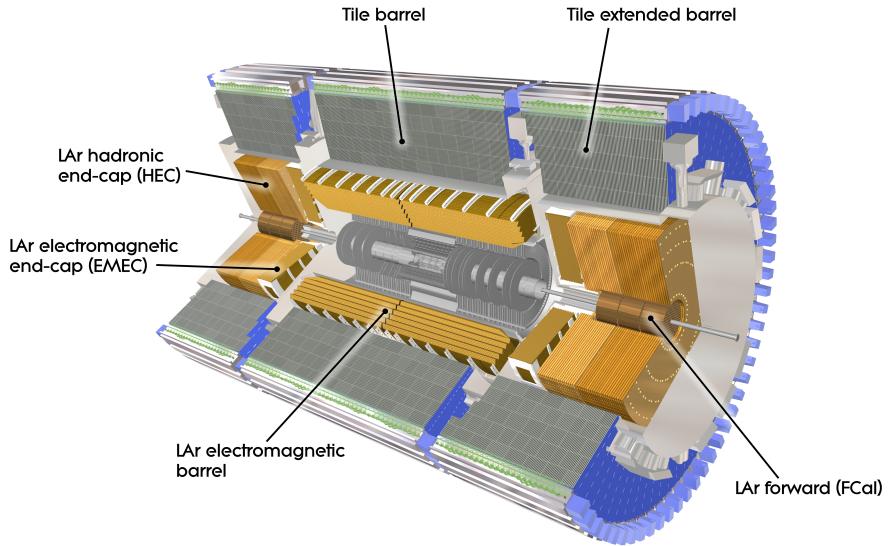


Figura 2.5: Sistema de calorímetros de ATLAS, mostrando el Calorímetro Electromagnético (ECAL) y el Calorímetro Hadronico (HCAL) [20].

Como se ha mencionado anteriormente, el sistema ID está rodeado por dos calorímetros: el Calorímetro Electromagnético (ECAL) y el Calorímetro Hadronico (HCAL), como se muestra en la Figure 2.5. Estos calorímetros están diseñados para medir la energía y la posición de las partículas incidentes, a través de la energía depositada por las cascadas de partículas secundarias producidas por las incidentes. Cubre todo el rango ϕ y hasta el $|\eta| < 4.9$, con una granularidad más fina en la región que coincide con el ID. El sistema de calorímetro permite discriminar entre fotones y electrones de hadrones (jets). Además, permite medir el desequilibrio energético (gracias a su cobertura total y hermiticidad) y proporciona al sistema de trigger la información necesaria para la selección de eventos.

Ambos calorímetros son denominados calorímetros de muestreo, con capas alternas de material absorbente y activo. La capa absorbente desencadena una lluvia de interacciones consecutivas con el material detector, la capa activa detecta la señal. El desarrollo de la lluvia y sus propiedades son de vital importancia para la identificación de las partículas, como se verá más adelante. Dos magnitudes importantes en relación con los calorímetros son la longitud de radiación, X_0 , y la longitud de interacción λ . La longitud de radiación se refiere a la distancia después de la cual la energía de una partícula (electrones por ejemplo) se ha reducido a $1/e$ de su energía inicial. La longitud de interacción describe el camino libre medio antes de que se produzca una interacción hadrónica.

La resolución de diseño del sistema sobre la energía calorimétrica viene dada por

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E}} \oplus b \oplus \frac{c}{E} \quad (2.6)$$

donde \oplus significa que los términos se suman en cuadratura. El término estocástico $\frac{a}{\sqrt{E}}$ está relacionado con las fluctuaciones en los desarrollos de la lluvia, el término constante b tiene

en cuenta las inhomogeneidades del detector, y el último término está asociado con el ruido electrónico y es proporcional a $\frac{1}{E}$. El valor de los coeficientes a y b depende de los objetos incidentes. Para el caso de los electrones en el **ECAL**, $a \sim 10\% \text{ GeV}^{1/2}$ y $b \sim 0.7\%$, mientras que los de los piones cargados en el centro del detector son $a \sim 50\% \text{ GeV}^{1/2}$ y $b \sim 5\%$ [21].

Calorímetro Electromagnético (**ECAL**)

El **ECAL** está especializado en la detección de electrones, positrones y fotones, que depositan su energía en lluvias relativamente densas: electrones energéticos que irradian fotones Bremsstrahlung, mientras que los fotones energéticos se convierten en pares electrón-positrón al atravesar el material denso. El material absorbente está hecho de plomo (Pb) con láminas de acero inoxidable, mientras que el Argón Líquido (**LAr**) se utiliza como material activo con electrodos de cobre y kapton para la lectura.

El calorímetro tiene una geometría de acordeón que proporciona una simetría completa ϕ sin fisuras azimutales. Está dividido en dos medios barriles que cubren la región central del detector ($|\eta| < 1.475$), con un pequeño hueco (4 mm) en $z = 0$ y una tapa final a cada lado del haz ($1.375 < |\eta| < 3.2$). La región de transición entre el barrel y end-cap se denomina región *crack*, y la mayoría de los análisis físicos que utilizan el **ECAL** requieren que los fotones y electrones se encuentren fuera de ella. Además, el **LAr** se utiliza para las tapas de los calorímetros hadrónicos, así como en el Calorímetro Forward (**FCAL**) ($3.1 < \eta < 4.9$).

El grosor de **ECAL** es superior a 22 longitudes de radiación (X_0) en la región del barrel, mientras que es superior a $24X_0$ en la región de end-caps. En el caso de los fotones, la distancia a la que la energía baja a $1/e$ es de $9/7X_0$, por lo que toda la energía electromagnética del fotón se deposita en el **ECAL**, y sólo una pequeña parte llega al **HCAL**.

El modo de medición es el siguiente. Las partículas incidentes interactúan con el medio absorbente (Pb), iniciando una lluvia de partículas cargadas y neutras. Las partículas cargadas ionizan el **LAr** y los electrodos, con la ayuda de un campo magnético aplicado, recogen los electrones producidos en el proceso de ionización. La señal total del medio activo es entonces proporcional a la energía real total de la partícula incidente.

Dentro de la región aceptada para las medidas de precisión ($|\eta| < 2.5$ excluyendo el crack), el **ECAL** se segmenta en tres capas longitudinales, mostradas en la Figure 2.6. La primera capa consiste en bandas de granularidad fina (también llamada *strip layer*) que ayuda a discriminar entre fotones aislados y pares de fotones espacialmente cercanos procedentes de decaimientos $\pi^0 \rightarrow \gamma\gamma$. Esta capa tiene un espesor constante de $\sim 6X_0$ en función de η (véase la Figure 2.7), y proporciona una medida precisa de esta variable. Para los fotones y electrones de alta energía, la mayor parte de su energía se recoge en la segunda capa, que tiene una granularidad lateral de 0.025×0.025 en (η, ϕ) y un espesor de $\sim 24X_0$. La tercera capa recoge la energía depositada por las colas de la lluvia electromagnética, con un espesor que varía entre 2 y $12X_0$. También hay un *presampler* (no se muestra en las figuras), que cubre la región $|\eta| < 1.8$ que mejora la medición de la energía para las partículas que comienzan la lluvia antes de entrar en el calorímetro.

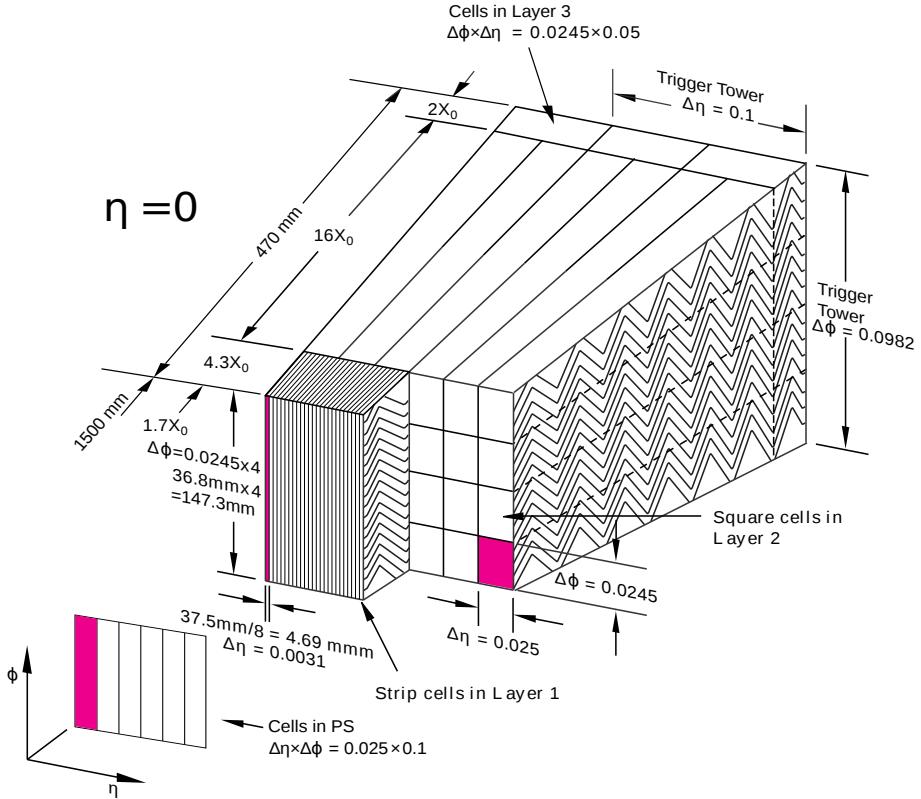


Figura 2.6: Segmento del [ECAL](#) mostrando la disposición de las capas y celdas del calorímetro. Además, se muestran las demisiones de las celdas en cada capa [7].

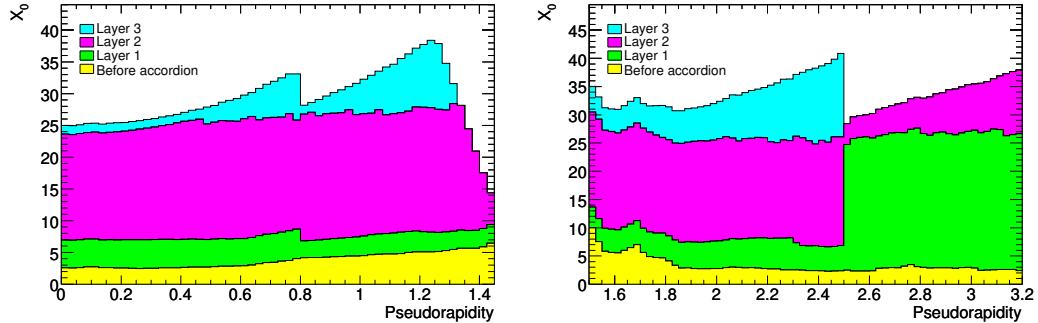


Figura 2.7: Longitudes de radiación en función de $|\eta|$ para cada capa del [ECAL](#) [7].

Calorímetro Hadrónico (HCAL)

Tres capas de calorímetro hadrónico rodean el [ECAL](#) y proporcionan discriminación adicional para electrones y fotones al medir la energía hadrónica. El [HCAL](#) se extiende en pseudorapidez hasta $|\eta| < 4.9$, permitiendo cubrir prácticamente la totalidad del ángulo sólido desde el punto de interacción. En la región del barrel ($|\eta| < 1.7$) se encuentra el primer calorímetro, el *Tile calorimeter*, un calorímetro de muestreo que utiliza acero como material absorbente y tejas centelladoras como material activo [?]. Está dividido en dos partes: $|\eta| < 1.0$ y $0.8 < |\eta| < 1.7$. Las tejas centelleadoras están dispuestas de una forma periódica y están conectadas a una fibra óptica que transporta la luz producida por las partículas que pasan a un tubo fotomultiplicador. Este arreglo se extiende, en R , de 2.28 a 4.25 m. En la región del end-cap ($1.5 < |\eta| < 3.2$)

hay un calorímetro de muestreo hadrónico, el Calorímetro Hadrónico del End-Cap (**HEC**), con placas de cobre como absorbente y argón líquido como material activo. Cada lado del endcap consiste en dos ruedas, una detrás de la otra con las placas planas de Cu dispuestas perpendicularmente al eje del haz, con un radio de 2,3 m. Finalmente está el **FCAL**, un calorímetro de muestreo que extiende la cobertura del sistema hasta $|\eta| < 4.9$, coaxial al eje del haz y situado a 4,7 m a cada lado del punto de interacción. El material principal de los módulos es el **LAr** (con cobre o tungsteno), y aunque no se utiliza para mediciones de precisión, proporciona información para el cálculo de la energía transversa faltante y la reconstrucción de jets en regiones muy cercanas al eje del haz.

El **HCAL** tiene un espesor superior a 7.7λ en la región del barrel (9.7 λ en total si se cuenta el **ECAL**). Análogamente a la longitud de radiación mencionada para el **ECAL**, se puede definir la longitud de interacción hadrónica como la distancia media a lo largo de la cual la energía de un hadrón se reduce a $1/e$ de su energía inicial. Así, toda la energía con la que los hadrones llegan al **HCAL** se deposita allí.

2.2.4. Espectrómetro de Muones (**MS**)

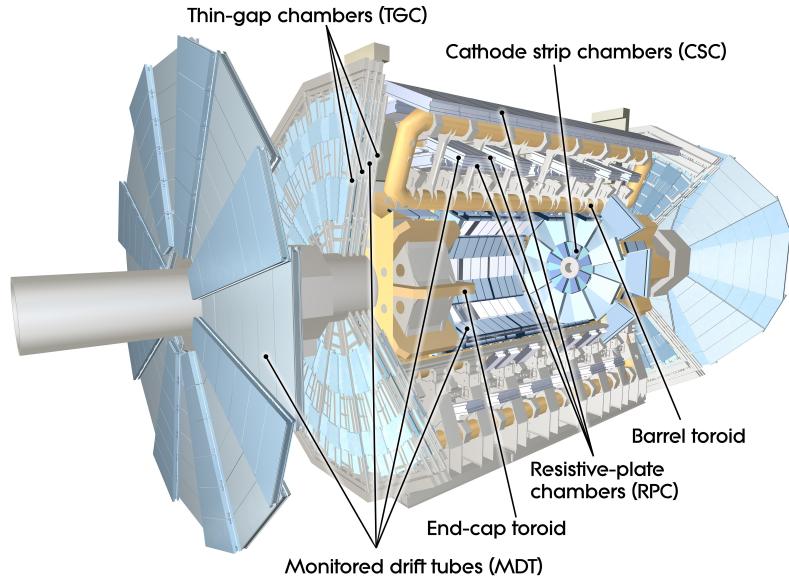


Figura 2.8: Diagrama del Espectrómetro de Muones (**MS**) [7].

Los muones de alto p_T generados en el punto de interacción tienen un poder de penetración muy elevado y son poco interactivos. Por lo tanto, el **MS** [22] está situado en la parte más externa del detector **ATLAS**, incrustado dentro del campo magnético de 4 T generado los imanes toroidales del barrel y end-caps, y está diseñado para obtener medidas de posición y momento de alta precisión de los muones de alto p_T en un rango de $|\eta|$ de $|\eta| < 2.7$. Se trata del mayor subdetector y el que da a **ATLAS** su tamaño. Éste subdetector se muestra en la Figure 2.8, destacando los subsistemas.

El **MS** se compone de diferentes tipos de cámaras de detección (véase la Figure 2.8). Los *Monitored Drift Tubes* (**MDTs**) son responsables de la mayoría de las medidas de precisión y cubren el

rango de pseudorapidez hasta $|\eta| < 2.7$. Funcionan de forma similar al [TRT](#), con tubos llenos de un gas ionizante y un ánodo central que recoge los electrones producidos, y el tiempo de deriva está asociado a la distancia a la traza dejada por la partícula. En la región del endcap se encuentran las *Cathode Strip Chambers* ([CSCs](#)) que tienen una alta resolución espaciotemporal y una cobertura de $|\eta| > 2.0$. Estas cámaras funcionan midiendo la carga depositada en un ánodo como resultado de la cascada de electrones creada cerca de él. Las *Resistive-Plate Chambers* ([RPCs](#)) proporcionan una estimación rápida del momento de los muones a nivel de trigger con una cobertura de $|\eta| < 1.05$ ¹. Las [RPCs](#) miden la descarga entre dos placas resistivas paralelas sometidas a una elevada diferencia de potencial, siguiendo la ionización del volumen de gas interno causada por el paso de muones energéticos. Por último, en la región del endcap, se encuentran los *Thin-Gap Chambers* ([TGCs](#)), de función similar a los [CSCs](#). También proporcionan información al sistema de trigger en esta región y tienen una cobertura de $|\eta| < 2.4$.

Si los hits en el [ID](#) y el [MS](#) se pueden asociar a un solo muón, se obtiene una muy buena resolución del momento de hasta

$$\frac{\sigma(p_T)}{p_T} = 0.02\% \cdot p_T [\text{GeV}] \oplus 2\%, \quad (2.7)$$

la cual se degrada si sólo se identifica una traza en uno de los dos sistemas.

2.2.5. El sistema de Trigger

El sistema de trigger de [ATLAS](#) [24–26] utiliza información del detector para rechazar eventos que no poseen una física interesante (física ya conocida, por ejemplo), reduciendo la frecuencia de eventos de 40 MHz (frecuencia de cruce de bunches mencionada en la [Section 2.1](#)) a alrededor de 1,5 kHz. Es necesario enfatizar aquí el papel central del sistema de trigger para el correcto funcionamiento de todo el experimento, siendo el responsable de decidir qué eventos se guardan y, en última instancia, qué física se encontrará (o no) durante el análisis de eventos. Sin un sistema de trigger eficiente, todos los subdetectores descritos anteriormente se desperdiciarían. Para lograr tal reducción en la frecuencia de eventos y, al mismo tiempo, tener una alta eficiencia en la selección de los de interés, el sistema de trigger se compone de dos niveles consecutivos capaces de realizar una identificación de partículas cada vez más compleja; un primer nivel de trigger basado en hardware, el Level-1 ([L1](#)), y luego un trigger de alto nivel basado en software, el High Level Trigger ([HLT](#)). Cada nivel permite analizar los eventos con mayor detalle, aumentando la precisión de los criterios de selección y la complejidad de los algoritmos utilizados.

Level-1 ([L1](#))

La decisión del trigger comienza con el [L1](#), basado en hardware [27], que identifica lo que se conoce como Región de interés ([ROI](#)). La [ROI](#) consiste en celdas vecinas en los [ECAL](#) y [HCAL](#),

¹ Durante el [LS2](#), la capa del end-cap más interna ha sido reemplazada por las *New Small Wheels* ([NSWs](#)) [23]. Presenta MicroMegas como rastreadores de precisión ya que proporcionan un mejor rendimiento a las altas tasas esperadas en las futuras operaciones del LHC.

y se define a partir de la posición en el calorímetro de cada objeto encontrado en un evento potencialmente interesante, que se extiende como un cono desde el punto de interacción a lo largo del detector. En cuanto a los muones, toma la información leída por el [MS](#), más concretamente por el [TGC](#) y el [RPC](#), y permite obtener una estimación rápida del p_T del muón. El [L1](#) también tiene una componente que permite tener en cuenta los requisitos topológicos, como las selecciones de masa invariante y las medidas de distancia, denominado el Level-1 Topológico ([L1Topo](#)).

El diseño del [L1](#) permite tener una aceptabilidad en el rango de $|\eta| < 2.5$ para electrones, fotones, muones y taus, hasta $|\eta| < 3.2$ para jets, y $|\eta| < 4.9$ para el cálculo del momento transverso faltante. Utilizando las [ROIs](#), el trigger [L1](#) debe tomar la decisión de guardar o descartar el evento, reduciendo la tasa de eventos de 40 MHz a menos de 100 kHz en aproximadamente $2,5 \mu\text{s}$, tiempo determinado en parte por el tamaño limitado de los buffers de memoria y en parte por el tiempo que tardan los muones producidos en el evento en llegar al [MS](#). Esta decisión final la toma el *Central Trigger Processor* ([CTP](#)), y luego pasa las [ROIs](#) al siguiente nivel de trigger: el [HLT](#).

High Level Trigger ([HLT](#))

Cuando un evento es aceptado por el [L1](#), el [HLT](#) [28] ejecuta una secuencia de algoritmos a partir de las [ROIs](#) definidas por el [L1](#), y permite reducir la tasa de eventos que se almacena a 1,5 kHz en 0,2 s. La reconstrucción e identificación de partículas candidatas en el [HLT](#) se evalúa en una secuencia de pasos donde se aplican diferentes algoritmos. Si la selección falla en un determinado paso, los pasos siguientes ya no se ejecutan para ahorrar tiempo de ejecución. En el [HLT](#), los algoritmos se agrupan en conjuntos de algoritmos de reconstrucción rápida que se ejecutan en primer lugar y, a continuación, se ejecuta un conjunto de algoritmos de reconstrucción de precisión similares a los utilizados *offline*. Los algoritmos de reconstrucción rápida utilizan la información del calorímetro y de las trazas del [ID](#) sólo dentro de la [ROI](#) para realizar la selección e identificación de candidatos, y llevar a cabo el rechazo del fondo lo más rápido y pronto posible. Si la partícula candidata supera los criterios definidos por la selección de reconstrucción rápida, se ejecutan los algoritmos de selección de precisión. Éstos tienen acceso a la información del detector fuera de la [ROI](#), con la máxima granularidad e incluyendo detalles sobre la calibración energética del calorímetro, la alineación del subdetector y el mapeo del campo magnético.

La secuencia exacta y el tipo de algoritmos considerados en el [HLT](#) se definen en el trigger *menu*. Esto comprende una base de datos de triggers, cada uno de los cuales define una secuencia de algoritmos y los requisitos de estos algoritmos para que un evento pase el [HLT](#).

Los requisitos de trigger se diseñan y presupuestan de forma que la tasa global del [HLT](#) no supere 1 kHz. En algunos casos, incluso la reducción de la tasa de eventos conseguida mediante los algoritmos del [HLT](#) para los requisitos de trigger deseados, como los trigger para objetos con bajo momento, es demasiado alta. Para mantener la tasa general del [HLT](#) por debajo de 1 kHz

en estos casos, los triggers pueden seguir incluyéndose en el menú, pero con una preescala. Un preescalado es un escalado artificial del trigger, que sólo acepta la N-ésima decisión de trigger si el factor de preescalado es N. Esto permite que los triggers con una alta tasa sigan recogiendo eventos.

Los algoritmos del [HLT](#) se ejecutan en aproximadamente 40.000 núcleos de CPU. Además, la construcción parcial de eventos se utiliza para el análisis a nivel de trigger, el monitoreo del detector y las calibraciones del subsistema detector. Finalmente, los eventos aceptados por el [HLT](#) se guardan en un disco y se distribuyen, disponibles *offline* para cualquier estudio o análisis.

2.3. Toma de datos durante el Run-2

El funcionamiento del [LHC](#) se organiza en distintos períodos conocidos como *runs*. Cada run suele durar varios años y se caracteriza por condiciones experimentales específicas, como la energía a la que colisionan los protones y la intensidad de los haces. Desde su puesta en marcha, el [LHC](#) se ha sometido a múltiples runs: Run-1 (2010-2013) operó a energías de colisión de hasta 8 TeV, Run-2 (2015-2018) a 13 TeV, y Run-3 (2022-presente) a 13,6 TeV. Cada período de toma de datos, una vez que el [LHC](#) anuncia haces estables, se divide en Bloques de Luminosidad ([LB](#)) de aproximadamente dos minutos. En cada [LB](#), la luminosidad instantánea es prácticamente constante y las condiciones del haz son estables. Debido a la alta complejidad del [LHC](#) y del detector [ATLAS](#), se espera que haya ineficiencias en los detectores y subdetectores y/o en la cadena de adquisición de datos. Durante cada run, cada parte del [ATLAS](#) es monitoreada y cualquier falla o problema es registrado, incluyendo componentes inactivos, o problemas en el haz del [LHC](#).

Para garantizar la alta calidad de los datos, libres de defectos significativos, los [LB](#) y los rangos dentro de ellos que superan todos los criterios de calidad se compilan en *Good Runs List (GRL)*. Las listas se elaboran y distribuyen de forma centralizada, con el fin de proporcionar a cualquier grupo de [ATLAS](#) la misma colección de [LBs](#). Dado que durante los períodos de tomas de datos están disponibles diferentes partes del detector (en un run óptimo, todos los subdetectores están disponibles), hay múltiples [GRLs](#) disponibles para utilizar. Cada análisis, entonces, selecciona qué [GRL](#) utilizar dependiendo de su tolerancia a las fallas de los subdetectores.

La presente tesis utiliza datos recolectados por [ATLAS](#) de colisiones pp durante el Run-2 (2015-2018), a una energía del centro de masa de $\sqrt{s} = 13$ TeV. Durante este run, el [LHC](#) entregó un total de 156 fb^{-1} , de los cuales [ATLAS](#) recolectó 147 fb^{-1} . La luminosidad integrada total disponible para análisis de física es de 140.07 fb^{-1} ², como se ve en la Figure 2.9a. La incertidumbre en la luminosidad integrada combinada para el Run-2 es de 0.83% [29], obtenida usando el detector LUCID-2 [30]. Combinando los años 2022, 2023 y 2024 de toma de datos del Run-3, se reconocieron 159 fb^{-1} de datos, mostrados en la Figure 2.9b [31, 32].

² Las primeras medidas y [GRL](#) iniciales sólo brindaban un total de 139 fb^{-1} disponibles para análisis

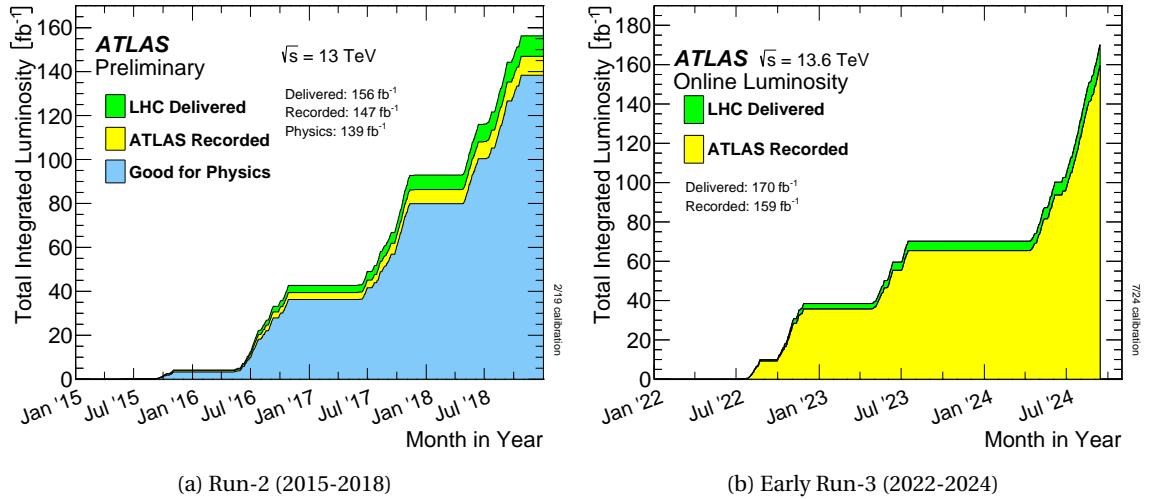


Figura 2.9: Luminosity delivered by the LHC and recorded by ATLAS during the Run-2 [29] and Run-3 data taking periods. For Run-2, the fraction of data good for physics analyses is also displayed.

Otro concepto importante en la adquisición de datos en ATLAS es el *pileup*, que se produce cuando las partículas producidas en más de una colisión pp llegan al detector al mismo tiempo, o más generalmente, cuando las señales se solapan de forma que no pueden separarse. Cuando colisionan haces de protones, la probabilidad de que se produzca una interacción es proporcional a la densidad de partículas, o mejor, al flujo de partículas, que se expresa mediante la luminosidad instantánea. El número real de colisiones de partículas que tienen lugar cuando dos haces se cruzan es una variable aleatoria que sigue una distribución de Poisson. Para luminosidades bajas, en la mayoría de los cruces de haces no se produce ninguna colisión, pero para luminosidades instantáneas altas, en la mayoría de los cruces se producen muchas colisiones de partículas al mismo tiempo. Dependiendo del subdetector y del tipo de medida, puede o no ser posible distinguir entre partículas procedentes de diferentes interacciones simultáneas. Es lo que se denomina como *in-time pileup*. Por el contrario, el *out-of-time pileup* incluye los efectos que surgen cuando el tiempo que el detector necesita para volver a su estado de espera es mayor que el tiempo entre cruces de haces. Una medida cuantitativa del pileup y de la actividad de eventos es el valor medio de interacciones inelásticas pp por bunch-crossing, $\langle\mu\rangle$.

Las luminosidades instantáneas máximas se multiplicaron por cuatro a lo largo de los cuatro años del Run-2, resultando en un aumento de $\langle\mu\rangle$ desde 10 hasta 60, como se muestra en la Figure 2.10a. Para el Run-3, el pileup aumentó drásticamente hasta valores de 57 para el año 2024, aumentando en promedio hasta 52 interacciones por bunch-crossing, mostradas en la Figure 2.10b.

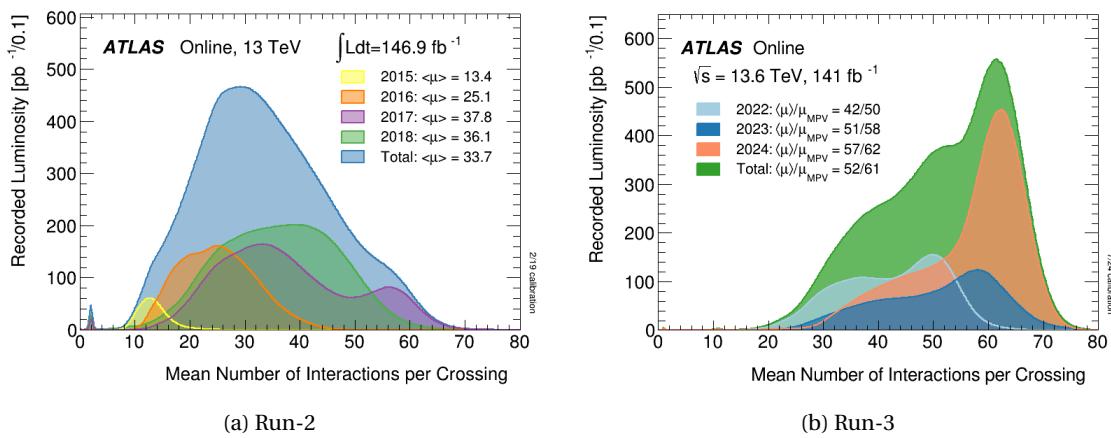


Figura 2.10: Condiciones de pileup durante el Run-2 y Run-3.

RECONSTRUCCIÓN E IDENTIFICACIÓN DE OBJETOS FÍSICOS

3

“Champions keep playing until they get it right.”

Billie Jean King

Las partículas (y los productos de sus decaimientos) producidas en cada colisión, interactúan con el detector de una manera particular según su naturaleza. La información recogida por todos los subdetectores descritos en el capítulo anterior permite reconstruir e identificar los objetos físicos presentes en cada suceso aceptado por el sistema de trigger. Existen dos tipos de reconstrucción e identificación. La *online*, se lleva a cabo al mismo tiempo que se producen las colisiones *pp*, y la *offline*, realizada después de que los eventos se guarden para su almacenamiento. La reconstrucción se realiza evento por evento, y se lleva a cabo del mismo modo para los eventos recolectados por el detector [ATLAS](#) y para los eventos simulados con Monte Carlo ([MC](#)). A continuación, se da un breve resumen de la reconstrucción offline y la identificación de los objetos utilizados en esta tesis.

3.1. Reconstrucción de trazas y vértices

En un evento con alto pileup, puede haber del orden de 1000 partículas cargadas pasando por el detector [ATLAS](#). La información del [ID](#) (Section 2.2.2) se utiliza para reconstruir las trayectorias de las partículas cargadas, denominadas *tracks*.

Dado que el [ID](#) es el detector más cercano al haz y está compuesto por material mínimamente ionizante con una granularidad elevada, este detector desempeña el papel principal en la reconstrucción de trazas. Éstas, permiten encontrar el momento y la trayectoria de las partículas cargadas, dado que dejan una señal en las diferentes capas del [ID](#). Además, como el campo solenoidal dentro del [ID](#) es homogéneo, la trayectoria resultante es circular en el plano *xy*. Cinco parámetros mostrados en la Figure 3.1 definen las trazas de las partículas cargadas:

- q/p_T : la relación entre la carga y el momento transverso que define la curvatura.

- d_0 : la distancia de máxima aproximación al vértice primario en el plano- xy que define el parámetro de impacto transversal.
- z_0 : el parámetro de impacto longitudinal a lo largo del eje z .
- ϕ_0 : el ángulo azimutal.
- θ_0 : el ángulo polar de la dirección de la partícula en el punto más cercano de aproximación [33].

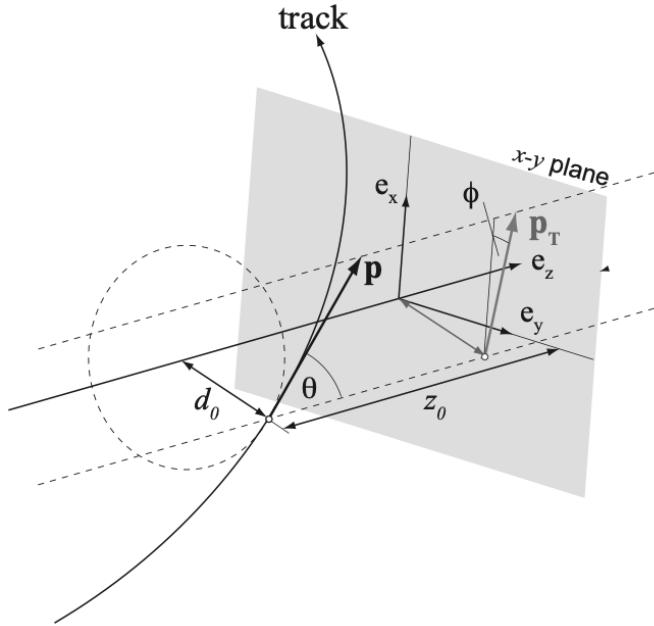


Figura 3.1: Esquema mostrando los parámetros usados para la reconstrucción de trazas [34].

La reconstrucción de trazas utilizada en durante el Run-2 utiliza dos enfoques complementarios: el enfoque *inside-out* y el *outside-in* [35].

El primer paso para la reconstrucción de trazas en el método inside-out es la búsqueda de *seeds*, en la que se buscan tres hits en el detector de silicio para comenzar la reconstrucción de la traza. A partir de estos tres hits y suponiendo un campo magnético uniforme, se obtiene una primera estimación de los parámetros de la traza. A partir de las seeds de la traza, ésta se extrae a las demás capas del detector de silicio, a partir de las cuales se utiliza un filtro combinatorio de Kalman para estimar los parámetros de la traza. En esta fase del proceso puede haber varias trazas candidatas para cada seed. Una vez formada la traza, se aplica un algoritmo de resolución de ambigüedades para reasignar los clusters compartidos a la traza con mejor coincidencia [36], y se ajusta la traza candidata final utilizando un método global χ^2 . La última parte del método inside-out consiste en extender las trazas hasta el **TRT**, e incluir los hits del **TRT**, para mejorar así su resolución del momento.

Para mejorar la eficiencia de las trazas de los decaimientos desplazados del punto de colisión original, se utiliza el método *outside-in*. Se utilizan los hits del **TRT** para comenzar la reconstrucción de la traza y luego se extiende para incluir los hits del detector de silicio, aplicándose

de nuevo un algoritmo para resolver las ambigüedades, mitigando así los hits compartidos entre múltiples trazas.

Los vértices primarios y secundarios son de vital importancia para la posterior reconstrucción de objetos en [ATLAS](#). En este paso, las trazas encontradas como se ha explicado anteriormente se utilizan como *input* para el algoritmo de reconstrucción de vértices [37, 38]. En primer lugar, el Vértice Primario ([PV](#)) se define como el lugar donde se da la interacción entre los dos protones inyectados por el [LHC](#). Los [PVs](#) se reconstruyen emparejando las trazas que se cruzan, lo que se realiza en tres pasos principales: búsqueda de seeds, asignación de trazas y ajustes. Finalmente, el vértice con el mayor $\sum p_T^2$ para todas las trazas asociadas, que también se denomina vértice de dispersión dura, se asigna como el [PV](#). Hay algunas partículas que decaen rápidamente tras su producción, como los leptones τ o los quarks más pesados (b o c -quarks), y su posición de decaimiento puede medirse. A partir de las trazas restantes originadas por estos decaimientos, es posible identificar vértices secundarios. Todos los vértices reconstruidos restantes se consideran pileup.

3.2. Fotones y electrones

La reconstrucción de electrones y fotones en el [ATLAS](#) se basa en la deposición de energía en el [ECAL](#). Como los electrones y los fotones dejan señales similares en este calorímetro, su reconstrucción se realiza simultáneamente, distinguiéndolos por la información de las trazas reconstruidas que se mostró anteriormente.

3.2.1. Reconstrucción

La reconstrucción de fotones y electrones *offline* [39, 40] hace uso de clusters dinámicos de tamaño variable, conectados topológicamente entre las celdas del [ECAL](#) y [HCAL](#) [41], denominados *topo-clusters*, que se agrupan además en *superclusters*. Durante el Run-1 [42–44], en contraste, los clusters eran de tamaño fijo, y que si bien tenían una respuesta lineal energética y estabilidad frente al pileup, no permitía reconstruir eficientemente la energía de fotones bremsstrahlung o de electrones/positrones producto de la creación de pares. La implementación de superclusters durante el Run-2, junto con la calibración de la energía descripta en la Ref. [45] permite solucionar esto sin perder la linealidad y estabilidad de los clusters de tamaño fijo. De esta forma, se distinguen tres tipos de objetos:

- Electrones: consiste en un cluster construido a partir de los depósitos de energía en el [ECAL](#) el cual tiene asignado una traza.
- Fotones convertidos: consiste en un cluster asignado con un vértice (o vértices) de conversión.
- Fotones no convertidos: un cluster que no se encuentra emparejado ni a una traza ni a un vértice de conversión.

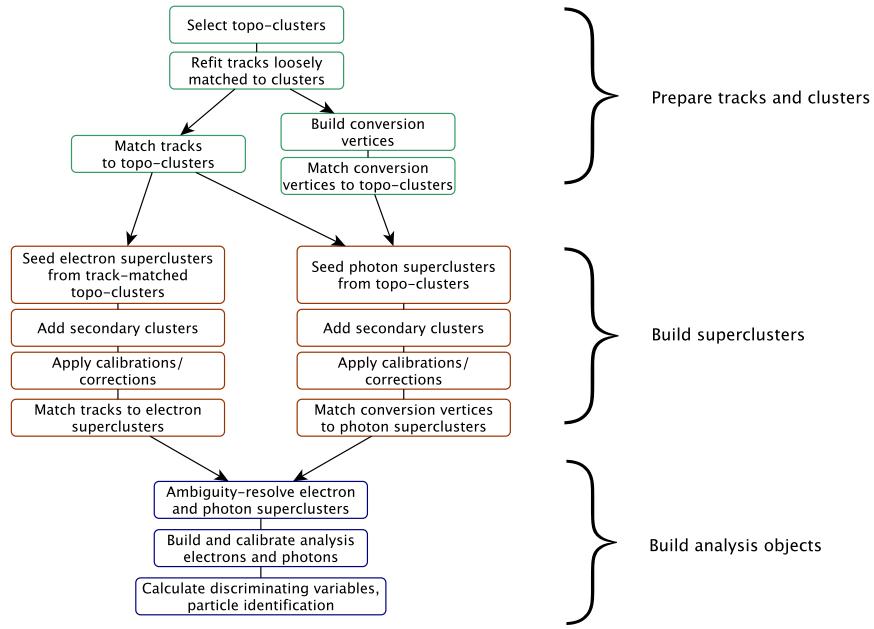


Figura 3.2: Diagrama mostrando el algoritmo de reconstrucción de electrones and fotones, extraído de Ref. [39]

El algoritmo para la reconstrucción de electrones y fotones procede como se muestra en la Figure 3.2. El proceso de reconstrucción comienza con la formación de topo-clusters. Primero, se forman proto-clusters en el ECAL y HCAL agrupando celdas que tienen una requerida y predefinida energía, y añadiendo posteriormente celdas vecinas en cuatro pasos consecutivos, obteniendo así los topo-clusters. Las reconstrucciones continúan sólo en aquellos casos en los que la energía de los topo-clusters en el ECAL es superior a 400 MeV y que la fracción de la misma con respecto a la energía total del topo-cluster sea mayor a 0.5, reduciendo gran parte los efectos de pileup.

El algoritmo también construye vértices de conversión a partir de las trazas reajustadas y los empareja con los topo-clusters seleccionados. Tras el ajuste inicial de las trazas y la construcción de las conversiones, los algoritmos de superclusters de electrones y fotones se ejecutan por separado y en paralelo. En la primera etapa, los topo-clusters se evalúan para su uso como candidatos a clusters semilla, que forman la base de los superclusters; en la segunda etapa, los clusters cercanos a los candidatos a clusters semilla se identifican como candidatos a clusters satélite, que pueden surgir de la radiación bremsstrahlung o de la división de los topo-clusters. Los clusters satélite se añaden a los candidatos a semilla para formar los superclusters finales, si superan unos ciertos criterios de selección necesarios. Tras aplicar correcciones de posición iniciales a los superclusters resultantes, el algoritmo de reconstrucción hace coincidir las trazas con los superclusters de electrones y los vértices de conversión a los superclusters de fotones.

Dado que un objeto puede reconstruirse como electrón y como fotón, se realiza una resolución de ambigüedad para eliminar parte del solapamiento. Sin embargo, se permite cierto solapamiento para mantener una alta eficiencia de reconstrucción de electrones y fotones, y luego se deja a que los análisis pueden aplicar sus propios criterios. Finalmente, se construyen y

calibran los electrones y fotones finales, lo que facilita el cálculo de las variables adicionales utilizadas para los cortes de calidad y la resolución de ambigüedades.

3.2.2. Identificación

Con el objetivo de poder discriminar los objetos *prompt*¹ de aquellos que no lo son, existen diferentes criterios de identificación. En [ATLAS](#), para fotones y electrones, esto se logra mediante una serie de variables denominadas Shower Shapes ([SSs](#)), que con ciertos algoritmos, se logra incrementar la pureza de los objetos deseados, al costo de tener una menor eficiencia de selección. Finalmente, se definen diferentes [WPs](#) que son derivados de forma central y luego distribuidos a toda la colaboración.

El objetivo principal de la identificación de electrones es separar los electrones prompt de los electrones producto del proceso de creación de pares a partir de los fotones, de los jets que depositan energía en el [ECAL](#), y de los electrones provenientes del decaimiento de hadrones de *heavy-flavor*. La identificación se basa en un método likelihood que utiliza algunas de las variables que serán descriptas en el Chapter 4, utilizando electrones provenientes de decaimientos de J/ψ y Z para bajo y alto E_T , respectivamente [46]. Se definen entonces 3 [WPs](#), denominados **Loose**, **Medium** y **Tight**, cuyas eficiencias de identificación de un electrón con $E_T = 40$ GeV son de 93 %, 88 % y 80 %, respectivamente [45].

Para distinguir los fotones reales (los procedentes de la colisión) de los fotones de fondo que tienen secciones transversales de producción mucho mayores (procedentes del decaimiento de hadrones, también llamados fotones falsos), es necesario basarse en un algoritmo de identificación con alta eficiencia de señal y rechazo de fondo, para fotones candidatos con $p_T \sim 10$ GeV hasta la escala TeV. Actualmente, la identificación de fotones en [ATLAS](#) se basa en un conjunto de cortes rectangulares en las [SSs](#) mencionadas anteriormente, que son calculadas a partir de la energía depositada en las celdas del cluster en la primera y segunda capa del [ECAL](#), y de la fuga hacia el [HCAL](#). Estas variables describen el paso de los fotones a través de los calorímetros, caracterizando las lluvias electromagnéticas laterales y longitudinales. El proceso completo de identificación de fotones se presenta en Chapter 4, donde las [SSs](#) se explican una a una. Además, en los Chapters 5 and 6 se presentan dos enfoques para corregir las diferencias observadas en estas variables entre los datos y [MC](#).

3.2.3. Aislamiento

Para mejorar aún más la selección de fotones y electrones se aplican criterios de aislamiento a estos objetos. A su vez, la presencia de otros objetos cerca del fotón o el electrón puede interferir en la correcta reconstrucción de las variables cinemáticas del mismo, como su energía. Para ello, se definen dos criterios de aislamiento: calorimétrico y de trazas.

¹ El término *prompt* hace referencia a aquellos objetos producidos rápidamente luego de la colisión, generalmente provenientes del vértice primario, para distinguirlos de aquellos producidos por el decaimiento tardío de otra partícula, como puede ser un hadrón.

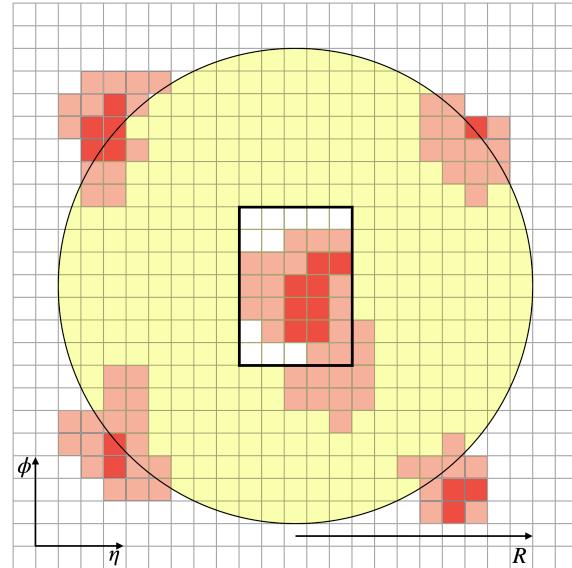


Figura 3.3: Diagrama mostrando el proceso del cálculo de la variable de aislamiento calorimétrico. Cuando se utiliza un cono con $R = 0.4$, se puede construir la variable $E_T^{\text{cone}40}$ mencionada en el texto.

El procedimiento para calcular la energía de aislamiento calorimétrico $E_T^{\text{cone}40}$ es el siguiente, y se muestra en la Figure 3.3. En primer lugar, se construye un cono de radio $\Delta R < 0.4$ alrededor del candidato a fotón o electrón, y se suman las energías de todas las celdas de los topo-clusters (introducidos en la Section 3.2.1) cuyos baricentros se encuentran dentro del cono. A continuación, a esta energía calculada, se le resta la energía de todas las celdas en una ventana de 5×7 (en unidades de $\eta \times \phi$ en la segunda capa del ECAL) centrada alrededor del candidato, con el fin de eliminar la energía del propio candidato. También se tienen en cuenta las contribuciones del pileup y las fugas de energía fuera del cono.

La variable de aislamiento de trazas, $p_T^{\text{cone}20}$, se obtiene sumando los p_T de las trazas de buena calidad en un cono de radio $\Delta R < 0.2$ alrededor del candidato a electrón o en la dirección del cluster de fotones convertidos. Se excluyen de este cómputo las trazas asociadas a la traza o al fotón convertido, así como aquellas trazas que no pasan el requisito de trazas de buena calidad. Una traza de buena calidad se define como aquella en la que el p_T es $p_T > 1 \text{ GeV}$, y tiene una distancia mínima al vértice primario a lo largo del eje z de $|z_0 \sin \theta| < 3 \text{ mm}$.

En general, para los fotones y electrones, no hay otra energía depositada en el cono alrededor del candidato, aparte de los objetos de baja energía originados por los restos de la colisión, las interacciones múltiples y el pileup. En cambio, para los falsos candidatos a fotones y los fotones no directos, se observa energía adicional dentro del cono, originada por los objetos que acompañan al jet.

Tabla 3.1: Resumen de los WP de aislamiento para electrones y fotones usados a lo largo de esta tesis.

Objecto	WP	Aislamiento Calorimétrico	Aislamiento de trazas
Fotón	FixedCutLoose	$E_T^{\text{cone}20} < 0.065 \times p_T$	-
	FixedCutTightCaloOnly	$E_T^{\text{cone}40} < 0.022 \times p_T + 2.45 \text{ GeV}$	-
	FixedCutTight	$E_T^{\text{cone}40} < 0.022 \times p_T + 2.45 \text{ GeV}$	$p_T^{\text{cone}20} / p_T < 0.05$
Electrón	Loose_VarRad	$E_T^{\text{cone}20} < 0.2 \times p_T$	$p_T^{\text{cone}30} / p_T < 0.15$
	HighPtCaloOnly	$E_T^{\text{cone}20} < \max(0.015 \times p_T, 3.5 \text{ GeV})$	-

A partir del aislamiento calorimétrico y de trazas se pueden definir diferentes WPs por separado tanto para electrones como para fotones. En el caso de los electrones, se definen dos estrategias: o bien conseguir una eficiencia fija, o bien aplicar cortes fijos en las variables de aislamiento. En el caso de los fotones, hay WPs que no utilizan ambas variables de aislamiento, como es el caso del WP que sólo utiliza el aislamiento calorimétrico. Las definiciones de los diferentes WPs utilizados a lo largo de esta tesis se muestran en la Table 3.1. Además, es común definir las siguientes variables para WP FixedCutTight del fotón:

$$E_T^{\text{iso}} = E_T^{\text{cone}40} - 0.022 \times E_T - 2.45 \text{ GeV} \quad (3.1)$$

$$p_T^{\text{iso}} = p_T^{\text{cone}20} / E_T \quad (3.2)$$

dejando por tanto las variables definiendo el WP FixedCutTight como:

$$E_T^{\text{iso}} < 0 \text{ GeV} \quad (3.3)$$

$$p_T^{\text{iso}} < 0.05. \quad (3.4)$$

3.3. Muones

La tasa de radiación bremsstrahlung es inversamente proporcional al cuadrado de la masa de una partícula. Dado que los muones son unas 200 veces más pesados que los electrones, interactúan principalmente con el material del detector a través de ionización. Por lo tanto, los muones son partículas mínimamente ionizantes que no crean lluvia electromagnética en los calorímetros y atraviesan todas las capas del detector ATLAS. Es por esta razón que la detección de muones depende de las mediciones de las trazas dejadas por ellos en el ID y el MS. La combinación de los dos subdetectores define cuatro tipos de muones, dependiendo de la información utilizada para la reconstrucción:

- Muones combinados (CB): muón reconstruido a partir de un reajuste global de las trazas del ID y del MS,
- Muones segmentados (ST): muón reconstruido a partir de una traza ajustada del ID que al extrapolarla al MS tienen un segmento en el MDT o el CSC,
- Muones calorimétricos (CT): muón reconstruidos a partir de la traza del ID ajustada a los depósitos de mínima energía ionizante en los calorímetros,
- Muones extrapolados (ME): muón reconstruido únicamente a partir de las trazas MS.

El solapamiento entre distintos tipos de muones se resuelve del siguiente modo. Cuando dos tipos de muones comparten la misma traza del ID, el orden de preferencia es: primero el CB, luego el ST y finalmente el CT. El solapamiento con ME se resuelve analizando los hits de las trazas, seleccionando aquellas trazas con mejor ajuste y mayor número de hits.

Para la identificación de muones, se aplican cortes de calidad para distinguir los muones aislados de los procedentes de procesos de fondo, principalmente del decaimiento de piones y

kaones. Las variables con buen poder discriminatorio utilizadas se describen en Ref. [47]. Se definen cuatro selecciones de identificación: Loose, Medium, Tight y High-pT. Las tres primeras categorías son inclusivas, siendo Medium la selección por defecto en [ATLAS](#). Por último, se pide a los candidatos a muones que van a ser utilizados por los análisis que satisfagan los requisitos de aislamiento, tanto a nivel de trazas como calorimétricos, de forma análoga a los detallados para los electrones y fotones en el apartado anterior. Para el aislamiento de trazas, se utiliza una variable similar a la empleada para los electrones fotones, pero con un cono de radio variable $\Delta R = \min(10 \text{ GeV}/p_T, 0.3)$ alrededor del momento del muón, excluyendo la traza del mismo. Para el aislamiento calorimétrico se utiliza la misma variable $E_T^{\text{cone}40}$, con la diferencia de utilizar un radio de $R = 0.2$, en lugar de 0.4. En base a estas variables, se definen 7 criterios de selección de aislamiento (7 WPs), optimizados para diferentes análisis.

3.4. Jets

Debido al confinamiento de color, un quark o gluón no puede existir por sí mismo y pasa por el proceso de hadronización para formar un chorro colimado de partículas de color neutro, denominados *jets*. Generalmente, los jets penetran a través del [ECAL](#) y son totalmente absorbidos por el material del calorímetro hadrónico. A continuación, se describe brevemente el método típico de agrupación adoptado por [ATLAS](#). También se describen los dos tipos existentes de reconstrucción de jets.

3.4.1. Algoritmo de clusterización de jets anti- k_t

Dado que los jets están constituidos por un elevado número de partículas que dejan deposiciones de energía en el [ECAL](#) y [HCAL](#) y trazas en el [ID](#), un algoritmo de clusterización agrupa los constituyentes en el evento para definir los jets. Dicho algoritmo se denomina algoritmo anti- k_t [48]. Del mismo modo que para los electrones y los fotones, la reconstrucción de los jets [ATLAS](#) se basa en la formación de topo-clusters: depósitos de energía agrupados en las celdas de los calorímetros mediante un algoritmo de combinación secuencial. Entonces, el algoritmo anti- k_t combina el topo-clusters con los siguientes pasos:

- Medir la distancia entre todos los topo-clusters entre sí, y de cada topo-cluster con el haz:

$$d_{ij} = \min(p_{T,i}^{-2}, p_{T,j}^{-2}) \frac{\Delta_{i,j}^2}{R^2} \quad (3.5)$$

$$d_{iB} = p_{T,i}^{-2} \quad (3.6)$$

donde $\Delta_{i,j}^2 = \Delta\phi_{i,j}^2 + \Delta\eta_{i,j}^2$ y R es el radio del jet.

- Si el mínimo de todas las distancias calculadas anteriormente es d_{iB} , el topo-cluster i se clasifica como jet, y se descarta en iteraciones sucesivas.

- Si el mínimo de todas las distancias es d_{ij} , topo-clusters i y j se combinan, todas las distancias se calculan de nuevo con este nuevo topo-cluster y la iteración se realiza de nuevo.

Este proceso se repite hasta que todas las partículas del evento se han agrupado.

El algoritmo anti- k_t comienza agrupando la radiación alrededor de la partícula más dura del evento, ya que la partícula con mayor p_T definirá el término $\min\left(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2}\right)$ en la definición de d_{ij} . Esto permite que los jets del evento tengan una dirección estable al principio del proceso de combinación. El algoritmo anti- k_t es preferible a otros algoritmos secuenciales de jets ya que los jets tienen formas regulares que son aproximadamente cónicas, mostrados en la Figure 3.4. Los jets que se originan a partir de quarks o gluones en general se denominan small- R jets y para su reconstrucción se utiliza un radio de $R = 0.4$. Por otro lado, los jets que representan partículas masivas que decaen hadrónicamente se denominan large- R jets, y se utiliza $R = 1.0$, dado que el uso de un cono más amplio ayuda a incluir la mayoría de las partículas producto del decaimiento.

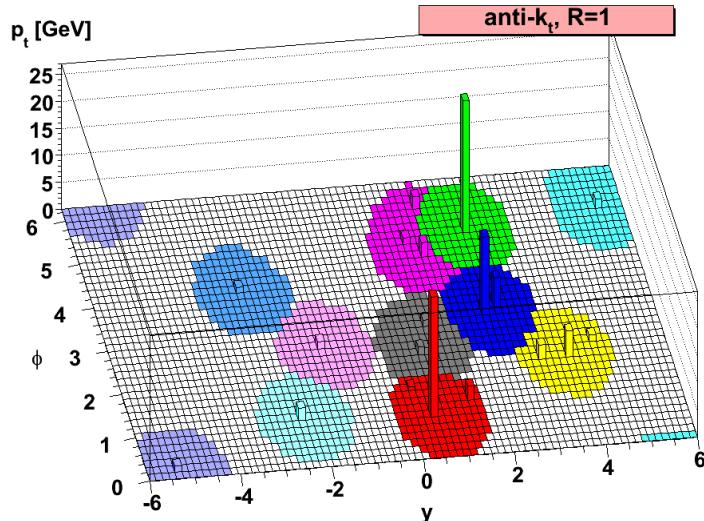


Figura 3.4: Representación esquemática del algoritmo anti- k_t para el proceso de clusterización de jets [48].

3.4.2. Jets Calorimétricos

Una forma de reconstruir los jets se basa en los depósitos de energía en el calorímetro. De forma similar a lo que se ha explicado para electrones y fotones en Section 3.2.1, los depósitos de energía en las celdas del ECAL y HCAL se utilizan para construir topo-clusters, que aproxima los depósitos de energía de hadrones individuales [40, 41]. Los jets reconstruidos de esta manera y agrupados con el algoritmo anti- k_t con un radio de $R = 0.4$ se denominan jets EMTOPO, y son los proxies de los quarks y gluones individuales. En la reconstrucción de jets, sólo se incluyen los topo-clusters con energía neta positiva.

3.4.3. Particle Flow (**PFlow**) Jets

Otro método para la reconstrucción de jets se basa en el algoritmo **PFlow** [49], en el que las mediciones del **ID** y del calorímetro se combinan para formar las señales, que idealmente representan partículas individuales. El algoritmo comienza vinculando cada traza del **ID** con un solo topo-cluster. Luego se calcula la energía esperada en el calorímetro depositada por cada partícula que también inició la traza. Luego, para cada sistema topo-cluster/traza, el algoritmo evalúa la probabilidad de que la energía de la partícula haya sido depositada en más de un topo-cluster, y decide si es necesario agregar más topo-clusters al sistema topo-cluster/traza para recuperar la energía total de la lluvia. Posteriormente, la energía depositada por la partícula que inicia la traza es sustraída celda por celda del conjunto de topo-clusters vinculados. Finalmente, si la energía remanente en el sistema es consistente con la esperada por las fluctuaciones de la lluvia de la señal de una sola partícula, los remanentes del topo-cluster son removidos.

El resultado de este algoritmo es un conjunto de trazas, y un conjunto de topo-clusters modificadas y no modificadas por el procedimiento anterior, que son los objetos **PFlow**. Los objetos **PFlow** también pueden agruparse con el algoritmo anti- k_t y el mismo $R = 0.4$ para formar los jets **PFlow**.

El algoritmo **PFlow** tiene bastantes ventajas sobre el EMTOP:

- La resolución en p_T del **ID** es significativamente mejor que la resolución de energía del calorímetro para partículas cargadas de baja energía.
- Permite una mayor aceptancia para partículas más *soft*. Las trazas se reconstruyen para partículas cargadas con un mínimo p_T de 400 MeV, el cual es menor que el requerido para la formación de topo-clusters.
- Mejora la resolución angular de una sola partícula cargada, ya que utiliza la información del *tracker* (Detector Interno (**ID**)) en lugar de la del calorímetro.
- Las partículas cargadas de bajo p_T que se originan dentro de un jet hadrónico son barridas fuera del cono del jet por el campo magnético para cuando alcanzan el calorímetro. Utilizando la coordenada azimutal de las trazas en el perigeo, estas partículas también son agrupadas en el jet.
- Es posible eliminar las trazas originadas por el pileup, sabiendo que éstas no proceden del **PV**.

Sin embargo, también introduce una complicación. Para cualquier partícula cuya medición de traza deba utilizarse, es necesario identificar correctamente y sustraer su señal en el calorímetro para evitar un doble conteo. En el algoritmo **PFlow**, se toma una decisión booleana sobre si utilizar la medición del tracker o del calorímetro. La capacidad de sustraer con precisión toda la energía de una sola partícula, sin eliminar la energía depositada por otras partículas, constituye el criterio clave de rendimiento sobre el que se optimiza el algoritmo.

En esta tesis, se consideran los **PFlow jets**, ya que han demostrado proporcionar una mejor reconstrucción del jet [49], principalmente para aquellos con bajo p_T y en la reconstrucción E_T^{miss} [50].

3.4.4. Calibración de jets

Una vez reconstruidos los jets, su cuadrimomento se corrige para que coincida con la cinemática de un *truth-jet*², como se muestra en la Figure 3.5. Las tres primeras correcciones tienen en cuenta la contaminación de la distribución de pileup subyacente y las fluctuaciones debidas al origen del jet [51]. La *Global Sequential Calibration* mejora la resolución de p_T de los jets (y las incertidumbres asociadas) eliminando secuencialmente la dependencia de la respuesta reconstruida del jet ($R = E^{\text{reco}}/E^{\text{truth}}$) en diversos observables. Por último, las diferencias residuales entre los datos y **MC** se tienen en cuenta midiendo el desequilibrio de momento en $Z + \text{jets}$, $\gamma + \text{jet}$ y eventos multijet.

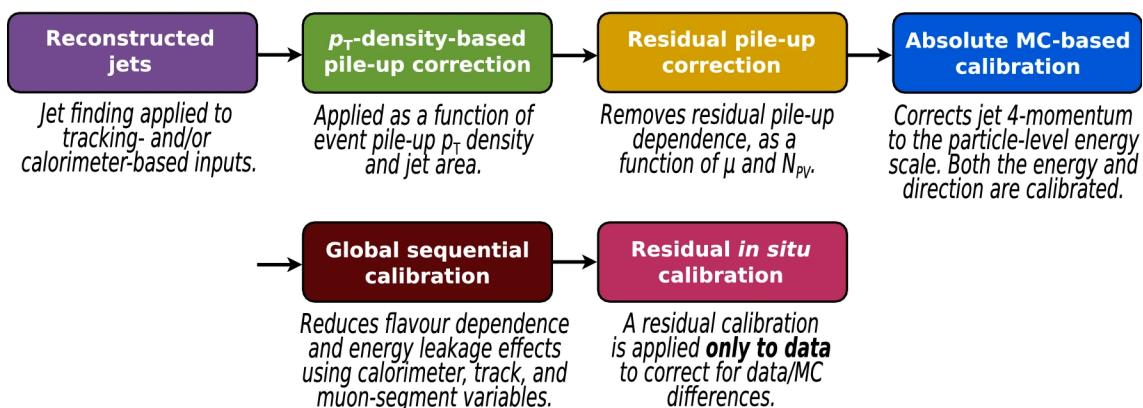


Figura 3.5: Pasos para la calibración del cuadrimomento de PFlow jets [51].

Para reducir el número de jets con una fracción considerable de energía procedente del pileup, se utiliza el algoritmo Jet vertex Tagger (**JVT**). Este algoritmo **update to NNJVT** reconstruye un discriminante multivariante que combina, entre otras cantidades, el Jet vertex Fraction (**JVF**) (fracción de las trazas p_T asociada a un jet originado por el **PV**, y el número total de trazas) y el número de **PVs** en el evento N_{PV} . Como los jets que no proceden de la interacción hard-scatter son generalmente más suaves, el corte **JVT** se aplica sólo a los jets con $p_T < 60 \text{ GeV}$ y $|\eta| < 2.4$. El **JVT WP** por defecto es 96 % eficiente para los jets de dispersión dura.

3.5. Jets de sabor pesado (*heavy flavor*)

Los decaimientos de hadrones pesados (de ahora en más *heavy-flavor*) se rigen principalmente por el hadrón más pesado en la cascada de decaimiento. Un hadrón *b* generalmente decae en cascada a un hadrón *c*, que a su vez decae a un hadrón *s*, etc., lo que conduce a la existencia de múltiples vértices.

² Los truth jets, o jets reales, provienen de partículas del estado final de simulaciones, luego de pasar por el algoritmo de clusterización anti- k_t .

Flavor Tagging ([FTAG](#)) es la clasificación de los jets que contienen hadrones b (b -jets), c (c -jets) o ni b ni c (jets livianos, o l -jets) utilizando algoritmos sensibles a las propiedades distintivas de las respectivas clases. Estos complejos algoritmos se basan en los múltiples vértices, en la elevada masa, la alta multiplicidad de decaimientos y los modos de decaimiento característicos de los hadrones b y c , así como en las propiedades de la fragmentación de los quarks pesados.

En [ATLAS](#) se emplea un proceso de dos etapas para reconstruir las características clave de los heavy-flavor jets. En la primera etapa, los algoritmos de bajo nivel utilizan métodos complementarios para extraer información sobre las trazas de las partículas cargadas vinculadas al jet. Algunos algoritmos se centran en las propiedades de las trazas individuales, mientras que otros analizan sus correlaciones o las combinan para reconstruir explícitamente los vértices desplazados. En la segunda etapa, las salidas de estos algoritmos se integran en un algoritmo de alto nivel que utiliza clasificadores multivariantes para optimizar el rendimiento. Con el tiempo, los algoritmos han evolucionado significativamente, empezando con discriminantes basados en likelihoods y *Boosted Decision Trees* ([BDTs](#)) durante el Run-1 del [LHC](#), y avanzando hacia métodos más avanzados como las redes neuronales recurrentes y profundas, lo que ha dado lugar a notables mejoras en el rendimiento de la identificación [52–55].

A partir del Run-3, el grupo de [ATLAS FTAG](#), desarrolla un novedoso algoritmo "GN2" basado en un *Transformer*. El algoritmo GN2 es un único modelo entrenado que sustituye a DL1d [56] y a los algoritmos de bajo nivel que lo alimentan. Se basa en GN1 [57], que se refinó rápidamente para pasar a ser GN2. GN2 sustituye la *Graph Attention Network* [58] utilizada por GN1 por un Transformador [59], y también se beneficia de otras optimizaciones de arquitectura y de un orden de magnitud más de estadística para su entrenamiento.

GN2 acepta directamente información sobre el jet y las trazas asociadas y, como tal, no depende de otros algoritmos de etiquetado de sabores (*flavor tagging*). GN2 mantiene los dos objetivos de entrenamiento auxiliares que se introdujeron con GN1: la agrupación de trazas que se originan en un vértice común y la predicción del proceso físico subyacente del que se originó cada traza.

Este nuevo algoritmo también está preparado para proporcionar la identificación de c -jets y jets procedentes de decaimientos τ . Las salidas de este tagger corresponden a las probabilidades de que un jet sea taggeado como un jet b , c , τ o *light*, denominadas como p_b , p_c , p_τ y p_w respectivamente.

3.5.1. Identificación y performance de b -tagging

Para evaluar la capacidad del tagger de identificar b -jets con una eficiencia constante, se mide la capacidad de rechazar los jets c , τ y *light*. Las probabilidades de salida del tagger se combinan para construir un único discriminante GN2_b , definido como

$$\text{GN2}_b = \log \left(\frac{p_b}{f_c p_c + f_\tau p_\tau + (1 - f_c - f_\tau) p_u} \right). \quad (3.7)$$

Los parámetros $f_{c(\tau)}$ son libres y determinan la importancia entre $p_{c(\tau)}$ y p_u en el discriminante. Los valores específicos de estos parámetros se determinan mediante un procedimiento de optimización basado en maximizar el rechazo de c -jets (τ -jets) y l -jets, y resultan ser 0.2 (0.01).

A partir de la valor discriminante del tagger, se pueden definir varios WPs, simplemente exigiendo que el valor $GN2_b$ esté por encima de un determinado umbral. El grupo FTAG de ATLAS proporciona de forma centralizada a toda la colaboración 5 WPs diferentes para lograr una eficiencia global fija de b -tagging: 65, 70, 77, 85 y 90%, y se muestran en la Figure 3.6. En dicha figura se comparan también las distribuciones de datos y MC del tagger GN2, donde las contribuciones de los distintos sabores se muestran con colores diferentes.

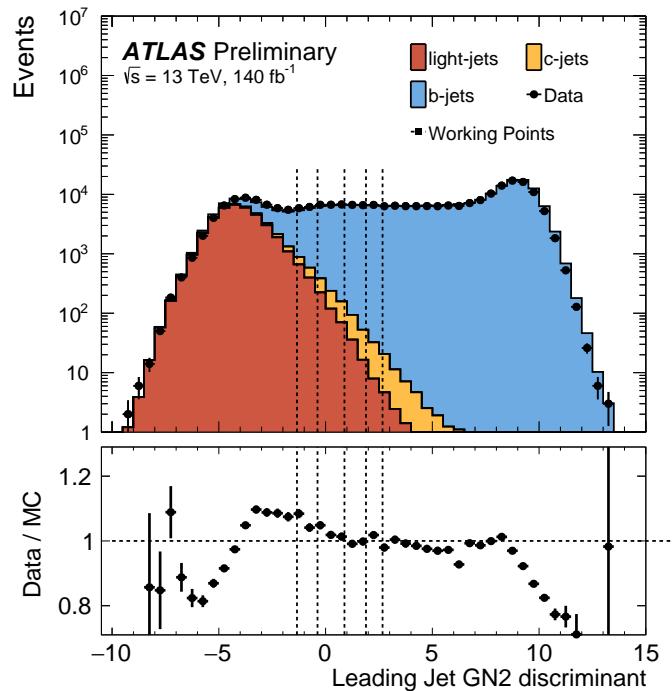


Figura 3.6: Comparación entre datos y simulación MC (eventos de $t\bar{t}$ de un sólo lepton) del discriminante del tagger GN2. Las contribuciones de los jets l , b y c se muestran con diferentes colores, y los 5 WPs de b -tagging se muestran con las líneas verticales punteadas. De izquierda a derecha, las líneas representan los WPs de 90, 85, 77, 70 y 65% de eficiencia. El panel inferior muestra el ratio entre los datos y toda la simulación MC [60].

Uno de los principales problemas del b -tagging es la disminución de la eficiencia a mayor p_T . En este régimen de p_T elevado, las partículas son más colimadas y tienden a viajar más lejos en el ID antes de decaer, lo que puede dar lugar a una traza de decaimiento con hits espurios. La degradación de la eficiencia se visualiza en la Table 3.2, donde se muestran las eficiencias de tagging para b -jets, junto con los rechazos c -jets, l -jets y τ -jets, en los regímenes de bajo y alto p_T . Los valores mostrados se calculan utilizando diferentes muestras, en las que $t\bar{t}$ se utiliza en la región bajo de p_T y eventos de decaimiento de Z' ³ se utilizan en la región alto de p_T . Puede verse que la eficiencia de b -tag cae en un 30% para jets de p_T más alto.

³ El modelo leptofóbico de vector axial Z' es un modelo de Materia Oscura simplificado en el cual el decaimiento teorizado es un par de quarks.

Tabla 3.2: Medidas de eficiencias de b -tagging, y de rechazos de c -jets, l -jets y τ -jets, en los regímenes de bajo y alto p_T .

Muestra	Rango de p_T [GeV]	Eficiencia de b -jet	Rechazo de c -jet	Rechazo de l -jet	Rechazo de τ -jet
$t\bar{t}$	$20 < p_T < 250$	0.76	17.52	448.61	71.15
Z'	$250 < p_T < 6000$	0.41	20.27	179.99	452.94

3.5.2. Identificación y performance de c -tagging

Al igual que con b -tagging, se puede construir un único discriminante a partir de las probabilidades dadas por el tagger para identificar c -jets frente a b -jets, τ -jets y l -jets:

$$\text{GN2}_c = \log\left(\frac{p_c}{f_b p_b + f_\tau p_\tau + (1 - f_b - f_\tau) p_u}\right) \quad (3.8)$$

donde ahora los valores $f_{b(\tau)}$ son los parámetros libres que controlan el rechazo entre jets b , τ y light. Utilizando el mismo procedimiento de optimización que para b -tagging, los valores para $f_{b(\tau)}$ resultan ser 0.3 (0.05).

Gracias a la gran eficiencia b -tagging conseguida por GN2, es posible diseñar un WP de c -tagging tras aplicar un veto de b -tagging, separando aún más los c -jets de los l -jets. Construyendo este WP de tagging simultáneo y asumiendo que la fracción de τ -jets es despreciable, se puede separar los jets b , c y livianos en tres regiones ortogonales. Partiendo de exigir que un jet *no* pase el WP de b -tagging de 77% de eficiencia (veto b -tag), se definen tres WPs diferentes de c -tagging definidos para eficiencias de 10, 30 y 50%, fijando el valor de GN2_c . Las medidas de eficiencia y rechazo de las dos muestras descritas anteriormente, tras aplicar el WP de c -tag de 50% de eficiencia se muestran en el Table 3.3.

Tabla 3.3: Medidas de eficiencia de c -tagging efficiencies para c -jets, y valores de rechazos de b -jets, l -jets y τ -jets en los regímenes de bajo y alto p_T . Los valores corresponden a aquellos luego de aplicar el veto del WP de b -tagging de 77% y de 50% de c -tagging. **rejection values not correct!**

Muestra	Rango de p_T [GeV]	Eficiencia de c -jet	Rechazo de b -jet	Rechazo de l -jet	Rechazo de τ -jet
$t\bar{t}$	$20 < p_T < 250$	0.467	17.52	448.61	71.15
Z'	$250 < p_T < 6000$	0.344	20.27	179.99	452.94

Parte III

Photon shower shape corrections

PHOTON IDENTIFICATION AND SHOWER SHAPES

4

“Champions keep playing until they get it right.”

Billie Jean King

The [ECAL](#) was described in Section 2.2.3, where the measurement mechanism was described. In this subdetector, photons deposit their energy via bremsstrahlung radiation and electron-positron pair creation, therefore creating an electromagnetic ([EM](#)) shower. The [ECAL](#) does a great job to compute the energy of the [EM](#) shower, but identifying the initiating particle remains a challenging task. However, by virtue of the different layers and granularities in the [ECAL](#), different characteristics of these [EM](#) showers can be studied, and are encoded by different variables called Shower Shape Variables ([SSVs](#)s).

4.1. Shower shapes

As mentioned in Section 3.2.2, photon identification relies on rectangular cuts applied to [SSVs](#) that can achieve excellent separation power between real isolated photons from fake photons originating from hadrons. These [SSVs](#) are computed from the photon candidates' energy deposits in the [ECAL](#) and [HCAL](#) cells, and serve to describe the passage of the photons candidates throughout the calorimeters, characterizing the lateral and longitudinal [EM](#) showers.

In general, real photons produce narrower energy deposits in the [ECAL](#), and have lower leakages to the [HCAL](#), compared to those photons provenient from hadrons, where the presence of additional neighbouring hadrons close to the fake photon tend to widen the showers. Furthermore, since the first layer of the [ECAL](#) consists on fine strips, it is possible to discriminate photon candidates coming from $\pi^0 \rightarrow \gamma\gamma$ decays, characterized by two local maxima due to the presence of two nearby photons.

In the following, the [SSVs](#) used for photon identification are detailed. The first variable makes use of the energy measured in the [HCAL](#):

- Hadronic leakage: is the transverse energy deposited in the [HCAL](#), normalized to the

energy deposited in the **ECAL**:

$$R_{\text{had}(1)} = \frac{E_{\text{T}}^{\text{had}}}{E_{\text{T}}^{\text{EM}}} \quad (4.1)$$

In order to minimize the effects of resolution degradation, in the barrel-endcap transition region of the **HCAL** ($0.8 \leq |\eta| \leq 1.37$) the energy deposit in the whole **HCAL** is used (R_{had}). On the remaining of the detector, only the energy deposited in first layer of the **HCAL** is used ($R_{\text{had } 1}$).

The following variables use the second-layer information of the **ECAL**:

- Lateral energy profile in η :

$$R_\eta = \frac{E_{3 \times 7}^{s2}}{E_{7 \times 7}^{s2}} \quad (4.2)$$

where $E_{i \times j}^{s2}$ is the energy sum in the second calorimeter layer contained in a window of $i \times j$ cells (units of $\eta \times \phi$ cells), centered at the most energetic cell. This variable gives a measure of the showers' width in the η direction.

- Lateral energy profile in ϕ :

$$R_\phi = \frac{E_{3 \times 3}^{s2}}{E_{3 \times 7}^{s2}} \quad (4.3)$$

defined in a similar way as R_η . However, this variable behaves very different for converted and unconverted photons. Due to the action of the magnetic field, the electrons and positrons are curved into opposite directions in ϕ , having as a result, **EM** showers much wider in the case of converted photons than those for unconverted ones.

- Lateral shower width in η :

$$w_{\eta^2} = \sqrt{\frac{\sum E_i \eta_i^2}{\sum E_i} - \left(\frac{\sum E_i \eta_i}{\sum E_i} \right)^2} \quad (4.4)$$

measures the proper width of the **EM** shower, where E_i is the energy in the i -th cell of the **ECAL**, measured in a window of 3×5 cells in $\eta \times \phi$.

The following variables use the information from the first **ECAL** layer, composed of the strip cells that allow for a high η resolution and allows for a good separation between isolated photons from photons product of the π^0 decay. The following FIGURE shows the difference in the energy deposited in the **ECAL** between the two cases mentioned previously.

- Lateral energy profile in η

$$f_{\text{side}} = \frac{E_7^{s1} - E_3^{s1}}{E_3^{s1}} \quad (4.5)$$

measures the energy outside the core of the three central strips within a window of 7 cells, divided by the energy in the three central cells.

- Lateral shower width in η (3 strips)

$$w_1 = \sqrt{\frac{\sum E_i(i - i_{max})^2}{\sum E_i}} \quad (4.6)$$

where i runs over all cells in a window of 3 cells around the highest-energy-cell. This variable measures the width of the **EM** shower in the first layer of the calorimeter.

- Lateral shower width in η (full). It is defined in a similar way as w_1 , but uses all the cells in a window of $\delta\eta \times \delta\phi = 0.0625 \times 0.2$, corresponding to approximately to 20×2 strips $\eta \times \phi$.
- Energy difference

$$\Delta E = E_{\max,2}^{s1} - E_{\min}^{s1} \quad (4.7)$$

represents the energy difference between the second maximum and the minimum reconstructed energy between the two maxima in the strip layer.

- Energy ratio

$$E_{\text{ratio}} = \frac{E_{\max,1}^{s1} - E_{\max,2}^{s1}}{E_{\max,1}^{s1} + E_{\max,2}^{s1}} \quad (4.8)$$

is the ratio of energy difference between the two maxima, normalized to the sum of those energies, in the strip layer.

4.2. Photon Identification

4.2.1. Optimisation

Starting from these discriminating **SSVs**, three **WPs** can be defined: *loose*, *medium* and *tight* **WPs** [46]. The loose **WP** is employs cuts to the variables defined in the second layer and to the hadronic leakage variable, used primarily by the trigger. The medium **WP** is a **WP** optimised to have a flat 95 % efficiency. This **WP** applies cut to all the previously defined variables (strip and middle layer and leaks to the **HCAL**). Finally, the tight **WP**, uses all the **SSVs** defined and provides an excellent background rejection. TABLE shows which variables are used for each **WP**.

ADD PLOTS OF ALL THE VARIABLES COMPARING REAL AND FAKE

The cuts on the **SSVs** for each identification **WP** are optimised as a function of the transverse energy and the pseudo-rapidity of the photon candidate, to account for the shape of the variables for different η and for variations in the amount of material and the geometry of the calorimeter. The three **WPs** are also optimised separately for converted and unconverted photons. The optimisation is performed with a Multivariate (**MV**) approach where signal efficiencies are scanned between 0 % and 100 % while trying to maximise the background rejection. The resulting, optimised, cut values are subject to fluctuations and therefore they are manually smoothed.

Two different **MC** samples are used for the optimisation procedure, representative at different p_T^γ . For photons with $10 < p_T < 25$ GeV, radiative Z decays ($Z \rightarrow \ell^+ \ell^- \gamma$) samples are used as signal, while $Z +$ jets events accounts for the background. Events used are selected by requiring two opposite charged leptons and a minimum angular separation between the photon and the lepton of $\Delta R_{\min}(\ell, \gamma) > 0.4$. To reject non-radiating Z bosons, the dilepton invariant mass has to satisfy $m_{\ell\ell} < 83$ GeV, and the three-body invariant mass $m_{\gamma\ell\ell}$ needs to approximate the Z boson mass: 80 GeV $< m_{\gamma\ell\ell} < 100$ GeV. Finally, the photon is required to have $|\eta| < 2.37$, excluding the crack region. Finally, for higher p_T photons, $p_T > 25$ GeV, the inclusive-photon ($\gamma+$ jet) signal events are compared against dijet backgrounds. The event selection used in this case is simply requiring the photon to be in the **ECAL** acceptance region (excluding the crack).

4.2.2. Efficiency measurements

Photon identification efficiency measurements are carried out using three different methods that are detailed in Ref. [?] and that are combined to yield correction factors for analyses. In all cases, photons are required to satisfy the Loose isolation criterion defined in Ref. [?] and therefore the photon efficiencies are measured relative to this isolation criterion. In the following paragraphs, a brief description of each method is given.

For the lower p_T range ($7 < p_T < 100$ GeV), photons from radiative Z decays are used as signal photons, selecting the events in the same way as for the **WPs** optimisation (Section 4.2.1). The only difference in this case, is an additional lower limit on the di-lepton invariant mass of $40 < m_{\ell^+ \ell^-} < 83$ GeV. To estimate the number of signal and background events, template fits to the observed three-body invariant-mass distribution are performed.

The second method to compute efficiencies relies on Smirnov transformations [61] to the electrons' **SSs** to resemble those of photons'. The samples used in this approach are $Z \rightarrow ee$ decays, in which the electrons are required to pass loose photon isolation. The candidate electrons in data contain a small background from $W +$ jets and multijet production; this background is subtracted by fitting simulated signal samples and background templates derived from data control regions to the m_{ee} data distributions. The electron candidates are counted for events in the range $70 < m_{ee} < 110$ GeV, and the efficiencies are measured using the tag-and-probe method described in Ref. [?]. The p_T range in which this method is implemented is $25 < p_T < 250$ GeV.

The final and third method uses higher p_T photons originating from Quantum Chromodynamics (**QCD**) $\gamma+$ jet production with transverse momenta in the range $50 < p_T < 1500$ GeV. The photons for this study are required to pass the loose identification **WP** employed in the trigger. This sample is dominated by background dijet events whose production cross section is orders of magnitude higher. The maxtrix method [?] is used in this case, which constructs four orthogonal regions that either pass or fail the tight identification **WP**, and pass or fail the track-isolation (described in Section 3.2.3). For each region, two unknowns arise: the number of signal and background events. If the track isolation efficiencies are known for the signal and background components, then it is possible to estimate the efficiency for loose photons pas-

sing the tight identification criteria. The isolation efficiencies for signal photons are estimated using **MC** samples, and the ones for backgrounds are obtained in a jet-enriched control region constructed by inverting the identification criteria. The efficiency measurements in data for the tight identification **WP** then reads:

$$\epsilon^{\text{tight-ID}} = \frac{\frac{\hat{\epsilon}_{\text{ID}} - \hat{\epsilon}_{\text{ID}}^b}{\hat{\epsilon}_{\text{ID}}^s - \hat{\epsilon}_{\text{ID}}^b} \cdot N_{\text{ID}}^T}{\frac{\hat{\epsilon} - \hat{\epsilon}^b}{\hat{\epsilon}^s - \hat{\epsilon}^b} \cdot N^T}, \quad (4.9)$$

where N^T accounts for the totality of photons in the inclusive sample which consists on N^s prompt photons (or signal photons) and N^b fake photons (background photons). The number N_{ID}^T is the subset of N^T that pass the identification requirement. Data, signal and background track isolation efficiencies are represented by $\hat{\epsilon}$, $\hat{\epsilon}^s$ and $\hat{\epsilon}^b$, respectively. Similarly, the track isolation efficiencies for those photons passing tight identification are shown as $\hat{\epsilon}_{\text{ID}}$, $\hat{\epsilon}_{\text{ID}}^s$ and $\hat{\epsilon}_{\text{ID}}^b$, respectively. The measured efficiencies for photons with $p_{\text{T}} > 150$ GeV is between 90 and 96%.

Since data and simulation measured efficiencies do not match, **MC** needs to be corrected to account for these differences. In the ideal case where one expects perfect agreement between both samples, ratios of the data efficiencies to simulation efficiencies in each $p_{\text{T}} - \eta$ -conversion status bin should be 1.0. These ratios are referred as Scale Factors (**SFs**) and are computed separately for each one of the methods described. Then, the different methods' **SFs** are combined using a weighted average in each bin, assuming the statistical and systematic uncertainties to be uncorrelated between the methods. Resulting **SFs** in all cases are consistent with 1.0, only deviating by a maximum of 2%. The only exception to this case is in the first p_{T} -bin ($7 < p_{\text{T}} < 10$ GeV) where deviations of up to 30% take place.

4.3. Shower shapes variables differences between data and MC

The **ATLAS MC** simulation does not perfectly describes data. This is clearly seen when computing the previously mentioned **SFs**, whose values were different from 1, meaning that different efficiencies are obtained in data and in **MC**. In particular, when comparing the **SSs** distributions, it is seen that **MC** distributions are shifted or even the whole shape differs.

The main differences on the distributions arise for the η shower profiles, where broader distributions were seen in data compared to **MC**. Part of the effect was corrected in 2010 after moving to detailed description of the material composition in the accordion absorbers in GEANT4. However, the remaining data-**MC** disagreements are still under study and could be due to several potential effects:

- Detector geometry description of the lead thickness (including possible variations of due to gravity) or material composition, material before the **ECAL**, a decrease of the width of cells caused by calorimeter contraction due to temperature (mainly in the first layer).
- Mismodeling of the electric field in the **LAr** gaps.

- Mismodeling of the cross-talk effect (energy sharing between calorimeter cells due to electronics possible in η direction).

To account for the differences in the **SSs**, historically, corrections were made in the form of shifts to each one of the **MC** distributions. These shifts comprised the so-called Fudge Factors (**FFs**), and were determined using a χ^2 minimisation on the comparison of data and **MC SSs** [?,?]. Even though the differences decreased substantially after these corrections, some of them remained, shown in FIGURE. It is seen from the distributions that the main differences that remained are related to the shape of the distributions, therefore needing for higher order corrections. In Chapter 5 a detailed description of newly derived corrections is presented. Since **SSs** are built from energy deposits on the **ECAL** cells, another possible way of correcting the current disagreement between data and **MC SSs** is to directly correct the energies on **MC** at a cell-level, fixing the differences in all **SSVs** at once. This new approach is studied in Chapter 6.

4.4. Samples and event selection for the SS correction studies

As mentioned above, the improved **FF** method and a novel cell-based reweighting method is presented in Chapters 5 and 6.

Similar to what had been done for the identification optimisation studies, two photon samples are used for the **FFs** calculation. For photons with $7 \leq p_T^\gamma \leq 50$ GeV, Final State Radiation (**FSR**) photons from Z -boson decay are considered, while photons from **QCD** γ +jet events are used for photons with $p_T^\gamma \geq 50$ GeV, hereinafter referred as Radiative Z (**RZ**) photons and Single Photon (**SP**) samples, respectively. On the other hand, for the cell-level corrections to the **SSVs**, only **RZ** photons are used. In what follows, event selection for both types of samples is detailed.

For both types of corrections, the **MC** samples are reweighted to match the luminosity of the collected **ATLAS** data, and also pileup re-weighted to match the pileup profile shown in Section ??.

4.4.1. Radiative Z boson decays

For low- p_T photons, **RZ** photons are used as signal photons, while backgrounds are modeled by $Z \rightarrow \ell\ell$ events. The photons are required to pass the following selection:

- **ECAL** $|\eta|$ acceptance region. First of all, the photons are required to be inside the **ECAL** acceptance region excluding the crack, detailed in Section 2.2.3, given by $|\eta| < 1.37$ or $1.52 < |\eta| < 2.37$.
- **Isolation**. Fake photon candidates are removed by imposing an isolation requirement on the calorimetric isolation variable with the `FixedCutTightCaloOnly` **WP**.
- **FSR selection**. As shown in FIGURE, the vast majority of events correspond to $m_{\gamma\ell\ell} > 100$ GeV and $m_{\ell\ell} \sim m_Z \approx 91$ GeV, which represent Initial State Radiation (**ISR**) photons

(photons radiated from the initial quarks). Photon candidates from **ISR** are largely affected by the $Z + \text{jets}$ background, where a jet fakes a photon¹. However, a second peak appears in the distribution where the three-body invariant mass approximates the Z mass ($m_{\gamma\ell\ell} \approx m_Z$). These particular type of events are referred as **FSR** photons, characterised by high real photon purity, and are the ones of interest for the correction studies.

- **Photon-lepton overlap.** By requiring a minimum angular distance between the photon and the closest lepton ($\Delta R_{\min} > 0.4$), biases in the **ECAL** deposits by the objects are avoided.
- **Truth-matching.** [give description here](#)

Special selection for cell-based reweighting corrections

For the cell-energy reweighting method to correct the **SS**s, a special selection needs to be applied to the **EM** clusters. For the current studies, only **SS**s built from the second layer are studied. Clusters of 7×11 cells in $\eta \times \phi$ are considered, shown in FIGURE with the current cell arrangement used.

In this work, only "healthy clusters." are considered, that is, events need to be associated to clusters of 77 cells (no cells missing) and the central cell must be the one with highest energy in the corresponding cluster ("hottest cell"). In FIGURE, an example of the averaged energies at each cell, for events with unconverted photons in data is shown.

For these particular studies, the photon isolation requirement is relaxed to the **FixedCutLoose WP** is used.

4.4.2. Inclusive photons

Photon+jet events are used for the high- p_T regime of the **FFs** corrections. The events pass loose identification trigger requirements and for the nominal values of the corrections tight identification is applied. This selection is applied to reduce the vast di-jet background, which has much higher production cross-sections. As for the **RZ** photon samples, the photons are required to be within the **ECAL** acceptance region excluding the crack.

¹ The production cross-section of $Z + \text{jets}$ is about three orders of magnitude higher than that of $Z + \gamma$, and a non-negligible fraction of jets contains high- $p_T \pi^0$'s, decaying to collimated photon pairs

FUDGE FACTORS

5

“Champions keep playing until they get it right.”

Billie Jean King

The **FF** method has been applied to photons in **ATLAS** as simple shifts of the distributions. In this chapter, a detailed explanation of the calculation is shown, with the addition of improvements derived for this type of corrections.

5.1. Calculation

FFs are computed in a series of steps that starts by preparing the input datasets up to the systematic uncertainties calculation. In the first step, histograms of all the **SSVs** are created and then smoothed, which are then used to perform the actual optimisation of the **FFs**. This process is repeated for different isolation and identification selection requirements, in order to compute the systematic uncertainties. In the following a step by step description of the process is described.

The calculation is performed separately for the two considered samples: **RZ** for photons with $7 \leq p_T \leq 50$ GeV and **SP** for photons with $p_T > 50$ GeV. Since **SSs** distributions vary as a function of p_T and $|\eta|$, the computation is done in bins of these mentioned variables:

$$p_T^\gamma : \begin{cases} \textbf{RZ} : [7, 15, 20, 30, 50] \text{ GeV} \\ \textbf{SP} : (50, 60, 80, 100, 150, 300, 600, \infty] \text{ GeV} \end{cases} \quad (5.1)$$

$$|\eta| : [0, 0.6, 0.8, 1.15, 1.37, 1.52, 1.81, 2.01, 2.37]. \quad (5.2)$$

Furthermore, as mentioned in Section 4.1, there are variables very sensitive to the conversion status of the photons, that is, whether if the photons are converted or unconverted. For this reason, the calculation is done separately for converted and unconverted photons. A total of nine variables are corrected using this method: E_{ratio} , f_{side} , R_η , R_ϕ , R_{had} , R_{had1} , w_1 , $w_{\eta 2}$ and $w_{\text{s tot}}$; as they are the ones in which the largest discrepancies are seen between data and **MC**.

For each mentioned [SSV](#), histograms of [MC](#) and data of 100 bins are created. The choice of the binning is done based on having sufficient statistics at each bin and also to capture all the features of the variables.

After that, each histogram is smoothed using the Kernel Density Estimator ([KDE](#)) tool from TMVA [62]. The [KDE](#) method consists of estimating the shape of a Probability Density Function ([pdf](#)) by the sum over smeared events. The [pdf](#) $p(x)$ of a variable x is

$$p(x) = \frac{1}{N} \sum_{i=1}^N K_h(x - x_i) \quad (5.3)$$

where N is the number of events, $K_h(t) = K(t/h)/h$ is the kernel function, and h is the bandwidth of the kernel. The basic idea is that each event is considered as a Dirac- δ -function, which are replaced by a Kernel function (Gaussian) and finally they are summed altogether to form the final Parton Distribution Function ([PDF](#)). The [KDE](#) smoothing can be applied in two forms: non-adaptive [KDE](#) or adaptive [KDE](#), as seen in [FIGURE](#). In the former, the bandwidth is constant for the entire sample h_{NA} , while in the latter, it uses the value from non-adaptive [KDE](#), but it varies as a function of $p(x)$ as

$$h_A = \frac{h_{NA}}{\sqrt{p(x)}} \quad (5.4)$$

Adaptive [KDE](#) improves the shape of the estimated [pdf](#) in regions of low statistics, however, in high statistics regions it can give rise to ".over-smoothing". The degree of smoothing is tuned by multiplying the bandwidth h by fine factors. These fine factors are user-defined parameter which are tuned to allow the [pdf](#) to retain the important features of the [SS](#) but to also avoid statistical fluctuations. Higher values indicate broader Kernel functions and therefore de [pdf](#) catches less statistical fluctuations. Examples of the smoothing procedure applied to R_η are shown in [FIGURE](#) for cases in which original histograms have low and high statistics.

Once the data and [MC pdfs](#) are created for a given variable, p_T , $|\eta|$ and conversion type, the [MC pdf](#) is normalised to data's and a χ^2 value is computed between both, excluding the underflow and overflow bins, as:

$$\chi^2 = \sum_{i=1}^N \frac{(w_{MC,i} W_{data} - w_{data,i} W_{MC})^2}{s_{MC,i}^2 W_{data}^2 + s_{data,i}^2 W_{MC}^2}. \quad (5.5)$$

N is the number of bins in the [pdfs](#), $w_{MC,i}$ and $w_{data,i}$ are the event numbers of [MC](#) and data at each bin, respectively, $s_{MC,i}$ and $s_{data,i}$ are the bin errors and finally W_{data} and W_{MC} are the sum of weights for data and [MC](#), respectively.

5.1.1. Shift-only corrections

Taking into account only the mean's correction of the [SSs](#), the [MC pdf](#) is shifted to the left and right one bin at a time. As a consequence of this procedure, the shift [FF](#) resolution directly depends on the bin-width of the [pdfs](#), and since the [pdfs](#) have been generated with high accuracy given the relatively small bin-width of the histograms, the [pdfs](#) are built to have a total of 5000 bins. The starting number of bins that the [MC](#) distribution needs to be shifted is estimated by

computing the difference on the means between data and simulation. From this starting value, shifts of 100 bins to each side are considered.

For each bin the distribution has been shifted, the aforementioned χ^2 value is computed and recorded. Assuming that the measurements errors $s_{MC,i}$ and $s_{data,i}$ have a normal gaussian distribution¹, and that the parameters for each χ^2 value are independent, it is expected that the shape followed by the χ^2 values is approximately paraboloidal, which can be seen from FIGURE.

To extract the FFs, the χ^2 scan near the minimum is fitted with a parabolic function (5 bins to each side of the minimum bin) and the shift FF is obtained from the fit minimum (see FIGURE).

Finally, the SSs can be corrected as

$$SS_{\text{new}} = SS_{\text{old}} + \text{shift},$$

shown for f_{side} in FIGURE.

5.1.2. Shift+stretch corrections

In order to improve the agreement between data and MC corrections to the widths of the distributions are introduced. To apply both corrections, first the maximum of the MC pdf is found and then the pdf is stretched around it. The stretch value is varied between 0.3 and 2.5 with a total of 300 steps. Since stretches are applied to the MC pdf, there might be cases in which the stretch is big enough to give rise to empty bins inbetween. To treat these empty bins, they are interpolated using the closest non-zero bins to each side. After the stretch has been applied, the MC pdf is shifted to the left and right, and then the same procedure as before is followed. As a result, a two-dimensional grid (shift and stretch plane) of χ^2 values is obtained, and the FFs are obtained from the center of the minimum bin. Finally, corrections are applied to MC SSs as

$$SS_{\text{new}} = \text{stretch} \times (SS_{\text{old}} - \text{stretch point}) + \text{shift} + \text{stretch point}. \quad (5.6)$$

This procedure is shown in FIGURE for the same shower shape shown in REFERENCE.

5.2. Uncertainties

5.2.1. Statistical uncertainties

To extract the statistical uncertainties on the shift and stretch FFs, the 1σ contour is needed. In the large sample limit and for two parameters x and y , the χ^2 values distribution near the minimum take the form

$$\chi^2 = \chi^2_{\min} + \frac{1}{1-\rho^2} \left[\left(\frac{x-x_0}{\sigma_x} \right)^2 + \left(\frac{y-y_0}{\sigma_y} \right)^2 - 2\rho \left(\frac{x-x_0}{\sigma_x} \right) \left(\frac{y-y_0}{\sigma_y} \right) \right], \quad (5.7)$$

¹ This requirement is satisfied as long as the bin contents of both pdfs are greater than 10.

where ρ is the correlation coefficient between both variables, and χ^2_{\min} is the χ^2 minimum value obtained from the 2D histogram.

The contour which gives the 1σ (68.3 % confidence level) uncertainty on the parameters is given by $\chi^2_{\min} + 2.3$. To obtain the points that make any desired contour, bin content and bin centers are used to interpolate the points position.

Using a quadric to fit the ellipse and transforming it to the most general canonical form

$$\frac{((x - x_0) \cos \theta + (y - y_0) \sin \theta)^2}{a^2} + \frac{((x - x_0) \sin \theta - (y - y_0) \cos \theta)^2}{b^2} = 1, \quad (5.8)$$

it is possible to extract the ellipse center (x_0, y_0) , its tilt angle θ and its semi-major and semi-minor axis, a and b , respectively. Using this information, the x and y upper and lower limits are obtained as follows:

$$x_{\text{limits}} = x_0 \pm \sqrt{a^2 \cos^2 \theta + b^2 \sin^2 \theta} = x_0 \pm \sigma_x \quad (5.9)$$

$$y_{\text{limits}} = y_0 \pm \sqrt{a^2 \sin^2 \theta + b^2 \cos^2 \theta} = y_0 \pm \sigma_y \quad (5.10)$$

giving then the statistical uncertainties on x and y as

$$\sigma_x = \sqrt{a^2 \cos^2 \theta + b^2 \sin^2 \theta} \quad (5.11)$$

$$\sigma_y = \sqrt{a^2 \sin^2 \theta + b^2 \cos^2 \theta}. \quad (5.12)$$

FIGURE shows the ellipse parameters and limits shown previously, and FIGURE shown an example of application of the fit to the $w_{\text{s tot}}$ SS. Finally, it is also possible to get the correlation coefficient ρ as

$$\rho = \tan(2\theta) \frac{\sigma_x^2 - \sigma_y^2}{2\sigma_x \sigma_y}. \quad (5.13)$$

5.2.2. Systematic uncertainties

The systematic uncertainties are derived by varying the preselection criteria, that is, photon identification and photon isolation. Changing different preselection criteria allows the SSVs to vary depending on the amount of background contamination, and therefore also the FFs will change. The different selections (also named configurations) are:

- Radiative Z (RZ) sample:

- Nominal configuration: No ID, FixedCutTightCaloOnly isolation.
- Loose ID, no isolation.
- Loose ID, FixedCutTightCaloOnly isolation.
- No ID, FixedCutLoose isolation.

- Single Photon (SP) sample:

- Nominal configuration: Tight ID, FixedCutLoose isolation.
- Tight ID, FixedCutTight isolation.

All other combinations (or lack thereof) of selection criteria would result in either a sample with too low statistics, or too low purity.

FFs are derived for each one of the previous configurations, and the difference between the ones in the nominal configuration with the remaining ones is calculated. The maximum difference is taken as the systematic uncertainty, as the most conservative approach.

5.3. Results

CELL-BASED ENERGY REWEIGHTING

6

“Champions keep playing until they get it right.”

Billie Jean King

yet another template (yat)

Parte IV

New Physics

ANALYSIS MOTIVATION AND STRATEGY

7

*"Champions keep playing until they
get it right."*

Billie Jean King

yet another template (yat)

SIGNAL AND BACKGROUND SAMPLES

8

“Champions keep playing until they get it right.”

Billie Jean King

yet another template (yat)

EVENT SELECTION

9

“Champions keep playing until they get it right.”

Billie Jean King

yet another template (yat)

BACKGROUND ESTIMATION**10**

“Champions keep playing until they get it right.”

Billie Jean King

yet another template (yat)

BACKGROUND MODELING**11**

“Champions keep playing until they get it right.”

Billie Jean King

yet another template (yat)

SYSTEMATIC UNCERTAINTIES

12

*“Champions keep playing until they
get it right.”*

Billie Jean King

yet another template (yat)

STATISTICAL ANALYSIS

13

“Champions keep playing until they get it right.”

Billie Jean King

yet another template (yat)

RESULTS

14

“Champions keep playing until they get it right.”

Billie Jean King

yet another template (yat)

CONCLUSIONS

Don't let anyone rob you of your imagination, your creativity, or your curiosity.

Mae Jemison

GLOSSARY

CERN Organización Europea para la Investigación Nuclear	6
ALICE A Large Ion Collider Experiment	
ATLAS A Toroidal LHC ApparatuS	ix
CMS Compact Muon Solenoid	
LHC Gran Colisionador de Hadrones	ix
LHCb Large Hadron Collider beauty	
LHCf Large Hadron Collider forward	
LS1 Long Shut down 1	11
LS2 Long Shut down 2	ix
TOTEM TOTal cross section, Elastic scattering and diffraction dissociation Measurement at the LHC	
MoEDAL Monopole & Exotics Detector At the LHC	
PS Proton Synchrotron	6
SPS Super Proton Synchrotron	6
L1 Level-1	vi
L1Topo Level-1 Topológico	17
LAr Argón Líquido	13
ECAL Calorímetro Electromagnético	vi
CSC Cathode Strip Chamber	16
CTP Central Trigger Processor	17
FCAL Calorímetro Forward	13
HEC Calorímetro Hadrónico del End-Cap	15
HCAL Calorímetro Hadronic	vi
HLT High Level Trigger	vi

Glossary	60
IBL <i>Insertable B-Layer</i>	10
ID Detector Interno	ix
SCT Semiconductor de Trazas	10
RPC <i>Resistive-Plate Chamber</i>	16
TGC <i>Thin-Gap Chamber</i>	16
TRT Detector de Radiación de Transición	10
NSW <i>New Small Wheel</i>	16
MDT <i>Monitored Drift Tube</i>	15
MS Espectrómetro de Muones	vi
GRL <i>Good Runs List</i>	18
LB Bloques de Luminosidad	18
PSB Proton Synchrotron Booster	6
BDT <i>Boosted Decision Tree</i>	32
MC Monte Carlo	x
MV Multivariate	38
KDE Kernel Density Estimator	44
ROI Región de interés	16
FF Fudge Factor	41
SF Scale Factor	40
SS Shower Shape	vii
SSV Shower Shape Variable	36
WP Working Point	x
pdf Probability Density Function	44
EM electromagnetic	36
FSR Final State Radiation	41
RZ Radiative Z	41
SP Single Photon	41
ISR Initial State Radiation	41
ME Matrix Element	27
PDF Parton Distribution Function	44
QCD Quantum Chromodynamics	39
SM Standard Model	
JVT Jet vertex Tagger	31

JVF	Jet vertex Fraction	31
PV	Vértice Primario	23
CB	Muones combinados	27
ST	Muones segmentados	27
CT	Muones calorimétricos	27
ME	Muones extrapolados	27
PFlow	Particle Flow	vii
FTAG	<i>Flavor Tagging</i>	31

BIBLIOGRAFÍA

- [1] O. S. Brüning, J. Poole, P. Collier, P. Lebrun, R. Ostojic, S. Myers, and P. Proudlock, *LHC Design Report*, CERN, Geneva **1** (2004) 548. <https://cds.cern.ch/record/782076>. 6
- [2] L. Evans and P. Bryant, *LHC Machine*, *Journal of Instrumentation* **3** no. 08, (2008) S08001. <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08001><https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08001/meta>. 6
- [3] E. Lopienska, *The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022*, <https://cds.cern.ch/record/2800984>. ix, 7
- [4] The ALICE Collaboration, *The ALICE experiment at the CERN LHC*, *Journal of Instrumentation* **3** no. 08, (2008) S08002. <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08002>. 6
- [5] The LHCb Collaboration, *The LHCb Detector at the LHC*, *Journal of Instrumentation* **3** no. 08, (2008) S08005. <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08005><https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08005/meta>. 6
- [6] The CMS Collaboration, *The CMS experiment at the CERN LHC*, *Journal of Instrumentation* **3** no. 08, (2008) S08004. <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08004><https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08004/meta>. 6
- [7] The ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *Journal of Instrumentation* **3** no. 08, (2008) S08003. <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08003>. ix, xi, 6, 9, 14, 15
- [8] The LHCf Collaboration, *The LHCf detector at the CERN Large Hadron Collider*, *Journal of Instrumentation* **3** no. 08, (2008) S08006. <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08006><https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08006/meta>. 6

- [9] The TOTEM Collaboration, *The TOTEM Experiment at the CERN Large Hadron Collider*, *Journal of Instrumentation* **3** no. 08, (2008) S08007.
<https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08007>
<https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08007/meta>. 6
- [10] MoEDAL Collaboration, *Technical Design Report of the MoEDAL Experiment.*, 6
- [11] R. M. Bianchi and A. Collaboration, “ATLAS experiment schematic or layout illustration.” 2022. ix, 8, 9
- [12] J. Pequenao, “Computer generated image of the ATLAS inner detector.” 2008. 10
- [13] J. Pequenao, “Computer generated image of the ATLAS inner detector.” 2008. 10
- [14] The ATLAS Collaboration, *Inner Detector Forward SCT Barrel SCT TRT Pixel Detectors Technical Design Report*, tech. rep., 1997.
<https://cds.cern.ch/record/331063/files/ATLAS-TDR-4-Volume-I.pdf>. 10
- [15] T. A. Collaboration, *ATLAS Insertable B-Layer Technical Design Report Addendum*, Tech. Rep. May, Sep, 2012. <https://cds.cern.ch/record/1291633>
<https://cdsweb.cern.ch/record/1451888>. 11
- [16] K. Potamianos, *The upgraded Pixel detector and the commissioning of the Inner Detector tracking of the ATLAS experiment for Run-2 at the Large Hadron Collider*, [arXiv:1608.07850](https://arxiv.org/abs/1608.07850). <https://arxiv.org/abs/1608.07850v1>. 11
- [17] The ATLAS Collaboration, *ATLAS pixel detector electronics and sensors*, *Journal of Instrumentation* **3** no. 07, (2008) P07007.
<https://iopscience.iop.org/article/10.1088/1748-0221/3/07/P07007>
<https://iopscience.iop.org/article/10.1088/1748-0221/3/07/P07007/meta>. 11
- [18] T. Heim, *Status and performance of the ATLAS Pixel Detector after 3 years of operation*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **765** (2014) 227–231. 11
- [19] The ATLAS Collaboration, *The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: design and performance*, *Journal of Instrumentation* **3** no. 02, (2008) P02013.
<https://iopscience.iop.org/article/10.1088/1748-0221/3/02/P02013>
<https://iopscience.iop.org/article/10.1088/1748-0221/3/02/P02013/meta>. 11
- [20] J. Pequenao, “Computer Generated image of the ATLAS calorimeter.” 2008. ix, 12
- [21] V. Rossetti, *Performance of the ATLAS Calorimeters and Commissioning for LHC Run-2*, <https://cds.cern.ch/record/2037117>. 13
- [22] *ATLAS muon spectrometer : Technical Design Report.*,
<https://cds.cern.ch/record/331068>. 15

- [23] B. Stelzer, *The New Small Wheel Upgrade Project of the ATLAS Experiment*, Nuclear and Particle Physics Proceedings **273-275** (2016) 1160–1165.
<https://www.sciencedirect.com/science/article/pii/S2405601415006719>. 16
- [24] The ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2010*, The European Physical Journal C **72** no. 1, (2012) 1849. 16
- [25] A. ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2015*, arXiv:1611.09661. <http://arxiv.org/abs/1611.09661> http://dx.doi.org/10.1140/epjc/s10052-017-4852-3. 16
- [26] The ATLAS Collaboration, *Operation of the ATLAS trigger system in Run 2*, Journal of Instrumentation **15** no. 10, (2020), arXiv:2007.12539v2. 16
- [27] R. Achenbach, P. Adragna, V. Andrei, P. Apostologlou, B. Åsman, C. Ay, B. M. Barnett, B. Bauss, et al., *The ATLAS Level-1 Calorimeter Trigger*, Journal of Instrumentation **3** no. 03, (2008) P03001. <https://dx.doi.org/10.1088/1748-0221/3/03/P03001>. 16
- [28] ATLAS Collaboration, P. Jenni, M. Nessi, M. Nordberg, and K. Smith, *ATLAS high-level trigger, data-acquisition and controls: Technical Design Report*. Technical design report. ATLAS. CERN, Geneva, 2003. <https://cds.cern.ch/record/616089>. 17
- [29] The ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, The European Physical Journal C **83** no. 10, (2023) 982. ix, 18, 19
- [30] G. Avoni, M. Bruschi, G. Cabras, D. Caforio, N. Dehghanian, A. Floderus, B. Giacobbe, F. Giannuzzi, et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, Journal of Instrumentation **13** no. 07, (2018) P07017.
<https://dx.doi.org/10.1088/1748-0221/13/07/P07017>. 18
- [31] ATLAS Collaboration, *Preliminary analysis of the luminosity calibration of the ATLAS 13.6 TeV data recorded in 2022*, tech. rep., CERN, Geneva, 2023.
<https://cds.cern.ch/record/2853525>. 18
- [32] ATLAS Collaboration, *Preliminary analysis of the luminosity calibration for the ATLAS 13.6 TeV data recorded in 2023*, tech. rep., CERN, Geneva, 2024.
<https://cds.cern.ch/record/2900949>. 18
- [33] The ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, The European Physical Journal C **77** no. 10, (2017) 673. 22
- [34] T. G. Cornelissen, N. Van Eldik, M. Elsing, W. Liebig, E. Moyse, N. Piacquadio, K. Prokofiev, A. Salzburger, and A. Wildauer, *Updates of the ATLAS Tracking Event Data Model (Release 13)*, tech. rep., CERN, Geneva, 2007. <https://cds.cern.ch/record/1038095>. ix, 22

- [35] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig, E. Moyse, and A. Salzburger, *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, tech. rep., CERN, Geneva, 2007. <https://cds.cern.ch/record/1020106>. 22
- [36] T. A. collaboration, *A neural network clustering algorithm for the ATLAS silicon pixel detector*, *Journal of Instrumentation* **9** no. 09, (2014) P09009. <https://dx.doi.org/10.1088/1748-0221/9/09/P09009>. 22
- [37] The ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, *The European Physical Journal C* **77** no. 5, (2017) 332. 23
- [38] *Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13$ TeV*, tech. rep., CERN, Geneva, 2015. <https://cds.cern.ch/record/2037717>. 23
- [39] The ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data*, *Journal of Instrumentation* **14** no. 12, (2019) P12006. <https://dx.doi.org/10.1088/1748-0221/14/12/P12006>. ix, 23, 24
- [40] ATLAS Collaboration, *Electron and photon reconstruction and performance in ATLAS using a dynamical, topological cell clustering-based approach*, tech. rep., CERN, Geneva, 2017. <https://cds.cern.ch/record/2298955>. 23, 29
- [41] The ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, *The European Physical Journal C* **77** no. 7, (2017) 490. 23, 29
- [42] ATLAS Collaboration, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run-1 data*, *Eur. Phys. J. C* **76** (2016) 666, [arXiv:1606.01813 \[hep-ex\]](https://arxiv.org/abs/1606.01813). 23
- [43] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, *Eur. Phys. J. C* **74** (2014) 3071, [arXiv:1407.5063 \[hep-ex\]](https://arxiv.org/abs/1407.5063). 23
- [44] W. Lampl, S. Laplace, D. Lelas, P. Loch, H. Ma, S. Menke, S. Rajagopalan, D. Rousseau, S. Snyder, and G. Unal, *Calorimeter Clustering Algorithms: Description and Performance*, tech. rep., CERN, Geneva, 2008. <https://cds.cern.ch/record/1099735>. 23
- [45] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton–proton collision data*, *JINST* **14** (2019) P03017, [arXiv:1812.03848 \[hep-ex\]](https://arxiv.org/abs/1812.03848). 23, 25
- [46] The ATLAS Collaboration, *Electron and photon efficiencies in LHC Run 2 with the ATLAS experiment*, *Journal of High Energy Physics* **2024** no. 5, (2024) 162. 25, 38
- [47] The ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}=13$ TeV*, *The European Physical Journal C* **76** no. 5, (2016) 292. 28

- [48] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- kt jet clustering algorithm*, *Journal of High Energy Physics* **2008** no. 04, (2008) 063.
<https://dx.doi.org/10.1088/1126-6708/2008/04/063>. ix, 28, 29
- [49] The ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, *The European Physical Journal C* **77** no. 7, (2017) 466. 30, 31
- [50] ATLAS Collaboration, *E_T^{miss} performance in the ATLAS detector using 2015-2016 LHC p-p collisions*, tech. rep., CERN, Geneva, 2018. <https://cds.cern.ch/record/2625233>. 31
- [51] The ATLAS Collaboration, *Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *The European Physical Journal C* **81** no. 8, (2021) 689. ix, 31
- [52] *Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data*, tech. rep., CERN, Geneva, 2014. <https://cds.cern.ch/record/1741020>. 32
- [53] ATLAS Collaboration, *Measurement of the b-tag Efficiency in a Sample of Jets Containing Muons with 5 fb^{-1} of Data from the ATLAS Detector*, tech. rep., CERN, Geneva, 2012.
<https://cds.cern.ch/record/1435197>. 32
- [54] A. Lavrenov, *MV2-algorithm's clones*, CoRR **cs.CR/0208003** (2002).
<https://arxiv.org/abs/cs/0208003>. 32
- [55] M. Lanfermann, *Deep Learning in Flavour Tagging at the ATLAS experiment*, PoS **EPS-HEP2017** (2018) 764. 32
- [56] The ATLAS Collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset*, *The European Physical Journal C* **83** no. 7, (2023) 681. 32
- [57] ATLAS Collaboration, *Graph Neural Network Jet Flavour Tagging with the ATLAS Detector*, tech. rep., CERN, Geneva, 2022. <https://cds.cern.ch/record/2811135>. 32
- [58] S. Brody, U. Alon, and E. Yahav, *How Attentive are Graph Attention Networks?*, CoRR **abs/2105.14491** (2021), [2105.14491](https://arxiv.org/abs/2105.14491). <https://arxiv.org/abs/2105.14491>. 32
- [59] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, *Attention Is All You Need*, CoRR **abs/1706.03762** (2017), [1706.03762](https://arxiv.org/abs/1706.03762). [http://arxiv.org/abs/1706.03762](https://arxiv.org/abs/1706.03762). 32
- [60] Z. Wu, “b-jets calibration in atlas.” ATLAS-CMS Flavour Tagging Workshop, 2024. x, 33
- [61] L. Devroye, *Non-Uniform Random Variate Generation*. Springer New York, NY, 1986. 39
- [62] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, et al., *TMVA - Toolkit for Multivariate Data Analysis*, 2007. [arXiv:physics/0703039 \[physics.data-an\]](https://arxiv.org/abs/physics/0703039). 44