



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

13 de octubre 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,*  
*13 de octubre 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

<b>Índice general</b>	<b>vi</b>
<b>Índice de figuras</b>	<b>ix</b>
<b>Índice de tablas</b>	<b>xiii</b>
<b>Introduction</b>	<b>2</b>
<b>I Theory Motivation</b>	<b>3</b>
<b>1 The Standard Model and Beyond</b>	<b>4</b>
<b>II Experimental setup</b>	<b>5</b>
<b>2 El LHC y el experimento ATLAS</b>	<b>6</b>
2.1. LHC .....	6
2.2. ATLAS .....	8
2.2.1. Sistema de coordenadas de ATLAS .....	9
2.2.2. Detector Interno .....	9
2.2.3. Calorímetros .....	11
2.2.4. Espectrómetro de Muones (MS) .....	15
2.2.5. El sistema de Trigger .....	16
2.3. Toma de datos durante el Run-2 .....	17
<b>3 Reconstrucción e identificación de objetos físicos</b>	<b>20</b>
3.1. Reconstrucción de trazas y vértices .....	20
3.2. Fotones y electrones .....	22
3.2.1. Reconstrucción .....	22
3.2.2. Identificación .....	24
3.2.3. Aislamiento .....	24
3.3. Muones .....	26

---

3.4. Jets . . . . .	27
3.4.1. Algoritmo de clusterización de jets anti- $k_t$ . . . . .	27
3.4.2. Jets Calorimétricos . . . . .	28
3.4.3. Particle Flow (PFlow) Jets . . . . .	28
3.4.4. Calibración de jets . . . . .	30
3.5. Jets de sabor pesado ( <i>heavy flavor</i> ) . . . . .	30
3.5.1. Identificación y performance de <i>b</i> -tagging . . . . .	31
3.5.2. Identificación y performance de <i>c</i> -tagging . . . . .	32
<b>III Photon shower shape corrections</b>	<b>34</b>
<b>4 Shower shapes y la Identificación de fotones</b>	<b>35</b>
4.1. Shower Shapes (SSs) . . . . .	35
4.2. Identificación de fotones . . . . .	39
4.2.1. Procesos de interés y selección de eventos . . . . .	39
4.2.2. Optimización . . . . .	41
4.2.3. Estimación de las eficiencias . . . . .	41
4.3. Las diferencias de las Shower Shapes entre datos y Monte Carlo (MC) . . . . .	44
<b>5 Correcciones de las Shower Shapes (SSs)</b>	<b>46</b>
5.1. Fudge Factors (FFs) . . . . .	46
5.1.1. Muestras de datos y simulaciones MC . . . . .	46
5.1.2. Cálculo de Fudge Factors (FFs) . . . . .	47
5.1.3. Cálculo de incertezas . . . . .	50
5.1.4. Resultados . . . . .	52
5.2. Correcciones de energía de las celdas . . . . .	54
5.2.1. Selección de eventos . . . . .	55
5.2.2. Cálculo de las correcciones . . . . .	56
5.2.3. Resultados . . . . .	59
5.3. Conclusiones y trabajo futuro . . . . .	60
5.3.1. Trabajo a futuro . . . . .	61
<b>IV New Physics</b>	<b>63</b>
<b>6 Analysis motivation and strategy</b>	<b>64</b>
<b>7 Signal and Background samples</b>	<b>65</b>
<b>8 Event selection</b>	<b>66</b>
<b>9 Background estimation</b>	<b>67</b>
<b>10 Background modeling</b>	<b>68</b>

<b>11 Systematic uncertainties</b>	<b>69</b>
<b>12 Statistical analysis</b>	<b>70</b>
<b>13 Results</b>	<b>71</b>
<b>Conclusions</b>	<b>72</b>
<b>A Relaciones para el ajuste de una elipse</b>	<b>73</b>
<b>Glossary</b>	<b>74</b>
<b>Bibliografía</b>	<b>77</b>

# ÍNDICE DE FIGURAS

2.1.	Vista general del complejo de aceleradores del Gran Colisionador de Hadrones (LHC) [3]. . . . .	7
2.2.	Vista general del detector A Toroidal LHC ApparatuS (ATLAS) y de todos sus subdetectores, 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]. . . . .	13
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.	Luminosidad entregada por el LHC y recolectada por ATLAS durante el Run-2 [30] y el Run-3. En el caso de Run-2, también se muestra la fracción de datos recolectados que son útiles para análisis de física. . . . .	18
2.10.	Distribución del número de interacciones por bunch-crossing durante Run-2 (izquierda) y Run-3 (derecha). . . . .	19
3.1.	Esquema de los parámetros usados para la reconstrucción de trazas [35]. . . . .	21
3.2.	Diagrama del algoritmo de reconstrucción de electrones and fotones, extraído de Ref. [40] . . . . .	23
3.3.	Diagrama del 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. . . . .	25
3.4.	Representación esquemática del algoritmo anti- $k_t$ para el proceso de clusterización de jets [50]. . . . .	28

3.5. Comparación entre datos y simulación MC (eventos del proceso $t\bar{t}$ semileptónico) del discriminante del tagger GN2. Las contribuciones de los jets $l$ , $b$ y $c$ se muestran con diferentes colores, y los 5 <i>Working Points</i> (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 la suma de las simulaciones MC [61]. . . . .	32
4.1. Representación esquemática de las SS de fotones. Los valores $E_C^{S_N}$ representan la energía en la capa $N$ del ECAL en un cluster $C$ . . . . .	36
4.2. Depósitos de energía característicos para un fotón aislado (izquierda), y un evento $\pi^0 \rightarrow \gamma\gamma$ (derecha), que es posible distinguir gracias a la fina granularidad de la primera capa del ECAL [62]. . . . .	38
4.3. Diagramas de Feynman del proceso de radiación de un fotón en decaimientos $Z \rightarrow \ell\ell\gamma$ para los casos de Radiación de estado inicial (ISR) (izquierda) y Radiación de estado final (FSR) (derecha). . . . .	40
4.4. Masa invariante de los dos leptones en función de la masa invariante de ambos leptones juntos con un fotón en (a) datos, (b) fondos y (c) señal. La región en la cual se encuentra una gran concentración de eventos con $m_{\ell\ell} \sim m_Z$ corresponde a eventos de ISR, mientras que eventos de FSR están caracterizados por $m_{\ell\ell\gamma} \sim m_Z$ . . . . .	40
4.5. Distribuciones normalizadas de señal (azul) y fondo (naranja) de diferentes SSs utilizando las muestras de $Z$ Radiativo (RZ) y pasando la selección de eventos detallada en la Sección 4.2.1. . . . .	42
4.6. Comparación de las eficiencias calculadas para datos y MC utilizando los tres métodos diferentes para su cálculo. En ambas figuras, para cada método, se muestran dos conjuntos diferentes de mediciones MC: la nominal y la corregida (discutida en el texto). Los paneles inferiores muestran el cociente entre las eficiencias de los datos y las predicciones MC (denominadas Factor de escalas (SFs) en el texto). Las figuras fueron tomadas de la Ref. [64]. . . . .	43
4.7. SFs resultantes de la identificación de fotones en los diferentes bins de $p_T$ y $ \eta $ para fotones convertidos (izquierda) and no convertidos (derecha). <b>Ask fran for the plots in his presentation!</b> . . . . .	44
4.8. Comparaciones de las SSs entre los datos (puntos negros) y la simulación MC nominal (línea roja) y corregida (línea azul), utilizando la muestra RZ [47]. . . . .	44
5.1. Esquema del suavizado no adaptativo y adaptativo del método Kernel Density Estimator (KDE). . . . .	48
5.2. Suavizado utilizando el método KDE aplicado a la SS $R_{had}$ para fotones en $0.8 <  \eta  < 1.15$ bajo dos posibles escenarios: baja y alta estadística. El histograma original se muestra con los puntos azules y las correspondientes Función de Densidad de Probabilidades (PDFs) con la línea naranja. Además, se muestran los valores de los fine factors usados en cada caso. . . . .	48

5.3. Cálculo de los FFs de shift+stretch para $f_{\text{side}}$ utilizando fotones no convertidos con momento transverso de $15 < p_T < 20 \text{ GeV}$ y pseudorapidez $2.01 <  \eta  < 2.37$ . . . . .	50
5.4. Valores de los FFs de shift y stretch para las $R_\eta$ (izquierda) y $w_{\eta^2}$ (derecha) para fotones convertidos con $1.15 <  \eta  < 1.37$ , en función de $p_T$ . Los resultados obtenidos por las muestras de RZ están representados por el color negro, mientras que los resultados de Single Photon (SP) se muestran en azul. Los puntos y las líneas denotan los valores centrales con sus incertezas estadísticas, mientras que las regiones sombreadas representan las incertezas totales. Los valores de shift se muestran en el panel superior, los cuales son normalizados por el ancho de la distribución luego de ser estirada por el stretch, como se ha explicado en el texto. Este último valor se muestra en el panel inferior de las figuras. . . . .	53
5.5. Valores de los FFs de shift y stretch para $w_{s\text{tot}}$ en función de $ \eta $ utilizando fotones convertidos con $50 < p_T < 60 \text{ GeV}$ de las muestras de SP. La Figura a muestra los valores de shift normalizados, mientras que los no normalizados se encuentra en la Figura b. Los puntos con las líneas de color muestran los valores centrales y las incertezas estadísticas, mientras que las áreas sombreadas representan las incertezas totales en cada bin. Los valores de stretch se muestran en los paneles inferiores de cada figura. . . . .	53
5.6. Distribuciones de algunas SSs seleccionadas usando las muestras de RZ para fotones convertidos luego de aplicar las correcciones de los FFs en la simulación. Las distribuciones de las SS están separadas para fotones en la región del barrel (fila de arriba) y en la región del endcap (fila de abajo). Los puntos negros representan los datos recolectados por ATLAS, mientras que las simulaciones no corregidas y corregidas están mostradas por las líneas azules y verdes, respectivamente. Los paneles inferiores, en cada figura, muestra el cociente entre el histograma de datos con cada uno de los obtenidos de las simulaciones MC. . . . .	54
5.7. Ídem a la Figura 5.6 pero utilizando las muestras de SP. . . . .	55
5.8. Disposición de las celdas y distribución de la energía entre las celdas del cluster. . .	56
5.9. Correcciones a las energías de las celdas de la simulación MC utilizando el mismo método diseñado para electrones. <b>fix luminosity</b> . . . . .	58
5.10. Ejemplo de las matrices de corrección de shift (izquierda) y stretch (derecha). Los valores mostrados corresponden al cálculo de las correcciones utilizando fotones no convertidos. Los valores de shift son multiplicados por un factor de 100 para mejorar su visualización. . . . .	59
5.11. Distribuciones de las energías normalizadas de las celdas 28, 39 y 50 de cluster de 77 celdas, para fotones no convertidos. Los puntos azules y rojos corresponden a las distribuciones de la simulación MC con y sin las correcciones, respectivamente, mientras que el histograma gris representa los datos. . . . .	60

5.12. Distribucioens de las SSs calculadas en la segunda capa del ECAL para fotones no convertidos (fila superior) y convertidos (fila inferior) con pseudorapidez $ \eta  < 0.6$ , comparando los diferentes métodos de corrección con los datos. Los datos experimentales están representados por los histogramas grises. La simulación MC sin corregir se muestra con los puntos rojos, la simulación corregida por el método de corrección de energías con la línea azul y la corregida por el método de FFs con la línea verde. . . . .	61
A.1. Parámeteros de una elipse. . . . .	73

# ÍNDICE DE TABLAS

3.1.	Resumen de los WP de aislamiento para electrones y fotones usados a lo largo de esta tesis. . . . .	25
3.2.	Medidas de eficiencias de $b$ -tagging, y de rechazos de $c$ -jets, $l$ -jets y $\tau$ -jets, en los regímenes de bajo y alto $p_T$ . . . . .	32
3.3.	Medidas de eficiencia de $c$ -tagging para $c$ -jets, y valores de rechazos de $b$ -jets, $l$ -jets y $\tau$ -jets en los regímenes de bajo y alto $p_T$ . Los valores corresponden a aquellos luego de aplicar el voto del WP de $b$ -tagging de 77% y de 50% de $c$ -tagging. <b>rejection values not correct!</b> . . . . .	33
4.1.	Shower Shapes (SSs) utilizadas para la identificación de fotones. Las tres columnas de la derecha denotan si la variable es utilizada o no para los WPs <i>loose</i> (L), <i>medium</i> (L) o <i>tight</i> (T), descriptos en la Sección 4.2.2. . . . .	36

# 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 la Organización Europea para la Investigación Nuclear ([CERN](#)). En el presente capítulo, se ofrece una introducción al [LHC](#) en la Sección 2.1, seguida de una discusión del detector [ATLAS](#) en la Sección 2.2. Finalmente, en la Sección 2.3, se describe brevemente cuáles fueron las condiciones para la toma 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 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 Figura 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 inyectados en el [LHC](#). En el [LHC](#), 8 cavidades de radiofrecuencia pueden impulsar la energía de los protones hasta 14 TeV. Los cuatro puntos amarillos en la imagen Figura 2.1 son cuatro puntos de interacción entre las partículas acelaradas, que albergan los experimentos [ALICE](#) [4], [LHCb](#) [5], [CMS](#) [6], [ATLAS](#) [7], [LHCf](#) [8], [TOTEM](#) [9], [MoEDAL](#) [10], entre muchos otros.

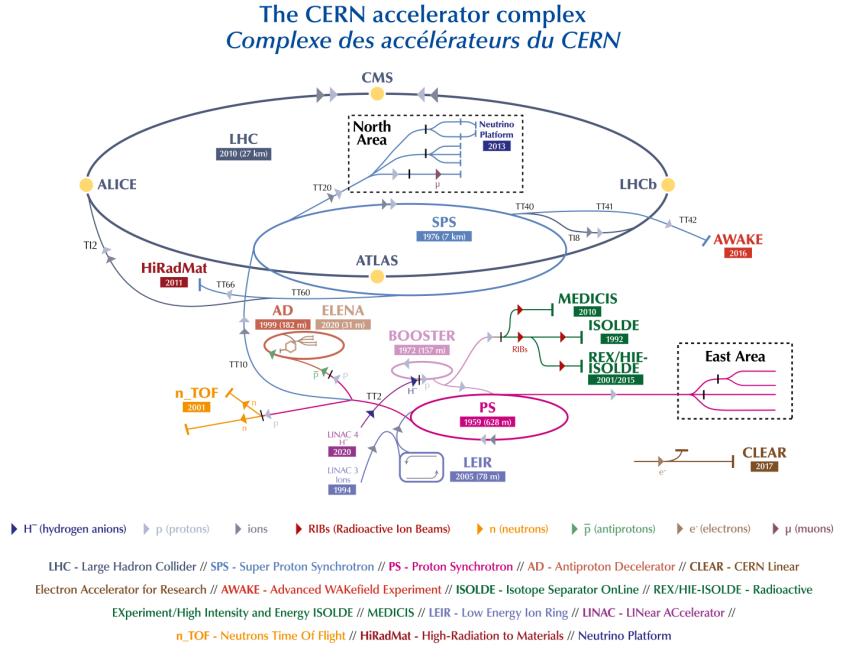


Figura 2.1: Vista general del complejo de aceleradores del LHC [3].

Los protones se inyectan en paquetes (*bunches*) 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 *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 área, 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,  $\gamma_r$  es el factor gamma relativista,  $\epsilon_n$  es la emitancia transversal normalizada del haz y  $\beta^*$  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  $\sim 11$  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  $\sim 40$  MHz.

La medida para el total de datos registrados se obtiene a partir de la luminosidad integrada a lo largo del tiempo y 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 Sección 2.3.

## 2.2. ATLAS

[ATLAS](#) es uno de los detectores multipropósito del [LHC](#). Fue diseñado y construido para estudiar las colisiones  $pp$  (y de iones pesados) y un gran espectro de procesos físicos en la escalada energía del TeV.

La forma general del detector es la de un cilindro, como se muestra en la Figura 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 (*barrel*), y las tapas exteriores (*end-caps*).

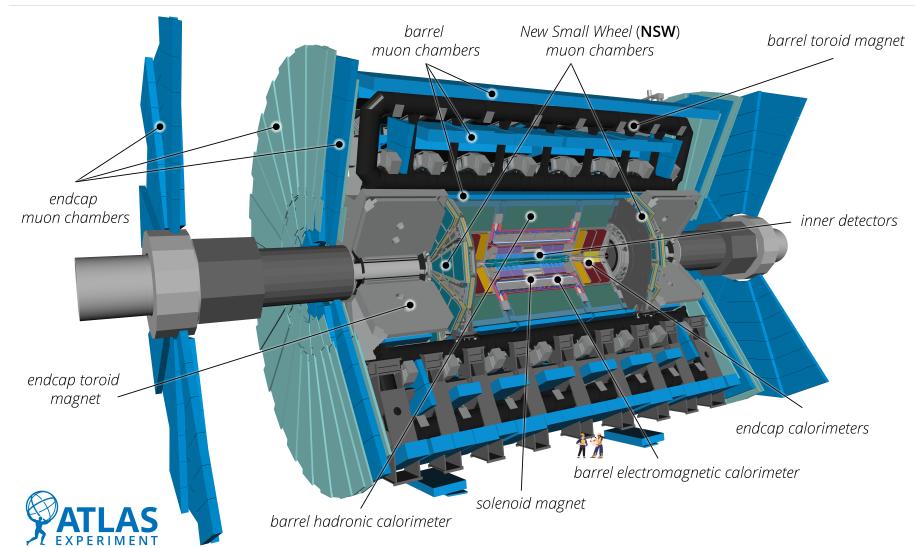


Figura 2.2: Vista general del detector [ATLAS](#) y de todos sus subdetectores, incluidos los sistemas añadidos durante el [LS2](#) [11].

[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  $\sim 2$  T para curvar las trazas 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 constituidas 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 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 también generan un campo magnético de hasta 4 T para los muones que salen en la dirección más próxima al haz.

### **2.2.1. Sistema de coordenadas de ATLAS**

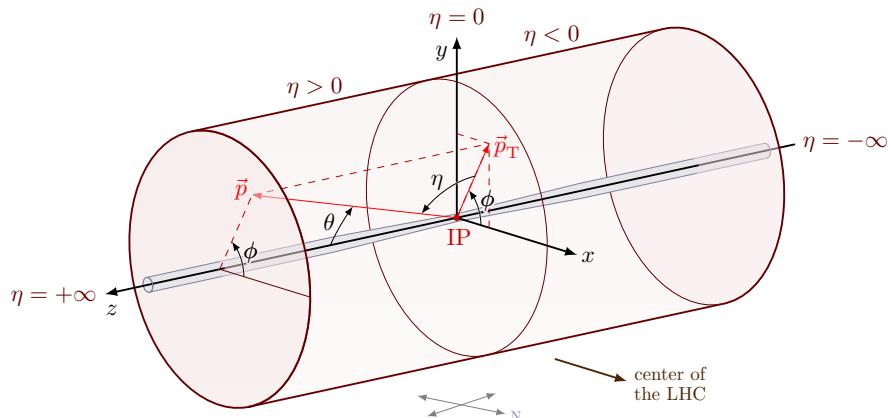


Figura 2.3: Sistema de coordenadas de ATLAS [11].

El sistema de coordenadas utilizado en [ATLAS](#), que se muestra en Figura 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  $\phi$ , y un ángulo polar  $\theta$  es el ángulo desde el eje del haz. En lugar de  $\theta$  se utiliza la rapidez  $y$  que 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:

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

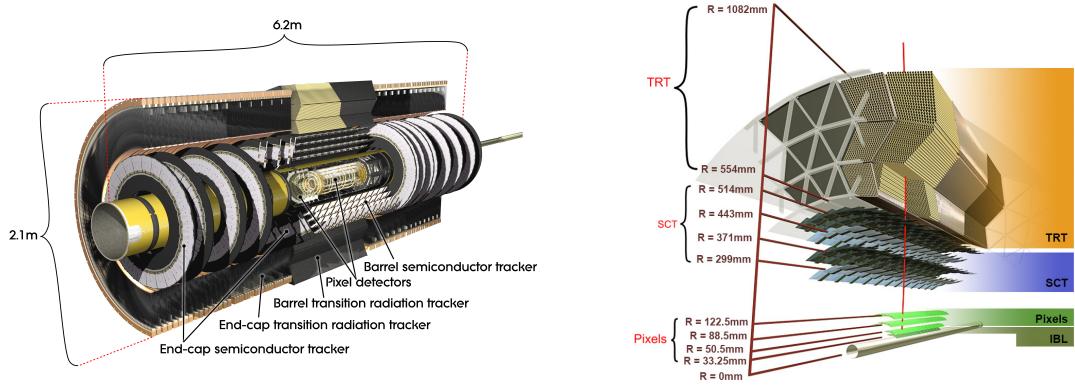
Para cuantificar la distancia entre dos objetos, se define  $\Delta R$ :

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

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**

El esquema de un corte transversal del Detector Interno (ID) [13] se muestra en la Figura 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



(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 [12].

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

(**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 reconstrucción 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 primarios, 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 que es crucial para la identificación de, por ejemplo, mesones *B* o leptones  $\tau$ . A continuación, se brinda una breve descripción de cada parte del **ID**.

**Insertable B-Layer (IBL)** Despues 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 [14, 15] 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 [16, 17]. 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 Figura 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  $50 \mu\text{m} \times 400 \mu\text{m}$  en el plano  $(\phi, z)$ , 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** [18]. 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 Figura 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 aproximadamente 50,000 tubos en la región del barrel, mientras que en los end-caps se sitúan aproximadamente 250,000 tubos.

### 2.2.3. Calorímetros

El sistema **ID** está rodeado por dos calorímetros: el Calorímetro Electromagnético (**ECAL**) y el Calorímetro Hadronico (**HCAL**), como se muestra en la Figura 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  $\phi$  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 partículas 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,  $X_0$ ,

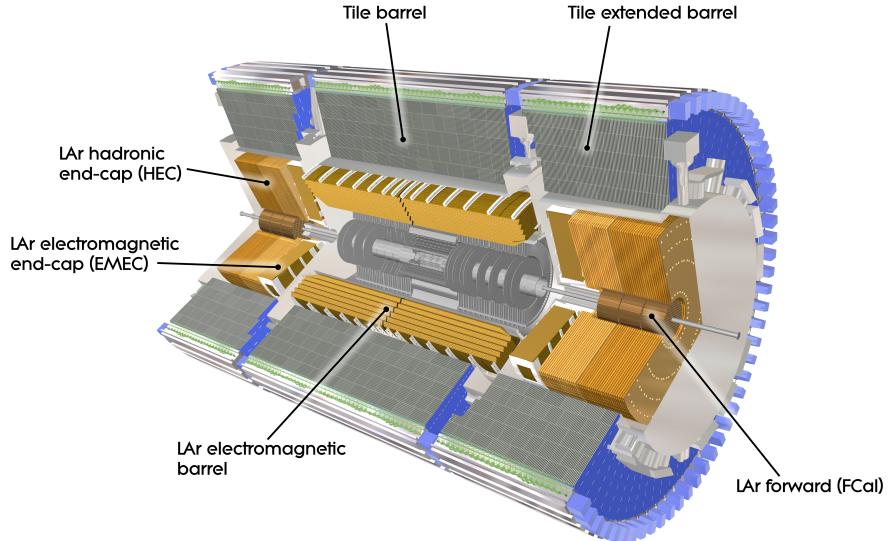


Figura 2.5: Sistema de calorímetros de ATLAS, mostrando el Calorímetro Electromagnético (ECAL) y el Calorímetro Hadronico (HCAL) [20].

y la longitud de interacción  $\lambda$ . 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  $\phi$  sin fisuras azimutales. Está dividido en dos medios bariles 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 (**F CAL**) ( $3.1 < \eta < 4.9$ ).

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 **H CAL**.

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 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.

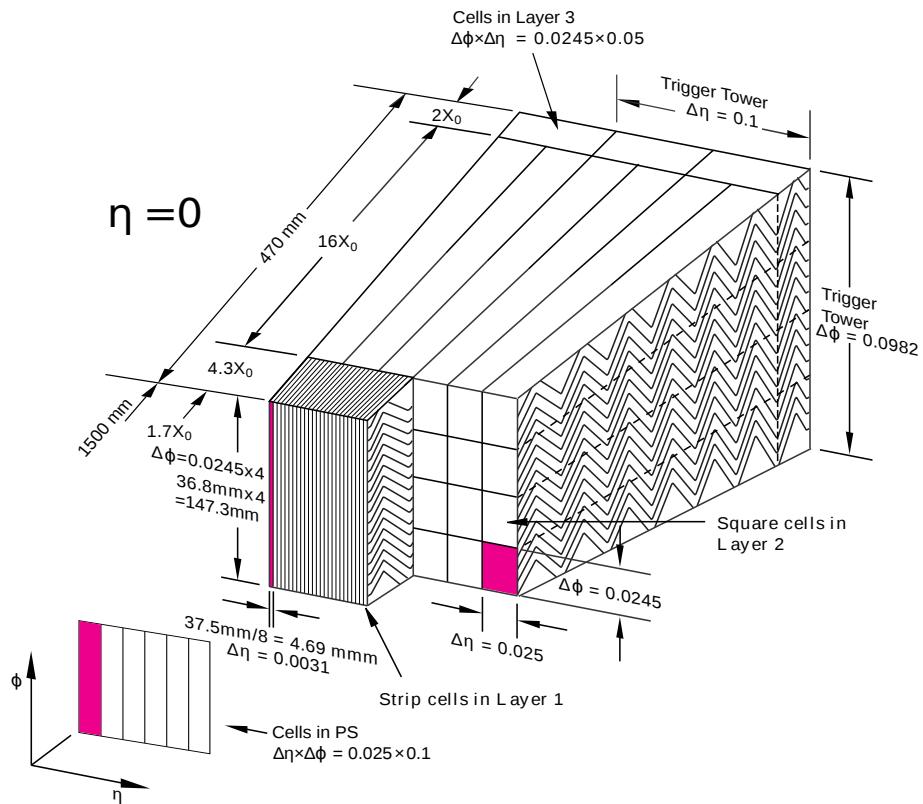


Figura 2.6: Segmento del **ECAL** mostrando la disposición de las capas y celdas del calorímetro. Además, se muestran las dimensiones de las celdas en cada capa [7].

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 Figura 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  $\eta$  (véase la Figura 2.7), y proporciona una medida precisa de  $\eta$ . 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  $0.025 \times$

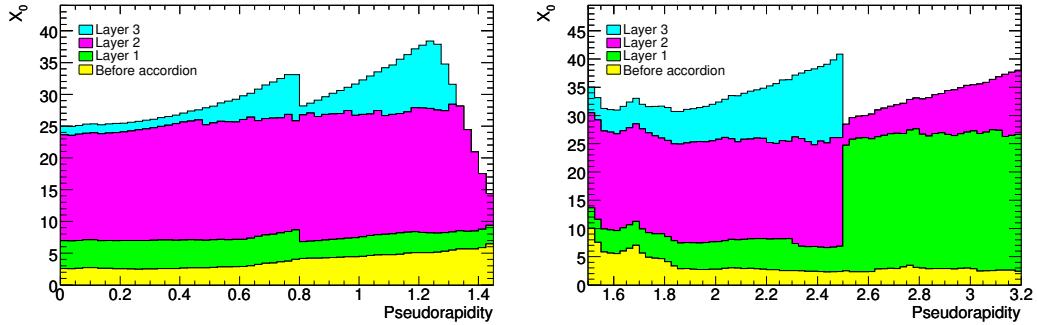


Figura 2.7: Longitudes de radiación en función de  $|\eta|$  para cada capa del [ECAL](#) [7].

0.025 en  $(\eta, \phi)$  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.

### 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 [22]. 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  $7.7\lambda$  en la región del barrel ( $9.7\lambda$  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)

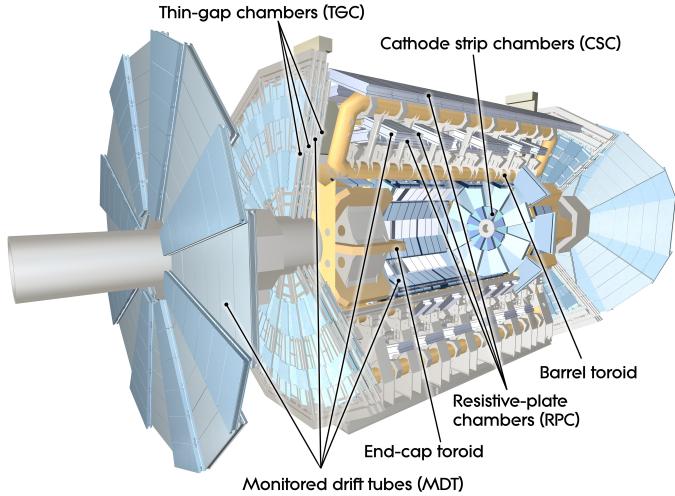


Figura 2.8: Diagrama del Espectrómetro de Muones (MS) [7].

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** [23] está situado en la parte más externa del detector **ATLAS**, insertado dentro del campo magnético de 4 T generado por los imanes toroidales del barrel y end-caps, y está diseñado para obtener medidas de posición y momento con 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 Figura 2.8, destacando los subsistemas.

El **MS** se compone de diferentes tipos de cámaras de detección (véase la Figura 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  $|\eta| < 1.05$ <sup>1</sup>. 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 reso-

<sup>1</sup> Durante el **LS2**, la capa del end-cap más interna ha sido reemplazada por las *New Small Wheels* (**NSWs**) [24]. Presenta MicroMegas como rastreadores de precisión ya que proporcionan un mejor rendimiento a las altas tasas esperadas en las futuras operaciones del LHC.

lución 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](#) [25–27] utiliza información del detector para rechazar eventos que no son de interés para el programa de [ATLAS](#) (colisiones *soft*, por ejemplo), reduciendo la frecuencia de eventos de 40 MHz (frecuencia de cruce de bunched mencionada en la Sección 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 almacenan para su posterior análisis, que podría llevar, por ejemplo, a un descubrimiento. 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 [28], que identifica lo que se conoce como Región de interés ([ROI](#)). La [ROI](#) consiste en celdas vecinas en los [ECAL](#) y [HCAL](#), 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 [29] 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 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 *menu* del trigger. 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 de forma tal 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 análisis a nivel de trigger, para el monitoreo del detector y las calibraciones del subsistema detector. Finalmente, los eventos aceptados por el HLT se almacenan 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*. Desde su puesta en marcha, se pueden distinguir los siguientes 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 periodo 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

dad 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  $pp$  del **LHC** 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  $156 \text{ fb}^{-1}$ , de los cuales **ATLAS** recolectó  $147 \text{ fb}^{-1}$ . La luminosidad integrada total disponible para análisis de física es de  $140.07 \text{ fb}^{-1}$ <sup>2</sup>, como se ve en la Figura 2.9a. La incertidumbre en la luminosidad integrada combinada para el Run-2 es de 0.83 % [30], obtenida usando el detector LUCID-2 [31]. Hasta el momento, combinando los años 2022, 2023 y 2024 de toma de datos del Run-3, se recogieron  $159 \text{ fb}^{-1}$  de datos, mostrados en la Figura 2.9b [32, 33].

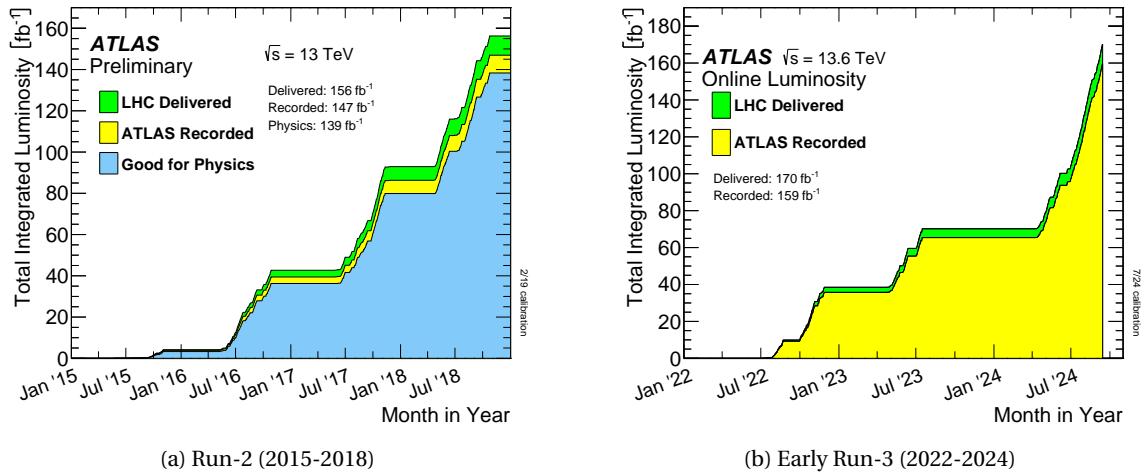


Figura 2.9: Luminosidad entregada por el **LHC** y recolectada por **ATLAS** durante el Run-2 [30] y el Run-3. En el caso de Run-2, también se muestra la fracción de datos recolectados que son útiles para análisis de física.

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,

<sup>2</sup> Las primeras medidas y **GRL** iniciales sólo brindaban un total de  $139 \text{ fb}^{-1}$  disponibles para análisis

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 Figura 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 Figura 2.10b.

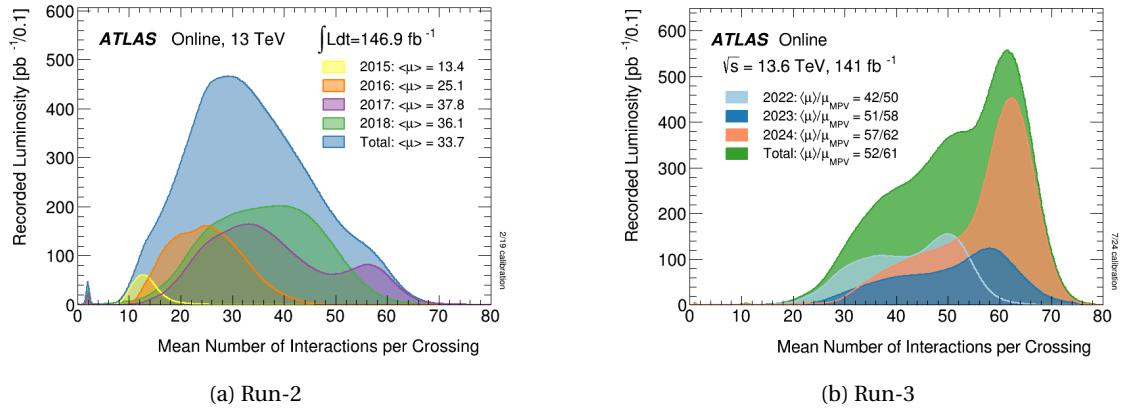


Figura 2.10: Distribución del número de interacciones por bunch-crossing durante Run-2 (izquierda) y Run-3 (derecha).

# RECONSTRUCCIÓN E IDENTIFICACIÓN DE OBJETOS FÍSICOS

# 3

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

---

Billie Jean King

Las partículas producidas en cada colisión y los productos de sus decaimientos, interactúan con el detector de una manera particular según su naturaleza. La información recogida por todos los subdetectores de [ATLAS](#) permite reconstruir e identificar los objetos físicos presentes en cada evento aceptado por el sistema de trigger (*selección online*). En este capítulo se describen los algoritmos de reconstrucción e identificación *offline*, que se lleva una vez que los eventos han sido aceptados por el trigger y almacenados.. 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](#) (Sección 2.2.2) se utiliza para reconstruir las trayectorias de las partículas cargadas, denominadas *tazas* (*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. Estas, permiten calcular 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 Figura 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  $z$ .
- $\phi_0$ : el ángulo azimutal.
- $\theta_0$ : el ángulo polar de la dirección de la partícula en el punto más cercano de aproximación [34].

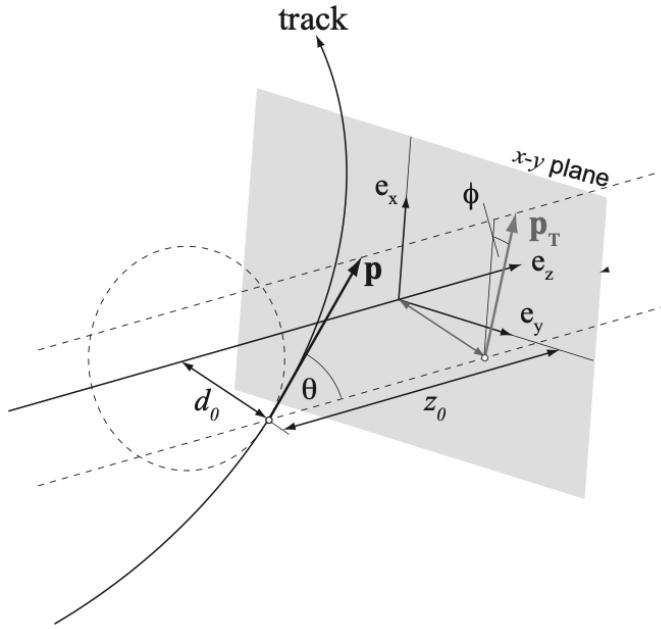


Figura 3.1: Esquema de los parámetros usados para la reconstrucción de trazas [35].

La reconstrucción de trazas utilizada durante el Run-2 utiliza dos enfoques complementarios: el enfoque *inside-out* y el *outside-in* [36].

El primer paso para la reconstrucción de trazas en el método *inside-out* es la búsqueda de semillas (*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 extrapola 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 [37] y se ajusta la traza candidata final utilizando un método global  $\chi^2$ . La última parte del método *inside-out* consiste en extender las trazas hasta el **TRT**, incluyendo los hits de este subdetector 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.

La identificación de los vértices de producción y decaimiento son de vital importancia para la posterior reconstrucción de objetos en [ATLAS](#). Para ello, un algoritmo de reconstrucción de vértices utiliza las trazas encontradas anteriormente [38, 39].

En primer lugar, el Vértice Primario ([PV](#)) se define como el lugar donde se da la interacción entre los dos protones injectados por el [LHC](#). Los [PVs](#) se reconstruyen en dos etapas: búsqueda de vértices, que asocia las trazas reconstruidas a los candidatos a [PV](#); y ajuste de vértices, donde las posiciones de los vértices es refinada de forma iterativa. En cada iteración se selecciona un conjunto de trazas y una posición semilla (*seed*) se usa para estimar el vértice, y las trazas incompatibles con este vértice son removidas para una futura iteración. 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  $\tau$  o los quarks más pesados ( $b$  o  $c$ -quarks), para los cuales, en [ATLAS](#), es posible determinar su vértice de decaimiento. 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](#) hace uso de las deposiciones 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.

### 3.2.1. Reconstrucción

La reconstrucción de fotones y electrones *offline* [40, 41] hace uso de clusters dinámicos de tamaño variable, conectados topológicamente entre las celdas del [ECAL](#) y [HCAL](#) [42], denominados *topo-clusters*, que se agrupan además en *superclusters*. Durante el Run-1 [43–45], en contraste, los clusters eran de tamaño fijo, y 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 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: un cluster construido a partir de los depósitos de energía en el [ECAL](#) el cual tiene asignado una traza.
- Fotones convertidos: un cluster asignado a 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.

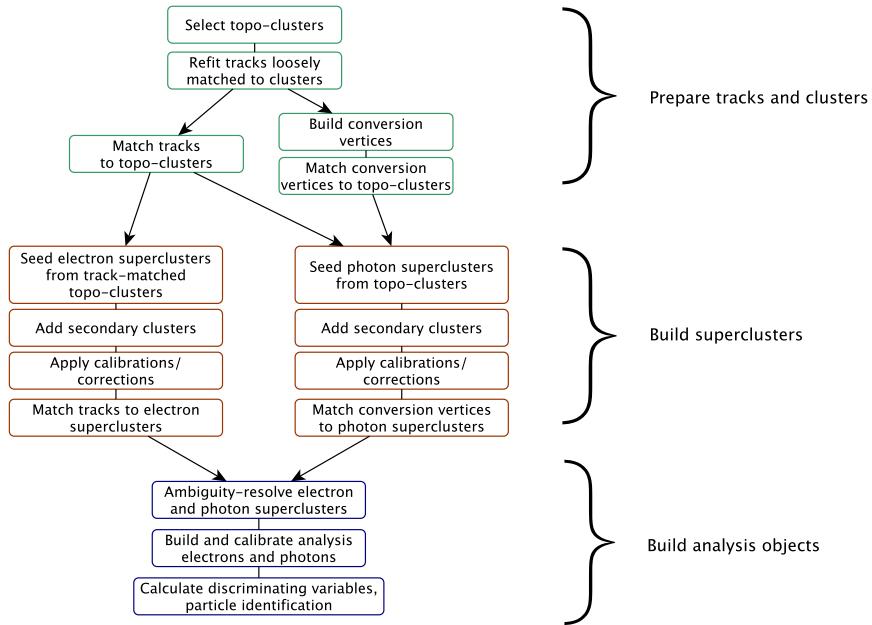


Figura 3.2: Diagrama del algoritmo de reconstrucción de electrones y fotones, extraído de Ref. [40]

El algoritmo para la reconstrucción de electrones y fotones procede como se muestra en la Figura 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 energía requerida y predefinida, y añadiendo posteriormente celdas vecinas, 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 la fracción de la misma con respecto a la energía total del topo-cluster es 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 que superan ciertos criterios de selección, se añaden a los candidatos a semilla para formar los superclusters finales. Finalmente 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 resuelve esta ambigüedad para eliminar parte del solapamiento. Sin embargo, se permite cierto solapamiento para mantener una alta eficiencia de reconstrucción de electrones y fotones, y para que en cada análisis de datos se apliquen criterios específicos a dicho estudio. Finalmente, se construyen y calibran los electrones y fotones finales.

### 3.2.2. Identificación

Con el objetivo de poder discriminar los objetos *prompt*<sup>1</sup> de aquellos que no lo son, existen diferentes criterios de identificación. En **ATLAS**, la identificación de fotones y electrones se logra mediante una serie de variables denominadas Shower Shapes (**SSs**) (descriptas en detalle en Capítulo 4). Estas, son variables que describen el paso de los fotones y electrones a través del **ECAL** y **HCAL**, caracterizando las lluvias electromagnéticas laterales y longitudinales, y son calculadas a partir de la energía depositada en las celdas de estos calorímetros. Mediante ciertos algoritmos que hacen uso de las **SSs**, 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 originados por quarks de sabores pesados (*heavy-flavor*). La identificación se basa en un método likelihood que utiliza algunas de las **SSs**, utilizando electrones provenientes de decaimientos de  $J/\psi$  y  $Z$  para bajo y alto  $E_T$ , respectivamente [46]. Se definen entonces **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 [47].

Para distinguir los fotones prompt/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. El proceso completo de identificación de fotones se presenta en Capítulo 4, donde las **SSs** se explican una a una. Además, en el Capítulo ?? se presentan dos enfoques para corregir las diferencias observadas en estas variables entre los datos y **MC**, uno de los principales objetivos de esta tesis.

### 3.2.3. Aislamiento

Para mejorar aún más la selección de fotones y electrones se aplican criterios de aislamiento a estos objetos. Para ello, se definen dos criterios de aislamiento: calorimétrico y de trazas.

El procedimiento para calcular la energía de aislamiento calorimétrico  $E_T^{\text{cone}40}$  se esquematiza en la Figura 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 Sección 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

---

<sup>1</sup> 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.

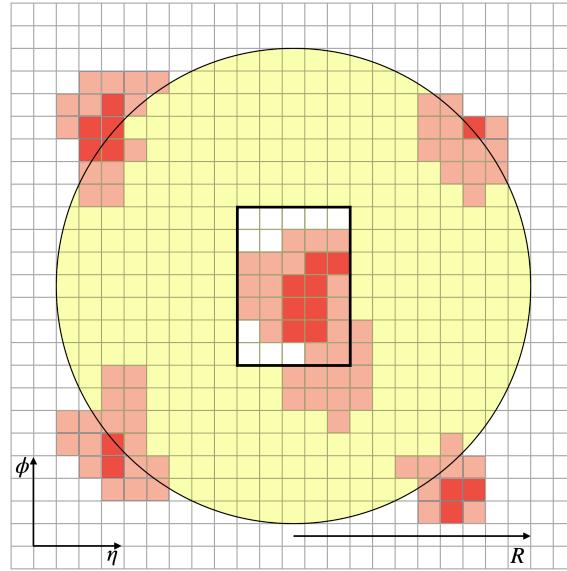


Figura 3.3: Diagrama del proceso del cálculo de la variable de aislamiento calorimétrico. Cuando se utiliza un cono con  $R = 0.4$ , se puede construir la variable  $E_T^{\text{cone}40}$  mencionada en el texto.

$5 \times 7$  (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 realizan correcciones para tener en cuenta las fugas de energía fuera del cono y las contribuciones de pileup [48]. La forma final de la energía de aislamiento calorimétrico resulta:

$$E_T^{\text{cone}40} = E_{T,\text{raw}}^{\text{isol}40} - E_{T,\text{core}} - E_{T,\text{leakage}} - E_{T,\text{pileup}}$$

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  $p_T > 1 \text{ GeV}$ , 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 es-

trategias: o bien conseguir una eficiencia fija, o bien aplicar cortes fijos en las variables de aislamiento. En el caso de los fotones, hay **WP**s que no utilizan ambas variables de aislamiento, como es el caso del **WP** que sólo utiliza el aislamiento calorimétrico. Las definiciones de los diferentes **WP**s utilizados a lo largo de esta tesis se muestran en la Tabla 3.1. Además, es común definir las siguientes variables para **WP FixedCutTight** del fotón:

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

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

lo que resulta para el **WP FixedCutTight** en:

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

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

### 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. [49]. 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

Los quarks y gluones no pueden detectarse de manera aislada, sino que por un proceso denominado hadronización, una vez producidos dan lugar a un chorro colimado de partículas que se denomina *jet*. Estos penetran a través del [ECAL](#) y son totalmente absorbidos por el material del calorímetro hadrónico. A continuación, se describen los métodos de reconstrucción de jets utilizados en [ATLAS](#)

### 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- $k_t$  [50]. Del mismo modo que para los electrones y los fotones, la reconstrucción de los jets en [ATLAS](#) comienza 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 los topo-clusters con los siguientes pasos:

- Determinación de 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 un valor fijo del algoritmo, que define el radio 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  $d_{ij} < d_{iB}$  se combinan los topo-clusters  $i$  y  $j$ , todas las distancias se calculan nuevamente 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- $k_t$  tiende a unificar las partículas *soft* con las *hard* y a separar a las partículas *hard* entre sí, 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 Figura 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.

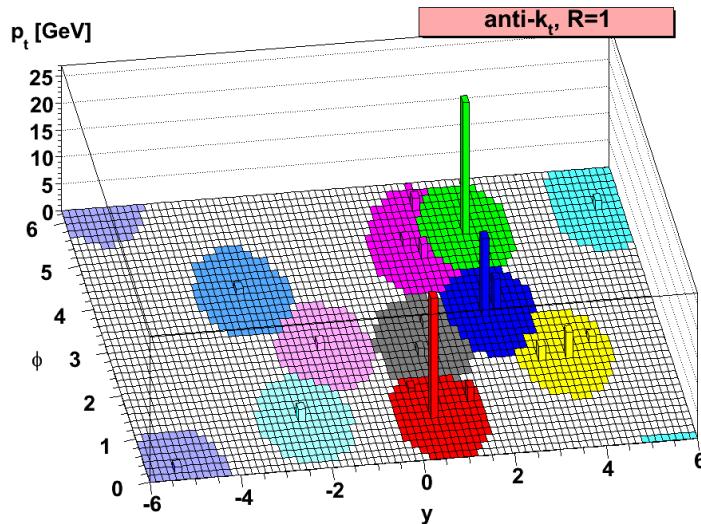


Figura 3.4: Representación esquemática del algoritmo anti- $k_t$  para el proceso de clusterización de jets [50].

### 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 la Sección 3.2.1, los depósitos de energía en las celdas del **ECAL** y **HCAL** se utilizan para construir topo-clusters, que se aproximan los depósitos de energía de hadrones individuales [41, 42]. Los jets reconstruidos de esta manera y agrupados con el algoritmo anti- $k_t$  con un radio de  $R = 0.4$  se denominan jets EM-TOPO. 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** [51], 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 y 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, el conjunto de topo-clusters, y un otro conjunto de topo-clusters modificados por el procedimiento anterior. La combinación de estos 3 conjuntos definen un objeto **PFlow**. Estos objetos también pueden ser agrupados con el algoritmo anti- $k_t$  y con  $R = 0.4$  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 **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**.

Cabe mencionarse, sin embargo, que para cualquier partícula cuya 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 **ID** 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 jets **PFlow**, ya que han demostrado proporcionar una mejor reconstrucción del jet [51].

### 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*<sup>2</sup>. Las tres primeras correcciones tienen en cuenta la contaminación de la distribución de pileup subyacente y las fluctuaciones debidas al origen del jet [52]. 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.

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  $N_{\text{PV}}$ . 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.

*Flavor Tagging* (FTAG) es la clasificación de los jets dependiendo del sabor (*flavor*) de los quarks por los que fueron iniciados, utilizando algoritmos sensibles a las propiedades distintivas de las respectivas clases. Entre ellos se consideran jets iniciados quarks  $b$  ( $b$ -jets),  $c$  ( $c$ -jets) o ni  $b$  ni  $c$  (jets livianos, también referidos como light-jets o  $l$ -jets). 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

---

<sup>2</sup> Los truth jets, o jets reales, son jets asociados a una partícula específica (jet iniciado por un quark de sabor liviano, por ejemplo) proveniente del estado final de una simulación, luego de pasar por el algoritmo de clusterización anti- $k_t$ .

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 [53–56].

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 [57] y a los algoritmos de bajo nivel que lo alimentan. Se basa en GN1 [58], que se refinó para pasar a ser GN2. GN2 sustituye la *Graph Attention Network* [59] utilizada por GN1 por un Transformador [60], y también se beneficia de otras optimizaciones de arquitectura y de poder realizar el entrenamiento de la red con un orden de magnitud más de datos para su entrenamiento.

El algoritmo cepta directamente información sobre el jet y las trazas asociadas y, como tal, no depende de otros algoritmos de etiquetado de sabores (*flavor tagging*). 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  $\tau$ . Las salidas de este tagger corresponden a las probabilidades de que un jet sea etiquetado (*taggado*) como un jet  $b$ ,  $c$ ,  $\tau$  o *light*, denominadas como  $p_b$ ,  $p_c$ ,  $p_\tau$  y  $p_u$ , 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  $c$ ,  $\tau$  y *light*. Las probabilidades de salida del tagger se combinan para construir un único discriminante  $\text{GN2}_b$ , 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 maximizando el rechazo de  $c$ -jets ( $\tau$ -jets) y  $l$ -jets, y resultan ser 0.2 (0.01).

A partir del valor discriminante del tagger, se pueden definir varios WPs, simplemente exigiendo que el valor  $\text{GN2}_b$  esté por encima de un determinado umbral. El grupo FTAG de ATLAS proporciona de forma centralizada a toda la colaboración cinco WPs diferentes para lograr una eficiencia global fija de  $b$ -tagging: 65, 70, 77, 85 y 90%, y se muestran en la Figura 3.5. 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.

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 están 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 Tabla 3.2, donde se muestran las eficiencias de

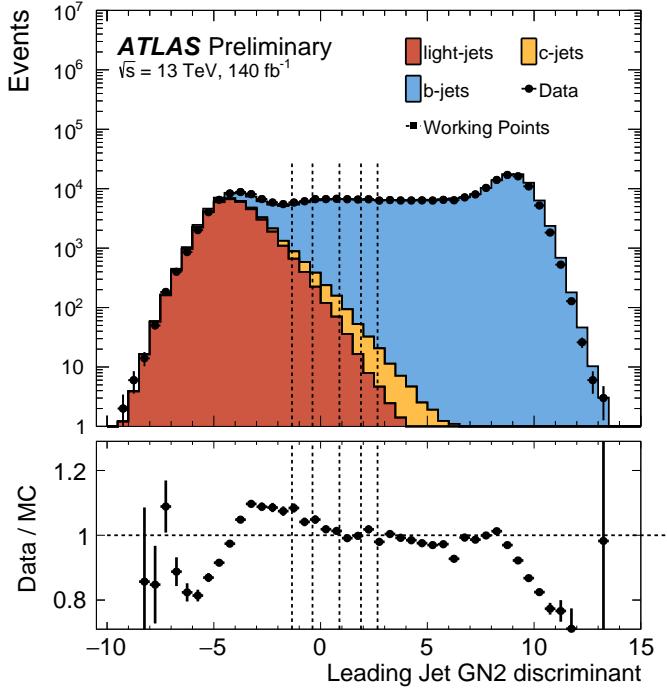


Figura 3.5: Comparación entre datos y simulación MC (eventos del proceso  $t\bar{t}$  semileptónico) del discriminante del tagger GN2. Las contribuciones de los jets  $l$ ,  $b$  y  $c$  se muestran con diferentes colores, y los 5 WPs de  $b$ -tagging se muestran con las líneas verticales punteadas. 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 la suma de las simulaciones MC [61].

tagging para  $b$ -jets, junto con los rechazos a  $c$ -jets,  $l$ -jets y  $\tau$ -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  $Z'$ <sup>3</sup> 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.

Tabla 3.2: Medidas de eficiencias de  $b$ -tagging, y de rechazos de  $c$ -jets,  $l$ -jets y  $\tau$ -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 $\tau$ -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,  $\tau$ -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  $b$ ,  $\tau$  y light. Utilizando el mismo procedimiento de optimización que para  $b$ -tagging, los valores para  $f_{b(\tau)}$  resultan ser 0.3 (0.05).

<sup>3</sup> 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.

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  $\tau$ -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  $\text{GN2}_c$ . Las medidas de eficiencia y rechazo de las dos muestras descriptas anteriormente, tras aplicar el **WP** de  $c$ -tag de 50% de eficiencia se muestran en el Tabla 3.3.

Tabla 3.3: Medidas de eficiencia de  $c$ -tagging para  $c$ -jets, y valores de rechazos de  $b$ -jets,  $l$ -jets y  $\tau$ -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 $\tau$ -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**

# SHOWER SHAPES Y LA IDENTIFICACIÓN DE FOTONES

# 4

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

---

Billie Jean King

El [ECAL](#) se presentó brevemente en la Sección 2.2.3, donde se describió el mecanismo que utiliza para la medida de energía y posición de fotones y electrones. En este subdetector, los fotones depositan su energía mediante la creación de pares electrón-positrón y la radiación bremsstrahlung, creando una lluvia electromagnética ([EM](#)). El [ECAL](#) hace un buen trabajo para calcular la energía de la lluvia [EM](#), pero identificar la partícula iniciadora sigue siendo una tarea difícil. Sin embargo, en virtud de las diferentes capas y granularidades en el [ECAL](#), pueden estudiarse diferentes características de estas lluvias [EM](#), codificadas por diferentes variables llamadas Shower Shapes ([SSs](#)).

El capítulo comienza con la descripción de todas las [SSs](#), que son centrales para poder identificar a los fotones reales de los falsos. La optimización del algoritmo de identificación de fotones está basada en las [SSs](#), y se encuentra descripta en la Sección 4.2. Además, en dicha sección, se presentan los métodos usados para la estimación de las eficiencias de identificación de fotones. Por último, en la Sección 4.3 se describe brevemente el mal modelado de las [SSs](#) en las simulaciones, un problema que tiene implicancias directas en los cálculos de eficiencias. Este mal modelado, sin embargo, será estudiado en detalle en el Capítulo 5.

## 4.1. Shower Shapes ([SSs](#))

Como se menciona en la Sección 3.2.2, la identificación de fotones se basa en cortes rectangulares aplicados a las diversas [SSs](#) que conducen a una excelente capacidad de separación entre fotones reales aislados de fotones falsos procedentes de hadrones. Estas [SSs](#) se calculan a partir de los depósitos de energía de los candidatos a fotones en las celdas del [ECAL](#) y [HCAL](#), y sirven para describir el paso de ellos a través de los calorímetros, caracterizando las lluvias [EM](#) laterales y longitudinales.

En general, los fotones reales producen depósitos de energía más estrechos en el **ECAL**, y tienen menores filtraciones hacia el **HCAL**, en comparación con aquellos fotones procedentes de hadrones, donde la presencia de hadrones vecinos adicionales cerca del fotón falso tienden a ensanchar las lluvias. Además, dado que la primera capa del **ECAL** presenta una gran resolución en  $\eta$ , es posible discriminar los candidatos a fotones procedentes de decaimientos  $\pi^0 \rightarrow \gamma\gamma$  que están caracterizados por dos máximos locales debidos a la presencia de dos fotones cercanos.

Tabla 4.1: Shower Shapes (SSs) utilizadas para la identificación de fotones. Las tres columnas de la derecha denotan si la variable es utilizada o no para los WPs *loose* (L), *medium* (M) o *tight* (T), descriptos en la Sección 4.2.2.

Categoría	Descripción	Nombre	L	M	T
Filtración hadrónica	Cociente entre el $E_T$ en la primera capa del <b>HCAL</b> y el $E_T$ del cluster <b>EM</b> ( $ \eta  < 0.8$ y $ \eta  > 1.52$ )	$R_{\text{had}1}$	✓	✓	✓
	Cociente entre el $E_T$ en todo el <b>HCAL</b> y el $E_T$ del cluster <b>EM</b> ( $0.8 <  \eta  < 1.37$ )	$R_{\text{had}}$	✓	✓	✓
<b>ECAL</b> (2 <sup>da</sup> capa)	Cociente entre la energía sumada en $3 \times 7$ celdas en $\eta \times \phi$ y la energía en $7 \times 7$ celdas, centradas alrededor del centro del cluster	$R_\eta$	✓	✓	✓
	Ancho lateral de la lluvia en dirección de $\eta$	$w_{\eta 2}$	✓	✓	✓
	Cociente de la energía sumada en $3 \times 3$ celdas en $\eta \times \phi$ y en $3 \times 7$ celdas, centradas alrededor del centro del cluster	$R_\phi$	✓	✓	
<b>ECAL</b> (1 <sup>ra</sup> capa)	Ancho lateral de la lluvia en 3 <i>strips</i> alrededor del máximo	$w_{\eta 1}$ or $w_1$	✓	✓	
	Ancho lateral total de la lluvia	$w_{\text{s tot}}$	✓	✓	
	Fracción de la energía fuera de las 3 strips centrales en un rango de 7 celdas, sobre la energía en las 3 celdas centrales	$f_{\text{side}}$	✓	✓	
	Diferencia entre la energía del segundo máximo con la energía mínima entre los dos primeros máximos.	$\Delta E$	✓	✓	
	Cociente de la diferencia de energía entre el primer y segundo máximo, sobre la suma de ambas energías	$E_{\text{ratio}}$	✓	✓	
	Cociente de la energía en la primera capa del <b>ECAL</b> y la energía total del cluster <b>EM</b>	$f_1$	✓	✓	

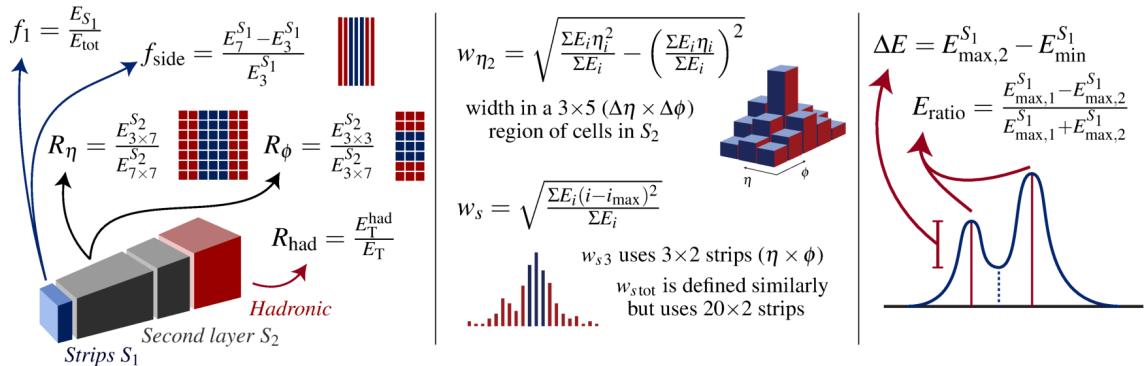


Figura 4.1: Representación esquemática de las SS de fotones. Los valores  $E_C^{S_N}$  representan la energía en la capa  $N$  del **ECAL** en un cluster  $C$ .

A continuación, se detallan las SSs utilizadas para la identificación de fotones, que se muestran resumidas en Tabla 4.1 y un esquema de cómo son calculadas se encuentra en la Figura 4.1. Las primeras variables hacen uso de la energía medida en el HCAL:

- Filtración hadrónica: es la energía trasversal depositada en el HCAL, normalizada respecto a la energía depositada en el ECAL:

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

Para minimizar los efectos de la degradación de la resolución, en la región de transición barrel-endcap del HCAL ( $0.8 \leq |\eta| \leq 1.37$ ) se utiliza la energía depositada en todo el HCAL ( $R_{\text{had}}$ ). En el resto del detector, sólo se utiliza la energía depositada en la primera capa del HCAL ( $R_{\text{had } 1}$ ).

Las siguientes variables utilizan la información de la segunda capa del ECAL:

- Perfil de energía lateral en  $\eta$ :

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

donde  $E_{i \times j}^{s2}$  es la suma de energía en la segunda capa del calorímetro contenida en una ventana de  $i \times j$  celdas (unidades de  $\eta \times \phi$ ), centrada en la celda más energética. Esta variable da una medida del ancho de las lluvias en la dirección  $\eta$ .

- Perfil de energía lateral en  $\phi$ :

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

definida de forma similar a  $R_{\eta}$ . Sin embargo, esta variable se comporta de forma muy diferente para fotones convertidos y no convertidos. Debido a la acción del campo magnético, los electrones y positrones se curvan en direcciones opuestas en  $\phi$ , por lo que se producen lluvias EM más anchas para los fotones convertidos que para los no convertidos.

- Ancho de la lluvia lateral en  $\eta$ :

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

mide el ancho propio de la lluvia EM, donde  $E_i$  es la energía en la  $i$ -ésima celda del ECAL, medida en una ventana de  $3 \times 5$  celdas en  $\eta \times \phi$ .

Las siguientes variables utilizan la información de la primera capa del ECAL, compuesta por las celdas *strips* que permiten una alta resolución en  $\eta$  y permite una buena separación entre fotones aislados de fotones producto del decaimiento de  $\pi^0$ . La Figura 4.2 muestra la diferencia en la energía depositada en el ECAL entre los dos casos mencionados anteriormente.

- Perfil de energía lateral en  $\eta$ :

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

mide la energía fuera del núcleo de las tres strips centrales dentro de una ventana de 7 celdas, dividida por la energía en las tres celdas centrales.

- Ancho de la lluvia lateral en  $\eta$  (3 strips)

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

donde  $i$  corre sobre todas las celdas en una ventana de 3 celdas alrededor de la de mayor energía. Esta variable mide el ancho de la lluvia EM en la primera capa del calorímetro.

- Ancho de la lluvia lateral en  $\eta$  (total). Se define de forma similar a  $w_1$ , pero utiliza todas las celdas en una ventana de  $\Delta\eta \times \Delta\phi = 0.0625 \times 0.2$ , que corresponde aproximadamente a  $20 \times 2$  strips en  $\eta \times \phi$ .
- Diferencia energética

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

representa la diferencia de energía entre el segundo máximo y la energía mínima reconstruida entre los dos máximos de la primera capa del ECAL.

- Cociente de energía

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

es la relación de la diferencia de energía entre los dos máximos, normalizada con respecto a la suma de esas energías, en la primera capa del ECAL.

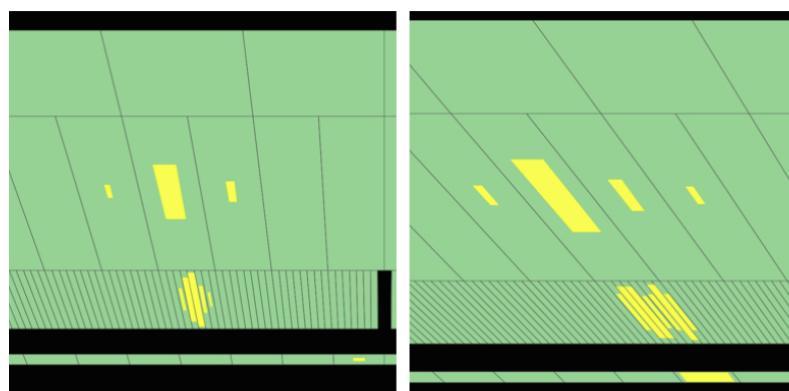


Figura 4.2: Depósitos de energía característicos para un fotón aislado (izquierda), y un evento  $\pi^0 \rightarrow \gamma\gamma$  (derecha), que es posible distinguir gracias a la fina granularidad de la primera capa del ECAL [62].

## 4.2. Identificación de fotones

La identificación de fotones prompt frente a fotones falsos en colisiones hadrónicas es un gran desafío. Los fotones falsos están ampliamente dominados por candidatos a fotones que surgen del decaimiento de hadrones en jets, mientras que una fracción más pequeña de candidatos falsos está asociada con hadrones que depositan una energía significativa en el [ECAL](#), imitando la de los fotones reales. Los procesos con fotones prompt en el estado final, que ocurren en colisiones  $pp$  en el [LHC](#), desempeñan un papel central en el programa de física de [ATLAS](#). Tanto para las búsquedas como para las medidas de precisión, es importante contar con algoritmos y técnicas para identificar los fotones reales frente a los falsos. Estas búsquedas o medidas de precisión se llevan a cabo en un rango muy amplio de energía del fotón, empezando, por ejemplo, por resonancias de baja masa del Higgs a un par de partículas tipo axión que decaen en 4 fotones ( $H \rightarrow aa \rightarrow 4\gamma$ ) [63], donde el momento transverso del fotón es de  $\sim 25$  GeV, hasta fotones de muy alto  $p_T$  en búsquedas de resonancias  $\gamma+jet$ , donde los fotones tienen un momento transverso mayor a 1 TeV [add reference once the chapter numbers are in place](#).

El identificación de fotones en [ATLAS](#) se basa en cortes en las [SSs](#) y se definen diferentes *Working Points* ([WPs](#)) con diferentes características: ya sea lograr un gran rechazo de fondo, o alta eficiencia de señal, o simplemente bajos tiempos de cómputo. En esta sección se describe el procedimiento utilizado para la optimización de estos [WPs](#) y luego se describen los métodos para medir las eficiencias correspondientes de cada [WP](#).

### 4.2.1. Procesos de interés y selección de eventos

Dado el amplio rango en el que se utilizan los fotones en [ATLAS](#), la optimización de los [WPs](#) utiliza dos procesos diferentes que permiten obtener muestras limpias de fotones en los regímenes de bajo y alto  $p_T$ . En el caso de bajo  $p_T$ , se utiliza una fuente muy limpia de fotones procedentes de decaimientos radiativos del bosón  $Z$ . Por otro lado, aunque con mayor contaminación de fondo, se emplean eventos Single Photon ([SP](#)) para fotones de alto  $p_T$ . En los siguientes párrafos se ofrece una descripción de cada de las muestra de fotones utilizadas.

**Decaimientos radiativos del bosón  $Z$**  En el rango de bajo  $p_T$ , los fotones procedentes del decaimiento radiativo del bosón  $Z$  ( $Z \rightarrow \ell\ell\gamma$ ) se seleccionan como fotones de señal. Hay dos modos de producción posibles para los procesos del Modelo Estándar ([SM](#)) de  $pp \rightarrow Z(\ell^+\ell^-)\gamma$ , donde  $\ell$  es un electrón o un muón. Estos son: Radiación de estado inicial ([ISR](#)) donde el fotón es radiado desde los quarks, y Radiación de estado final ([FSR](#)) (de ahora en más también referido como decaimiento  $Z$  Radiativo ([RZ](#))), donde el fotón es radiado desde uno de los leptones del estado final a través de bremsstrahlung. Ambos modos de producción se muestran en la Figura 4.3.

Ambos procesos [FSR](#) y [ISR](#) pueden identificarse fácilmente comparando la distribución de la masa invariante de los dos leptones ( $m_{\ell\ell}$ ) con la distribución de la masa invariante de los dos

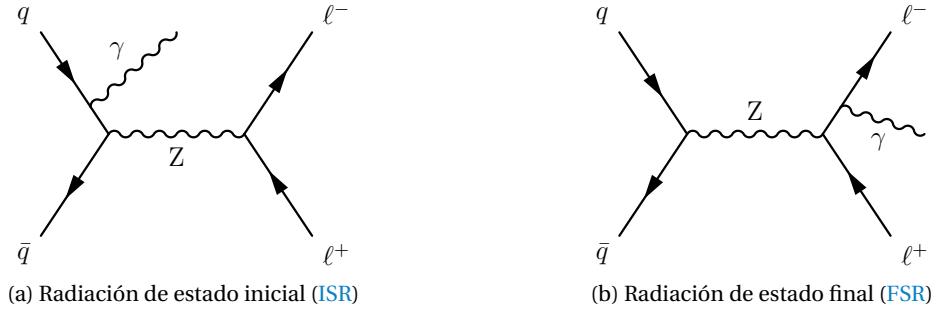


Figura 4.3: Diagramas de Feynman del proceso de radiación de un fotón en decaimientos  $Z \rightarrow \ell\ell\gamma$  para los casos de **ISR** (izquierda) y **FSR** (derecha).

leptones junto con el fotón ( $m_{\ell\ell\gamma}$ ). Para los eventos **ISR**,  $m_{\ell\ell}$  tiene su máximo en la masa del  $Z$ , y el fotón simplemente suma a la masa invariante de tres cuerpos ( $m_{\ell\ell\gamma}$ ) haciéndola mayor que 91 GeV. En el caso **FSR**, en cambio, la masa invariante de tres cuerpos  $m_{\ell\ell\gamma}$  presenta su máximo en la masa del  $Z$ , que se observa en la Figura 4.4a. Para los estudios de identificación de fotones, sólo se consideran los fotones de los eventos **FSR**. La razón detrás de la selección de **FSR** sobre **ISR** se debe a lo siguiente. Los eventos **ISR** también sufren la contaminación de fondos provenientes de eventos  $Z + \text{jets}$ , en los que el jet se identifica erróneamente como un fotón, y además la sección eficaz  $Z + \text{jets}$  es de órdenes de magnitud mayor a la del proceso  $Z + \gamma$ . A partir de las Figuras 4.4b y 4.4c, donde se muestran las distribuciones de  $m_{\ell\ell}$  en función de  $m_{\ell\ell\gamma}$  para los procesos  $Z \rightarrow \ell\ell$  y  $Z \rightarrow \ell\ell\gamma$ , respectivamente, se puede apreciar la separación entre estos dos procesos cuando se seleccionan fotones **FSR**.

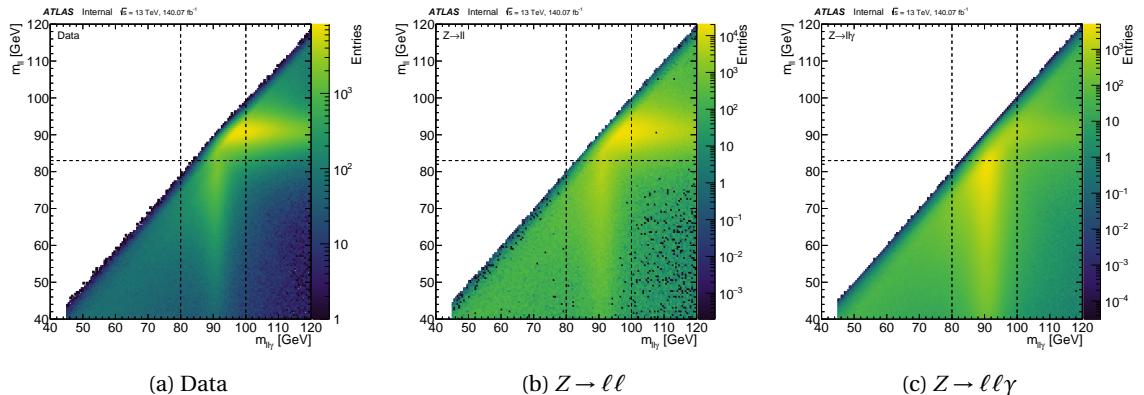


Figura 4.4: Masa invariante de los dos leptones en función de la masa invariante de ambos leptones junto con un fotón en (a) datos, (b) fondos y (c) señal. La región en la cual se encuentra una gran concentración de eventos con  $m_{\ell\ell} \sim m_Z$  corresponde a eventos de **ISR**, mientras que eventos de **FSR** están caracterizados por  $m_{\ell\ell\gamma} \sim m_Z$ .

Con la muestra **RZ**, se requiere que los fotones tengan un momento transverso  $p_T > 7$  GeV y una pseudorapidez en el rango de  $|\eta| < 1.37$  o  $1.52 < |\eta| < 2.37$ , evitando así la región del crack. Para los estudios de optimización no se aplica ningún requisito de aislamiento sobre los fotones, pero para las medidas de eficiencia se utiliza el **WP** de aislamiento *loose*, descripto en la Sección 3.2.3. Se requiere que los leptones tengan  $E_T > 10$  GeV, los muones una pseudorapidez  $|\eta| < 2.5$  y para los electrones  $|\eta| < 2.47$ , excluyendo el crack. Tanto a los electrones como a los muones se les exige que pasen los requisitos de aislamiento *loose* y que pasen el criterio de identificación *medium*.

El fotón **FSR** se selecciona entonces requiriendo  $80 < m_{\ell\ell\gamma} < 100$  GeV y  $40 < m_{\ell\ell} < 83$  GeV. Finalmente, para evitar cualquier sesgo en las **SS** del fotón y en sus variables de aislamiento, se requiere una distancia mínima de  $\Delta R > 0.4$  entre dicho fotón y el leptón más cercano.

**Single Photon (SP)** La muestra de fotones inclusivos, o Single Photon (**SP**), se recoge mediante triggers de un sólo fotón, cuyos umbrales varían entre 10 GeV y 140 GeV y requieren una identificación *loose*. Aunque los triggers utilizados para obtener esta muestra están preescalados (con la excepción del de 140 GeV), proporcionan un gran conjuntos de datos de fotones de alto  $p_T$ . Estos procesos incluyen eventos a Leading Order (**LO**) de  $\gamma + \text{jet}$  procedentes de la dispersión dura  $qg \rightarrow q\gamma$  y  $q\bar{q} \rightarrow g\gamma$ , así como fotones prompt procedentes de la fragmentación de quarks en eventos de dijet de **QCD**. Se requiere que estos fotones presenten una pseudorapidez de  $|\eta| < 2.37$  excluyendo el crack, y pasar el requisito de aislamiento *loose*. Las muestras de **SP** se utilizan tanto para los estudios de optimización como para la estimación de las eficiencias.

### 4.2.2. Optimización

A partir de las **SSs** anteriormente descriptas, se pueden definir tres **WPs** para los fotones: *loose*, *medium* y *tight* [46]. El **WP** *loose* emplea cortes a las variables definidas en la segunda capa y a la variable de filtración hadrónica, y es utilizado principalmente por el trigger. Los **WPs** *medium* y *tight* utilizan todas las variables definidas previamente. El primero está optimizado para tener una eficiencia fija de 95 %, mientras que el segundo proporciona un excelente rechazo de fondo. La Tabla 4.1 muestra qué variables se utilizan para cada **WP**.

Para la optimización de los **WPs** se utilizan las dos muestras definidas previamente: los eventos **RZ** para fotones con  $10 < p_T < 25$  GeV como señales y eventos de  $Z \rightarrow \ell\ell$  como fondos; y para el régimen de alto  $p_T$  ( $p_T > 25$  GeV) los eventos de **SP** se consideran como señal mientras que los eventos dijet son los fondos.

La identificación de fotones utiliza cortes en las **SSs**. En la Figura 4.5 se muestran ejemplos de tres de estas **SS** ( $R_\eta$ ,  $E_{\text{ratio}}$  y  $w_{\eta 2}$ ) comparando eventos de señal y de fondo utilizando las muestras de **RZ**, donde se observa un excelente poder discriminatorio. Los cortes en todas las **SSs**, para cada **WP** de identificación, se optimizan en función de la energía transversal y la pseudorapidez del candidato a fotón, para tener en cuenta la forma de las variables para diferentes  $\eta$  y para variaciones en la cantidad de material y la geometría del calorímetro. Los **WPs** *medium* y *tight* también se optimizan por separado para fotones convertidos y no convertidos. Los cortes se optimizan utilizando un enfoque Multivariable (**MV**), en el que las eficiencias de señal se escanean entre 0 % y 100 % mientras se intenta maximizar el rechazo de fondo.

### 4.2.3. Estimación de las eficiencias

Una vez optimizados los diversos **WPs** de identificación, es fundamental contar con las eficiencias ya sea de los datos y las simulaciones **MC**. Las estimaciones de eficiencias de fotones hace uso de tres métodos diferentes y característicos en diferentes rangos de  $p_T$ , que son detallados

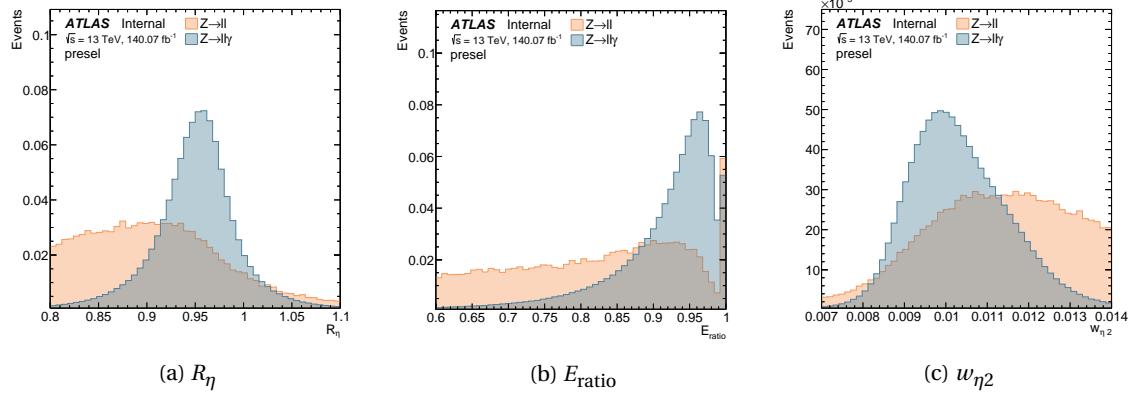


Figura 4.5: Distribuciones normalizadas de señal (azul) y fondo (naranja) de diferentes SSs utilizando las muestras de RZ y pasando la selección de eventos detallada en la Sección 4.2.1.

en la Ref. [64], y son brevemente descriptos en los próximos párrafos. Para los tres métodos, se requiere que los fotones satisfagan el criterio de aislamiento loose definido en la Sección 3.2.3 y, por tanto, las eficiencias de los fotones se miden con respecto a este criterio de aislamiento.

Para el rango de bajo  $p_T$  ( $7 < p_T < 100 \text{ GeV}$ ), los fotones procedentes del pproceso RZ se utilizan como fotones de señal. El método para estimar las eficiencias se basa en el uso de ajustes a la distribución observada de la masa invariante de tres cuerpos ( $m_{\ell\ell\gamma}$ ) antes y después de aplicar el criterio de identificación tight. El número de eventos de señal y de fondo puede contarse a partir de los ajustes, y las purezas de señal se calculan antes ( $P^{\text{total}}$ ) y después ( $P^{\text{pass}}$ ) de la aplicación de la identificación tight. La eficiencia final en los datos viene dada por:

$$\epsilon_{ID} = \frac{P^{\text{pass}} N_{\text{data}}^{\text{pass}}}{P^{\text{total}} N_{\text{data}}^{\text{total}}}.$$

El segundo método para calcular eficiencias se basa en transformaciones de Smirnov [65] para que las distribuciones de las SSs de los electrones se parezcan a los de los fotones. Las muestras usadas en este método son decaimientos  $Z \rightarrow ee$ , en los que se requiere que los electrones pasen el criterio de aislamiento de fotones loose. También es necesario tener en cuenta la contribución de una pequeña fracción de fondo de procesos  $W + \text{jets}$  y producción multijet. Estos fondos son tratados mediante ajustes a la distribución de  $m_{ee}$  de datos, utilizando señales simuladas y formas funcionales que describen el resto de los fondos, obtenidas en regiones de control (regiones en donde las contribuciones de estos fondos son dominantes y separables). Luego, los candidatos a electrones se cuentan a partir de eventos en el rango  $70 < m_{ee} < 110 \text{ GeV}$ , y las eficiencias se miden utilizando el método tag-and-probe descrito en la Ref. [40]. El rango  $p_T$  en el que se aplica este método es  $25 < p_T < 250 \text{ GeV}$ .

El último y tercer método utiliza muestras de SP con fotones en el rango  $50 < p_T < 1500 \text{ GeV}$ . En este caso se utiliza el *Matrix Method* [64], que construye cuatro regiones ortogonales que pasan o no el WP de identificación tight, y pasan o no el aislamiento de trazas (descripto en la Sección 3.2.3). Para cada región, surgen dos incógnitas: el número de eventos de señal y de fondo. Si se conocen las eficiencias de aislamiento de trazas para los componentes de señal y de fondo, entonces es posible estimar la eficiencia de los fotones loose que pasan los criterios

de identificación tight. Las eficiencias de aislamiento para los fotones de señal se estiman utilizando muestras de **MC**, y las de fondo se obtienen en una región de control enriquecida con jets construida a partir de la inversión de los criterios de identificación. Las eficiencias en datos para el **WP** de identificación tight son entonces:

$$\varepsilon^{\text{tight-ID}} = \frac{\frac{\hat{\varepsilon}_{\text{ID}} - \hat{\varepsilon}_{\text{ID}}^b}{\hat{\varepsilon}_{\text{ID}}^s - \hat{\varepsilon}_{\text{ID}}^b} \cdot N_{\text{ID}}^T}{\frac{\hat{\varepsilon} - \hat{\varepsilon}^b}{\hat{\varepsilon}^s - \hat{\varepsilon}^b} \cdot N^T},$$

donde  $N^T$  representa la totalidad de fotones en la muestra inclusiva que consiste en  $N^s$  fotones de señal (o fotones prompt) y  $N^b$  fotones falsos (fotones de fondo). El número  $N_{\text{ID}}^T$  es el subconjunto de  $N^T$  que pasa el requisito de identificación. Las eficiencias de aislamiento de traza de datos, señal y fondo se representan con  $\hat{\varepsilon}$ ,  $\hat{\varepsilon}^s$  y  $\hat{\varepsilon}^b$ , respectivamente. Del mismo modo, las eficiencias de aislamiento de traza para los fotones que superan la identificación tight se muestran como  $\hat{\varepsilon}_{\text{ID}}$ ,  $\hat{\varepsilon}_{\text{ID}}^s$  y  $\hat{\varepsilon}_{\text{ID}}^b$ , respectivamente. Las eficiencias medidas para los fotones con  $p_T > 150$  GeV están entre 90 y 96 %.

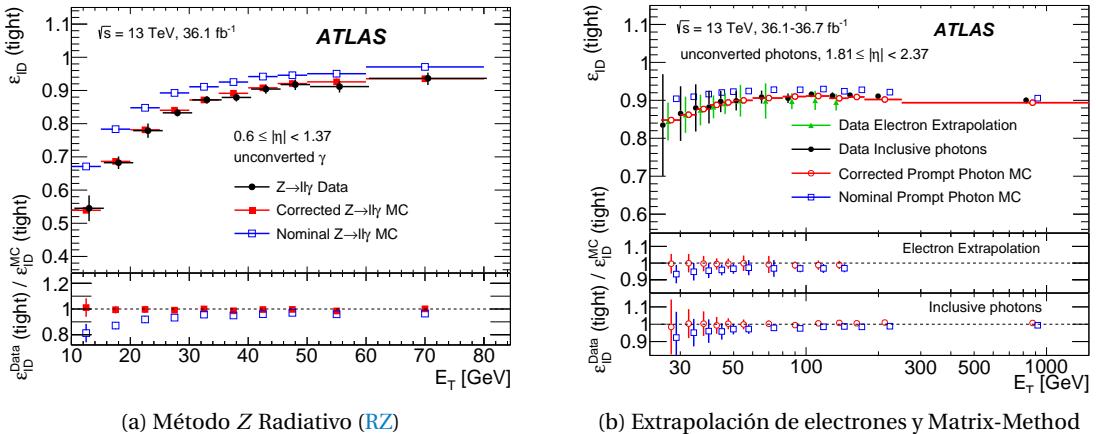


Figura 4.6: Comparación de las eficiencias calculadas para datos y **MC** utilizando los tres métodos diferentes para su cálculo. En ambas figuras, para cada método, se muestran dos conjuntos diferentes de mediciones **MC**: la nominal y la corregida (discutida en el texto). Los paneles inferiores muestran el cociente entre las eficiencias de los datos y las predicciones **MC** (denominadas Factor de escalas (**SFs**) en el texto). Las figuras fueron tomadas de la Ref. [64].

En la Figura 4.6a se muestra un ejemplo de las eficiencias de identificación en función del  $p_T$  del fotón utilizando el método **RZ**. Las eficiencias de los datos están representadas por los puntos negros, mientras que el **MC** nominal se muestra con cuadrados azules vacíos. Los cocientes de datos con el **MC** nominal (también denominados Factor de escala (**SF**)) mostrados en el panel inferior difieren hasta en un 20 % de 1, lo que demuestra que las eficiencias calculadas con **MC** difiere de las calculadas con los datos. Sin embargo, también se muestra otro conjunto de eficiencias, pertenecientes a simulaciones de **MC** corregidas, mejorando drásticamente el acuerdo entre los datos y la simulación, como se ve de los **SFs**. La razón por la que se necesitan estas correcciones y cómo se implementaron en **ATLAS** se explica en la siguiente sección (Sección 4.3), y cómo se corrigen en el Capítulo 5. La Figura 4.6b muestra las medidas de eficiencia usando los dos métodos restantes (extrapolación de electrones y Matrix Method), donde se obtienen las mismas mejoras en los **SFs** cuando se usa la simulación corregida.

Como ya se ha mencionado, estos cocientes entre datos y eficiencias MC se denominan SFs y encapsulan las diferencias entre datos y simulación. Se calculan por separado para cada uno de los tres métodos y después se combinan utilizando una media ponderada [66] en cada *bin* y asumiendo que las incertezas estadísticas y sistemáticas no están correlacionadas entre los métodos. Los resultados actuales de estos SFs, calculados utilizando el conjunto completo de datos de Run-2, se muestran en la Figura 4.7. La obtención de los SFs es de vital importancia, ya que luego son factores que se aplican a los eventos de MC para ser corregidos, y así lograr una comparación justa con los datos recolectados.

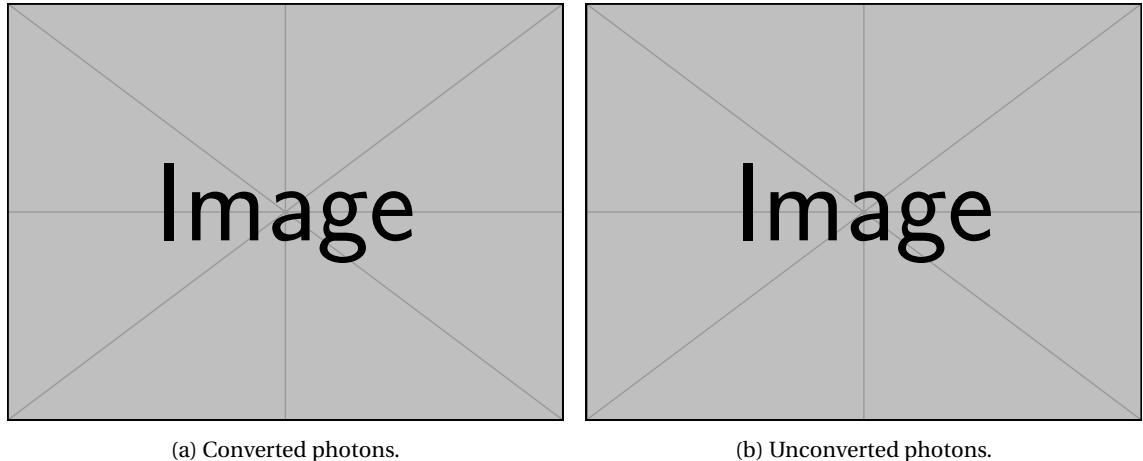


Figura 4.7: SFs resultantes de la identificación de fotones en los diferentes bins de  $p_T$  y  $|\eta|$  para fotones convertidos (izquierda) y no convertidos (derecha). [Ask fran for the plots in his presentation!](#)

### 4.3. Las diferencias de las Shower Shapes entre datos y MC

Como se ha mostrado anteriormente, la simulación MC no describe perfectamente los datos, lo cual puede verse de los valores de los SFs los cuales difieren de 1. En particular, al comparar las distribuciones de las SSs, se observa que las distribuciones MC están desplazadas o incluso la forma difiere, como se muestra en la Figura 4.8, al comparar los datos (puntos negros) con el histograma representado por la línea roja, correspondiente al MC.

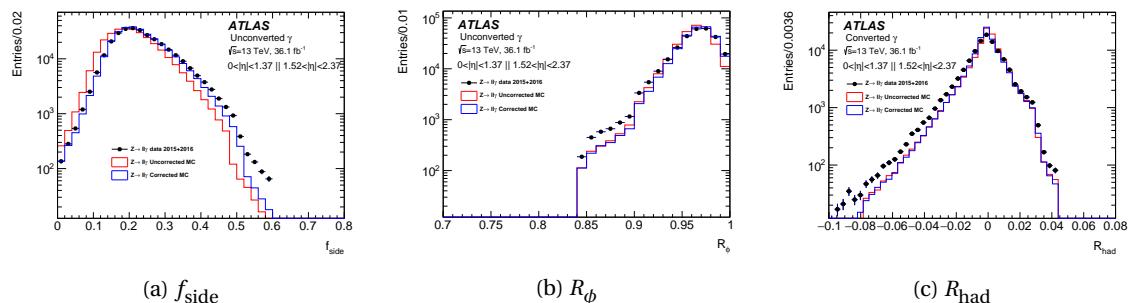


Figura 4.8: Comparaciones de las SSs entre los datos (puntos negros) y la simulación MC nominal (línea roja) y corregida (línea azul), utilizando la muestra RZ [47].

Las principales diferencias en las distribuciones surgen para los perfiles de la lluvia en la dirección de  $\eta$ , donde se observan distribuciones más amplias en los datos en comparación con MC.

Parte del efecto se corrigió en 2010 tras pasar a una descripción detallada de la composición del material absorbente del [ECAL](#) en [GEANT4](#). Sin embargo, algunas discrepancias entre los datos y [MC](#) aún permanecen y son motivos de estudio para la colaboración. Algunas razones potenciales puede ser:

- Descripción geométrica del grosor del plomo en el [ECAL](#) (incluyendo posibles variaciones debidas a la gravedad).
- Modelado erróneo del campo eléctrico en los huecos de [LAr](#).
- Modelado erróneo del efecto de *cross-talk* (intercambio de energía entre las celdas del calorímetro debido a la electrónica).

Para tener en cuenta estas diferencias en las [SSs](#), históricamente, se hicieron correcciones en forma de desplazamientos de cada una de las distribuciones de [MC](#). Estos desplazamientos comprendían los denominados Fudge Factors ([FFs](#)), y se determinaban utilizando una minimización de  $\chi^2$  en la comparación de las distribuciones de las [SSs](#) entre datos y [MC](#) [40, 64]. Aunque las diferencias del valor medio disminuyeron sustancialmente tras estas correcciones, como se observa por ejemplo en el caso de  $f_{\text{side}}$  en la Figura 4.8a, quedaron diferencias residuales pero aún así notables. Se observa en las distribuciones que las principales diferencias que se mantuvieron están relacionadas con la forma de las mismas, por lo que se necesitan correcciones de orden superior. En el siguiente capítulo se presenta una descripción detallada de las correcciones. Además, dado que las [SSs](#) se construyen a partir de depósitos de energía en las celdas del [ECAL](#), otra forma posible de corregir el desacuerdo actual es corregir directamente las energías de las celdas en las simulaciones [MC](#), corrigiendo entonces todas las [SSs](#) en simultáneo. Este nuevo enfoque se describe también en el siguiente capítulo.

# CORRECCIONES DE LAS SHOWER SHAPES (**SSs**)

5

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

---

Billie Jean King

En el capítulo anterior se vio que los **SFs** (cociente entre las eficiencias de los datos y las obtenidas a partir de la simulación **MC**) se desvían de la unidad. Dado que la identificación de fotones se basa en los cortes de las **SSs** de fotones, se vio que las diferencias de hecho aparecen en estas variables. Desde el Run-1, estas se han corregido con lo que se conoce como Fudge Factors (**FFs**), que se han calculado como simples desplazamientos a las distribuciones **MC** y se ha visto que proporcionan muy buenas mejoras de los **SFs**. Sin embargo, como se ha visto antes, siguen habiendo discrepancias entre las distribuciones que hay que abordar para poder contar con una simulación aún mejor. En la la Sección 5.1, se presenta un enfoque más sofisticado basado en un cálculo de orden superior para corregir las **SSs**. Asimismo, en la la Sección 5.2 se estudia y aborda un nuevo enfoque que utiliza directamente las energías de las celdas. Los estudios presentados en este capítulo constituyen uno de los principales temas de trabajo de la presente tesis.

## 5.1. Fudge Factors (**FFs**)

### 5.1.1. Muestras de datos y simulaciones **MC**

Los **FFs** se calculan utilizando el conjunto completo de datos de Run-2, recolectados a una energía de centro de masa de  $\sqrt{s} = 13$  TeV y con una luminosidad integrada correspondiente a  $140 \text{ fb}^{-1}$ . Las muestras simuladas de **RZ** y **SP** se utilizan para este estudio, ya que representan rangos  $p_T$  complementarios. Los eventos de **RZ** se generan con SHERPA 2.2.11 [67], mientras que SHERPA 2.2.1 se utiliza para los eventos de fondo  $Z \rightarrow \ell\ell$ . Respecto a las muestras **SP**, los eventos se generan con PYTHIA 8.186 [68], que incluye eventos  $\gamma + \text{jet}$  de Leading Order (**LO**) procedentes tanto de procesos directos ( $qg \rightarrow q\gamma$  y  $q\bar{q} \rightarrow g\gamma$ ) como de fragmentación de fotones procedentes de eventos Quantum Chromodynamics (**QCD**) dijet.

En ambos casos, el detector [ATLAS](#) se simula utilizando GEANT4 [69] y los eventos [MC](#) se escalan para que sus distribuciones de pileup se asemejen a las de los datos, para cada año del periodo de toma de datos.

### 5.1.2. Cálculo de Fudge Factors (FFs)

El cálculo se realiza por separado para las dos muestras consideradas: [RZ](#) para fotones con  $7 \leq p_T \leq 50$  GeV y [SP](#) para fotones con  $p_T > 50$  GeV, que ya se discutieron en la la Sección 4.2.1. Dado que las distribuciones de las [SSs](#) varían en función de  $p_T$  y  $|\eta|$ , el cálculo se realiza en bines de estas variables:

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

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

Además, como se menciona en la Sección 4.1, hay variables muy sensibles al estado de conversión del fotón, es decir, si los fotones están convertidos o no. Por esta razón, el cálculo se hace por separado para fotones convertidos y no convertidos. En total se corrigen nueve variables con este método:  $E_{\text{ratio}}$ ,  $f_{\text{side}}$ ,  $R_\eta$ ,  $R_\phi$ ,  $R_{\text{had}}$ ,  $R_{\text{had}\,1}$ ,  $w_1$ ,  $w_{\eta\,2}$  y  $w_{\text{s}\,\text{tot}}$ ; ya que son en las que se observan las mayores discrepancias entre los datos y [MC](#).

Para cada [SS](#), se crean histogramas de [MC](#) y datos de 100 bines. La elección del *binneado* se basa en disponer de estadística suficiente en cada bin y también en capturar todas las características de las variables. Después, cada histograma se suaviza utilizando una herramienta del paquete de TMVA [70] denominada Kernel Density Estimator ([KDE](#)). El método [KDE](#) consiste en estimar la forma de una Función de Densidad de Probabilidad ([PDF](#)) mediante la suma sobre eventos suavizados. La [PDF](#)  $p(x)$  de una variable  $x$  es entonces

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

donde  $N$  es el número de eventos,  $K_h(t) = K(t/h)/h$  es la función kernel, y  $h$  es el ancho de banda del kernel. La idea básica es que cada evento se considera como una función Dirac- $\delta$ , que se sustituye por una función Kernel (Gaussiana) y finalmente se suman para formar la [PDF](#) final. El método de suavizado [KDE](#) puede aplicarse de dos formas: no adaptativo o adaptativo, como se ve en la Figura 5.1. En el primer caso, el ancho de banda es constante para toda la muestra  $h_{NA}$ , mientras que en el segundo, se utiliza el valor de [KDE](#) no adaptativo pero que varía en función de  $p(x)$  como

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

El método [KDE](#) adaptativo mejora la forma de la [PDF](#) especialmente en regiones de baja estadística, pero en regiones de alta estadística puede dar lugar a un exceso de suavizado o *oversmoothing*. El grado de suavizado se ajusta multiplicando el ancho de banda  $h$  por lo que se denominan *fine factors*. Estos factores son parámetros definidos por el usuario que se ajustan para permitir que la [PDF](#) conserve las características importantes del histograma original

y también para evitar fluctuaciones estadísticas. Los valores más altos de los factores indican funciones Kernel más amplias y, por lo tanto, la PDF capta menos fluctuaciones estadísticas. En la Figura 5.2 se muestran ejemplos del procedimiento de suavizado aplicado a  $R_{\text{had}}$  para casos en los que los histogramas originales tienen baja y alta estadística.

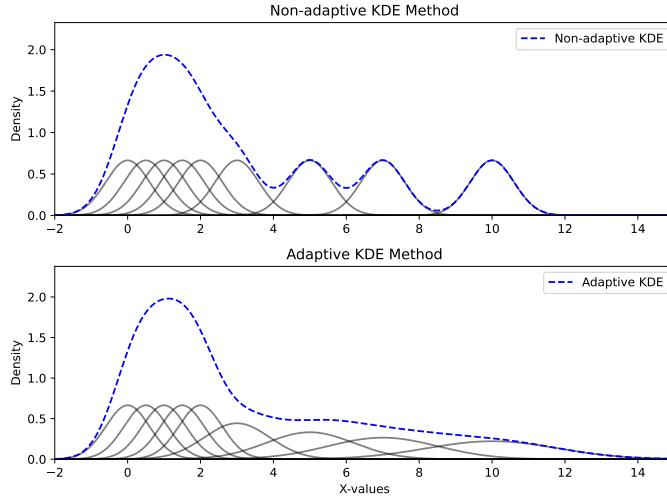


Figura 5.1: Esquema del suavizado no adaptativo y adaptativo del método KDE.

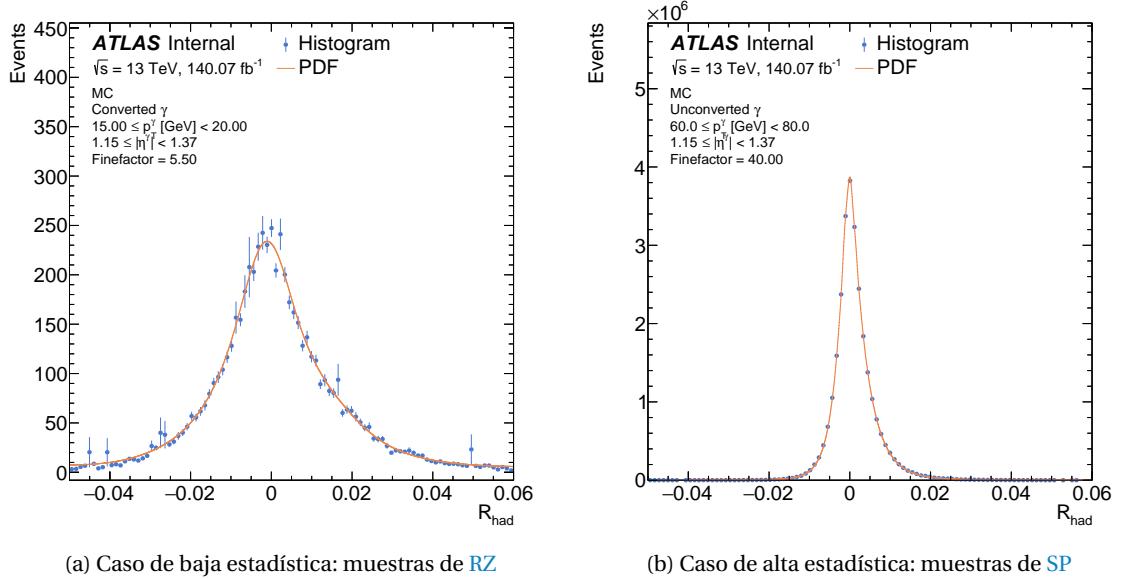


Figura 5.2: Suavizado utilizando el método KDE aplicado a la SS  $R_{\text{had}}$  para fotones en  $0.8 < |\eta| < 1.15$  bajo dos posibles escenarios: baja y alta estadística. El histograma original se muestra con los puntos azules y las correspondientes PDFs con la línea naranja. Además, se muestran los valores de los fine factors usados en cada caso.

Una vez creados las PDFs de los datos y la simulación MC para una dada variable, bin de  $p_T$  y  $|\eta|$ , y tipo de conversión, la PDF de MC se normaliza al de los datos y se calcula un valor  $\chi^2$  entre ambos como [71]:

$$\chi^2 = \sum_{i=1}^N \frac{(w_{\text{MC},i} W_{\text{data}} - w_{\text{data},i} W_{\text{MC}})^2}{s_{\text{MC},i}^2 W_{\text{data}}^2 + s_{\text{data},i}^2 W_{\text{MC}}^2}. \quad (5.1)$$

$N$  es el número de bins de las PDFs,  $w_{\text{MC},i}$  y  $w_{\text{data},i}$  son los números de eventos de MC y datos en cada bin, respectivamente,  $s_{\text{MC},i}$  y  $s_{\text{data},i}$  son los errores del bin y, por último,  $W_{\text{data}}$  y  $W_{\text{MC}}$  son la suma de los pesos en datos y MC, respectivamente.

### Correcciones *shift-only*

Como ha sido mencionado anteriormente, las correcciones a las **SSs** de **MC** han sido realizadas a partir de simples corrimientos de ellas. Estos corrimientos o desplazamientos se denominan, de aquí en adelante, *shift fudge-factors FF*, o simplemente *shifts*. Para ello, se desplaza a la **PDF** de **MC** a la izquierda y a la derecha un bin a la vez. El número inicial de bines que se debe desplazar a la distribución **MC** se calcula mediante la diferencia de los valores medios de las distribuciones de datos y **MC**. A partir de este valor inicial, se consideran shifts de 100 bines a cada lado. Como consecuencia de este procedimiento, la resolución del shift depende directamente del ancho del bin de la **PDFs**, por lo que bines más pequeños conducen a una mejor resolución del shift. Dado que los histogramas, en primer lugar, se construyen con bines relativamente anchos, la **PDFs** puede construirse utilizando bines pequeños de alta precisión para asegurar una alta resolución. Despues de pruebas de convergencia de los **FFs**, se decide construir las **PDFs** con 5000 bines.

Para cada bin que se ha desplazado la distribución, se calcula y se registra el valor  $\chi^2$  antes mencionado. Suponiendo que los errores  $s_{MC,i}$  y  $s_{data,i}$  tienen una distribución gaussiana estándar<sup>1</sup>, se espera que la forma seguida por los valores  $\chi^2$  cerca del mínimo sea aproximadamente parabólica.

Para extraer los **FFs**, se realiza un ajuste a los valores de  $\chi^2$  cercanos al mínimo (5 bines a cada lado del bin mínimo) utilizando una función parabólica y el **FF** de shift se obtiene a partir del mínimo ajustado. Por último, utilizando este valor, se puede corregir a la **SSs** evento a evento como:

$$x = x_{\text{old}} + \text{shift}.$$

donde  $x_{\text{old}}$  y  $x$  representan el valor original y el valor post-corrección de la variable la cual se quiere corregir, respectivamente.

### Correcciones *shift+stretch*

Se observó que incluso después de aplicar correcciones de shift a las **SSs** de **MC**, seguían existiendo diferencias en las formas de las mismas, y en algunos casos éstas pueden ser bastante sustanciales. Una forma de seguir mejorando el acuerdo entre los datos y **MC** es incluir otra corrección que se denomina *stretching*. Las dos correcciones, actuando en conjunto, son denominadas como correcciones *shift+stretch* (o desplazamiento+estiramiento), que pretenden corregir simultáneamente el valor medio y los anchos de las distribuciones de **MC**.

El método de corrección *shift+stretch* empieza por encontrar el máximo de la **PDF** de **MC**. Posteriormente, la **PDF** se estira alrededor del máximo calculando la nueva posición de cada bin por el producto  $\text{stretch} \times (x - \text{stretch point})$ , donde  $x$  es el centro del bin en cuestión. De este modo, el centro de cada bin conserva la distancia inicial al centro de la distribución, multiplicada por el factor de stretch. En el escenario en el que el stretch es  $> 1$ , puede haber casos en

---

<sup>1</sup> Este requisito se cumple siempre que los contenidos de los bines de ambas **PDFs** sean mayores que 10, lo que también se satisface puesto que los histogramas se construyen con bines relativamente amplios.

los que sea lo suficientemente grande como para dar lugar a bins vacíos. El contenido de estos bins vacíos se interpola linealmente a partir de los bins vecinos distintos de cero. Una vez *estirada* la PDF, se desplaza a izquierda y derecha siguiendo el mismo procedimiento que para el caso de shift-only, calculando los valores  $\chi^2$  para cada shift<sub>i</sub> después de aplicar el stretch<sub>j</sub>. Como resultado de este procedimiento, ahora se obtiene una grilla bidimensional de valores de  $\chi^2$  en el plano de shift-stretch. El par shift-stretch se obtiene del centro del bin mínimo, y comprenden ahora los FFs. Las correcciones pueden ser aplicadas a cada SS  $x$ , evento a event, como:

$$x = \text{stretch} \times (x_{\text{old}} - \text{stretch point}) + \text{shift} + \text{stretch point}, \quad (5.2)$$

donde nuevamente  $x_{\text{old}}$  representa el valor de la variable sin corregir.

Un ejemplo de los valores de  $\chi^2$  resultantes para la variable  $f_{\text{side}}$  se muestra en la Figura 5.3a, donde el shift está representado en el eje  $x$  y el stretch en el eje  $y$ . El valor óptimo de shift-stretch en este caso corresponde a shift = 0.03 y stretch = 1.09. En la Figura 5.3b se muestran las PDFs antes y después de aplicar las correcciones, donde se comparan con la PDF de los datos. Como se ve en la figura, hay una gran mejora y las distribuciones coinciden casi a la perfección.

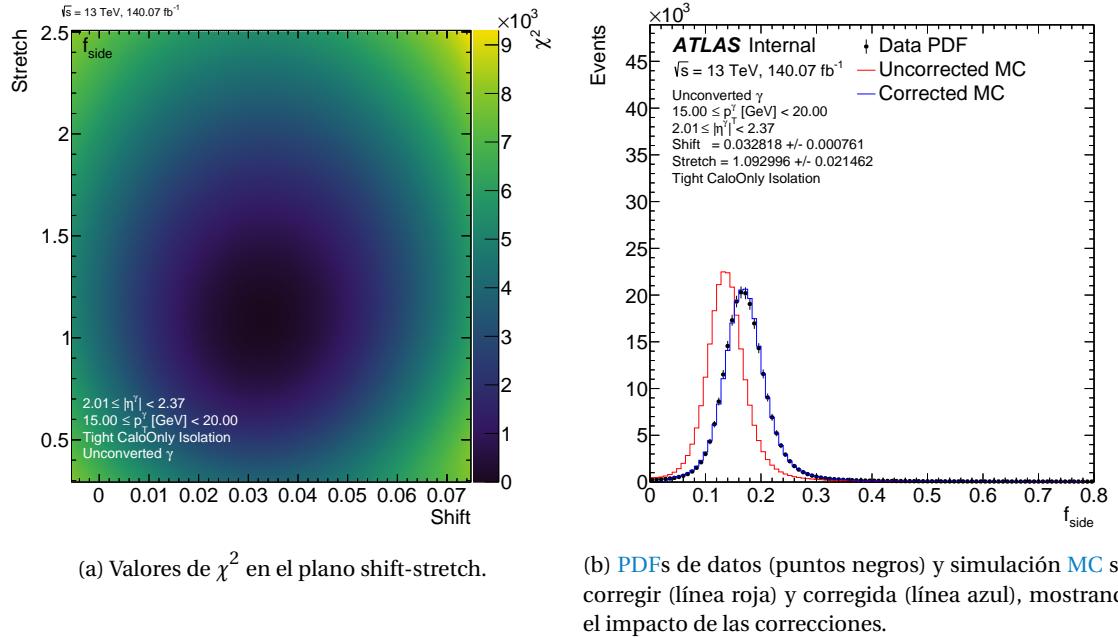


Figura 5.3: Cálculo de los FFs de shift+stretch para  $f_{\text{side}}$  utilizando fotones no convertidos con momento transverso de  $15 < p_T < 20 \text{ GeV}$  y pseudorapidez  $2.01 < |\eta| < 2.37$

### 5.1.3. Cálculo de incertezas

#### Incertezas estadísticas

Para extraer las incertezas estadísticas de los FFs de shift y stretch, se realiza un ajuste al contorno de  $1\sigma$  (nivel de confianza del 68.3 %) sobre los valores  $\chi^2$ . Este contorno representa una

elipse que toma la siguiente forma:

$$\chi^2 = \chi_{\min}^2 + \frac{1}{1-\rho^2} \left[ \left( \frac{x-x_0}{\sigma_x} \right)^2 + \left( \frac{y-y_0}{\sigma_y} \right)^2 - 2\rho \left( \frac{x-x_0}{\sigma_x} \right) \left( \frac{y-y_0}{\sigma_y} \right) \right], \quad (5.3)$$

donde  $\rho$  es el coeficiente de correlación entre ambas variables,  $\sigma_x$  y  $\sigma_y$  las incertezas sobre  $x$  y  $y$ , respectivamente,  $(x_0, y_0)$  es la posición del centro de la elipse, y  $\chi_{\min}^2$  es el valor mínimo de  $\chi^2$  obtenido del histograma bidimensional.

Extrayendo los semiejes mayor y menor de la elipse ajustada, y con el ángulo de inclinación de la misma, las incertezas estadísticas sobre dos variables  $x$  y  $y$  (que en este caso representan el shift y el stretch, respectivamente) son (véase el Apéndice A):

$$\sigma_x = \sqrt{a^2 \cos^2 \theta + b^2 \sin^2 \theta} \quad (5.4)$$

$$\sigma_y = \sqrt{a^2 \sin^2 \theta + b^2 \cos^2 \theta}. \quad (5.5)$$

### Incertezas sistemáticas

Las incertezas sistemáticas se obtienen variando los criterios de preselección, es decir, la identificación y el aislamiento de fotones. El cambio de los diferentes criterios de preselección permite que las SSs varíen dependiendo de la cantidad de contaminación de fondo, y en consecuencia también lo hacen los FFs. Las diferentes selecciones son, para cada muestra:

- *Z Radiativo (RZ):*

- Nominal: Sin criterio de identificación, aislamiento FixedCutTightCaloOnly.
- Identificación loose, sin aislamiento.
- Identificación loose, aislamiento FixedCutTightCaloOnly.
- Sin identificación, aislamiento FixedCutLoose.

- *Single Photon (SP):*

- Nominal: identificación tight, aislamiento FixedCutLoose.
- identificación tight, aislamiento FixedCutTight .

Todas las demás combinaciones (o falta de ellas) de criterios de selección darían como resultado una muestra con estadísticas demasiado bajas o una muy baja pureza.

Los FFs se derivan para cada una de las selecciones anteriores, y se calcula la diferencia entre la nominal y la variada. La diferencia máxima se toma como incerteza sistemática, como el caso más conservativo. Finalmente, las incertezas estadísticas y sistemáticas se suman en cuadratura.

### 5.1.4. Resultados

Debido al hecho de que los FFs se calculan en un amplio rango de  $p_T$  y utilizando dos muestras distintas que abarcan regiones complementarias, los resultados se concatenan en 50 GeV, que coincide con el límite entre ellas.

A continuación, los valores de shift y stretch para distintas SSs serán mostrados. Los valores de shift se normalizan utilizando la desviación estándar de la SS luego de aplicar el FF de stretch, ya que esta cantidad permite comprender cuánto se desplaza cada variable con respecto a su ancho. Además, proporciona una medida única para todas las variables consideradas, ya que cada una de ellas abarca rangos diferentes. No obstante, el ancho de las variables varía según los distintos bines de  $p_T$  y  $|\eta|$ , lo que puede dar lugar a grandes diferencias entre bines vecinos.

En la Figura 5.4, se presentan ejemplos de los FFs resultantes para las variables  $R_\eta$  y  $w_{\eta 2}$  utilizando fotones convertidos. Se puede observar que para ambas variables los FFs dependen de  $p_T$ , especialmente hacia momentos transversos más altos. Este comportamiento también se repite en todas las variables. Inspeccionando los comportamientos y tendencias de los FFs, también es posible recuperar información sobre el mal modelado de las SSs por el MC. Como se mencionó en la Sección 4.3, se observaron anchos y perfiles en  $\eta$  más amplios para los datos en comparación con la simulación. De hecho, esto se puede inferir dado los valores de stretch aumentan hacia valores más altos de  $p_T$ , estirando las simulaciones MC hasta el doble de su ancho inicial. En el caso de  $R_\eta$  ( $w_{\eta 2}$ ) mostrado, la simulación MC sobreestima (subestima) el valor central de la distribución en casi una desviación estándar después de corregir la ancho, lo que significa que las diferencias entre la distribución MC sin corregir con la de los datos son muy grandes.

También es útil visualizar los FFs en un bin de  $p_T$  fijo y en función de  $|\eta|$ , para así determinar qué tan dependientes de  $|\eta|$  son las correcciones. Esto se muestra para  $w_{s \text{ tot}}$  utilizando fotones convertidos con  $50 < p_T < 60$  GeV en la Figura 5.5. Como se puede notar, para  $|\eta| > 1.81$  (los dos últimos bines), los valores de shift normalizados son mayores que los de los bines anteriores en, al menos, un factor 2. Sin embargo, los valores de shift sin normalizar mostrados en la Figura 5.5b no presentan un cambio tan brusco, observándose sólo una pequeña dependencia en  $|\eta|$ . Como consecuencia de este comportamiento, se puede concluir que el cambio brusco observado es debido al cambio en el ancho de la distribución entre los distintos bines de  $|\eta|$ , tal como se había anticipado.

Para validar los FFs obtenidos, las correcciones se aplican a las SSs evento por evento. Las Figuras 5.6 y 5.7 muestran la aplicación de los FFs a algunas de las SS utilizando las muestras RZ y SP, respectivamente, divididas en las regiones barrel y endcap en  $|\eta|$ . En la región barrel, las correcciones mejoran el acuerdo entre datos y MC, pero la mejora no es tan significativa como en la región endcap, donde se observan excelentes acuerdos entre datos y MC. Tomando como ejemplo las variables  $w_1$  y  $w_{s \text{ tot}}$ , se observan grandes diferencias en las formas entre la simulación nominal y los datos, que los métodos shift+stretch consiguen corregir. El mismo comportamiento se observa con las muestras SP, en las que estas variables presentan dos o

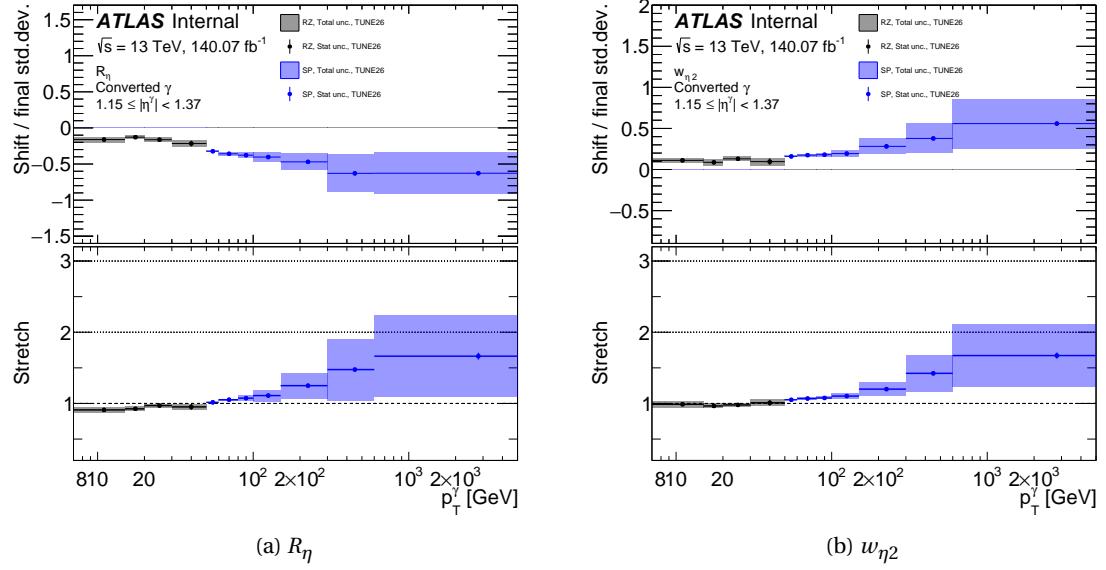


Figura 5.4: Valores de los FFs de shift y stretch para las  $R_\eta$  (izquierda) y  $w_{\eta 2}$  (derecha) para fotones convertidos con  $1.15 < |\eta| < 1.37$ , en función de  $p_T$ . Los resultados obtenidos por las muestras de RZ están representados por el color negro, mientras que los resultados de SP se muestran en azul. Los puntos y las líneas denotan los valores centrales con sus incertezas estadísticas, mientras que las regiones sombreadas representan las incertezas totales. Los valores de shift se muestran en el panel superior, los cuales son normalizados por el ancho de la distribución luego de ser estirada por el stretch, como se ha explicado en el texto. Este último valor se muestra en el panel inferior de las figuras.

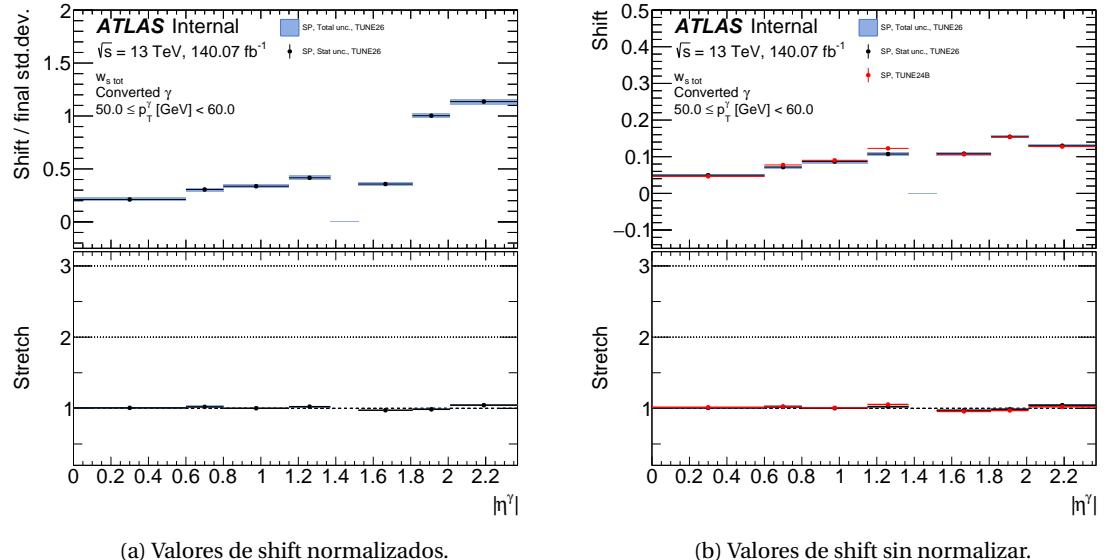


Figura 5.5: Valores de los FFs de shift y stretch para  $w_{s\text{ tot}}$  en función de  $|\eta|$  utilizando fotones convertidos con  $50 < p_T < 60$  GeV de las muestras de SP. La Figura a muestra los valores de shift normalizados, mientras que los no normalizados se encuentran en la Figura b. Los puntos con las líneas de color muestran los valores centrales y las incertezas estadísticas, mientras que las áreas sombreadas representan las incertezas totales en cada bin. Los valores de stretch se muestran en los paneles inferiores de cada figura.

más picos, y que se corrigen correctamente con el método de FFs. En todos los casos mostrados, el MC corregido y los datos son casi indistinguibles, lo que demuestra la importancia de estas correcciones y cómo logran un excelente acuerdo.

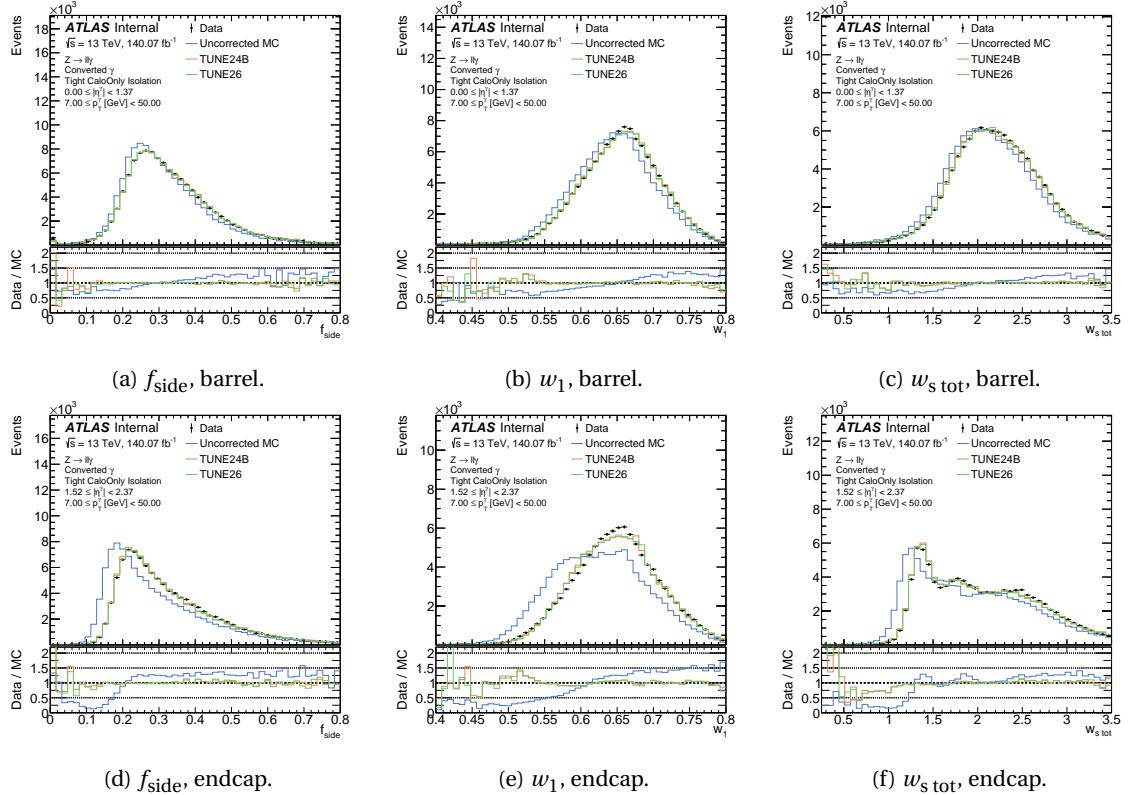


Figura 5.6: Distribuciones de algunas SSs seleccionadas usando las muestras de RZ para fotones convertidos luego de aplicar las correcciones de los FFs en la simulación. Las distribuciones de las SS están separadas para fotones en la región del barrel (fila de arriba) y en la región del endcap (fila de abajo). Los puntos negros representan los datos recolectados por ATLAS, mientras que las simulaciones no corregidas y corregidas están mostradas por las líneas azules y verdes, respectivamente. Los paneles inferiores, en cada figura, muestra el cociente entre el histograma de datos con cada uno de los obtenidos de las simulaciones MC.

## 5.2. Correcciones de energía de las celdas

El diseño y la funcionalidad del ECAL de ATLAS se describió en la Sección 2.2.3, así como el proceso a partir del cual los electrones y los fotones depositan sus energías en el ECAL: creación de pares y radiación bremsstrahlung. Luego, a partir de estas deposiciones de energía en el ECAL se construyen los SSs y se utilizan para la identificación de fotones. Sin embargo, el hecho de que las SSs calculadas mediante las simulación MC y los datos SSs no coincidan, significa que las deposiciones de energía son diferentes entre estos dos, lo que lleva a un desacuerdo a un nivel inferior.

Aunque el método de Fudge Factor (FF) descripto anteriormente condujo a una excelente mejora del acuerdo entre los datos y las distribuciones MC, sigue basándose en la modificación de variables de alto nivel y todas independientemente unas de otras. En cambio, otro enfoque diferente es el de corregir directamente los depósitos de energía de las celdas en la simulación MC. Esto permitiría calcular todas las Shower Shapes (SSs) y cualquier otra variable que utiliza la energía de las celdas ya corregidas.

El enfoque de corregir las energías de las celdas del ECAL se ha desarrollado y probado inicialmente para electrones [72], y posteriormente para fotones [73]. Para el caso de los electrones,

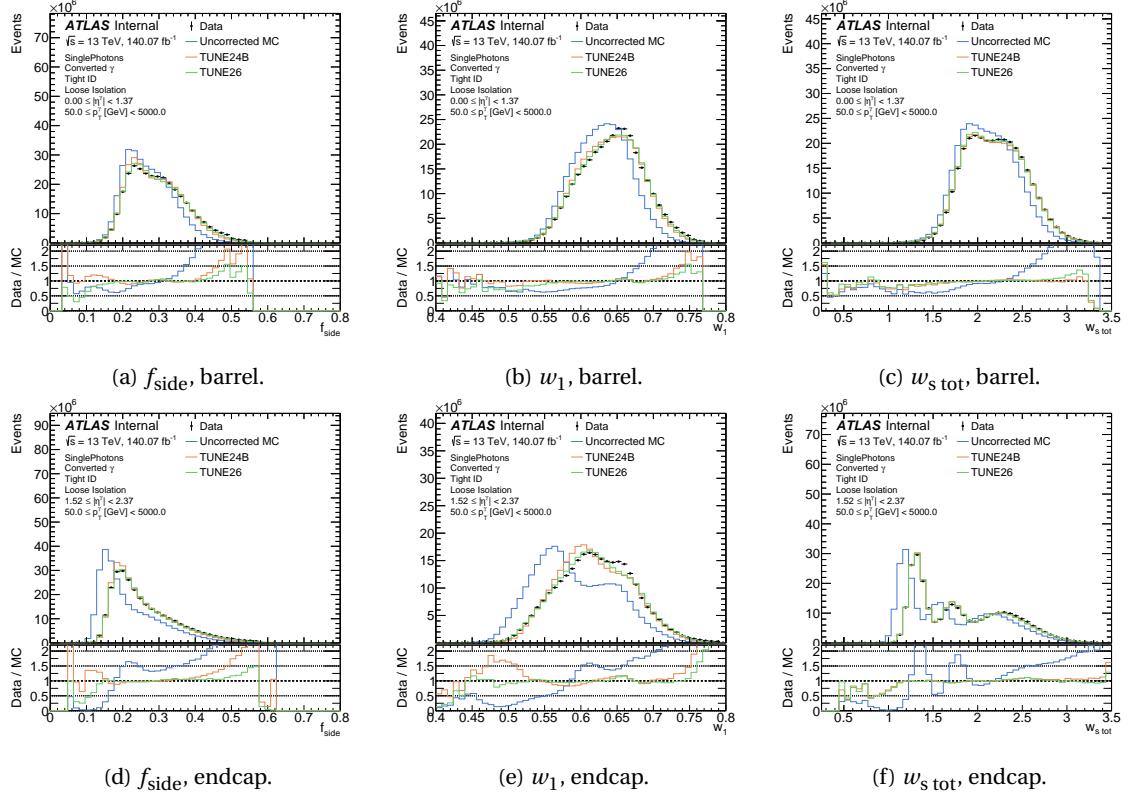


Figura 5.7: Ídem a la Figura 5.6 pero utilizando las muestras de SP.

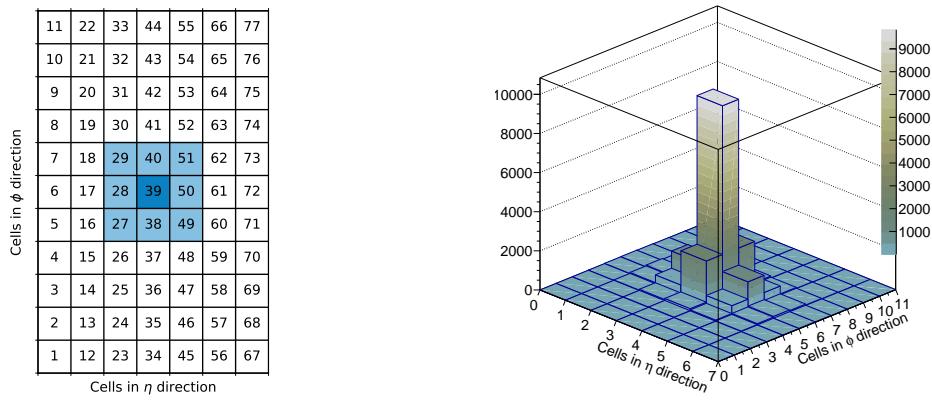
los resultados han sido muy prometedores, ya que se corrigieron sustancialmente las SSs de la segunda capa del calorímetro. Sin embargo, para los fotones, el mismo método que se utilizó para los electrones no funcionó de la forma que se esperaba, ya que sólo permitió corregir las energías en promedio. Otro enfoque para corregir la simulación se basó en hacer coincidir eventos de datos y eventos simulados, estudio que sólo fue probado pseudodatos y resulta técnicamente complicado, pero que condujo a mejores resultados [73].

En la presente sección, se estudia una nueva forma de corregir las energías de celda en MC, utilizando sólo la segunda capa del ECAL, por simplicidad. El método comparte similitudes con el método FF, lo que además facilita su comprensión. En primer lugar, se presenta la selección de eventos especiales utilizada para este estudio. Se discute brevemente el método de corrección de energías utilizado por los primeros estudios basados en electrones y fotones, y luego se presenta en detalle cómo se mejora este método.

### 5.2.1. Selección de eventos

Los estudios presentados en esta sección se llevan a cabo con el mismo conjunto de datos utilizado para el cálculo FF, descripto en la Sección 5.1.1. Sin embargo, en este caso sólo se utilizan las muestras de RZ. Los eventos se seleccionan como se describe en la Sección 4.2.1, utilizando fotones que pasan el criterio de aislamiento loose. Sin embargo, dado que estos estudios se basan en la información de la segunda capa del ECAL, es necesario tener en cuenta una selección especial de las celdas que la conforman.

Cuando un electrón o fotón entra en el calorímetro, su huella en la segunda capa es un grupo visible de celdas que rodean a la más energética y central (también denominada *hottest cell*). En este estudio, se consideran clusters de  $7 \times 11$  celdas en  $\eta \times \phi$ , mostradas en la Figura 5.8a donde también se muestra la disposición utilizada. Aproximadamente, el 90 % de la energía del cluster se reparte entre las 9 celdas centrales, resaltadas en azul en la Figura 5.8a. La energía media normalizada de los datos se muestra en la Figura 5.8b, visualizando cómo se distribuye la energía.



(a) Disposición de las celdas, mostrando para cada una su número. La celda central corresponde a la celda número 39 resaltada en azul oscuro, mientras que las 8 celdas vecinas se muestran resaltadas en celeste.

(b) Energía promedio en cada celda.

Figura 5.8: Disposición de las celdas y distribución de la energía entre las celdas del cluster.

En este trabajo, sólo se consideran los eventos en los que los clusters tienen el total de las 77 celdas. Además, se requiere en los eventos que la celda central sea la más energética.

### 5.2.2. Cálculo de las correcciones

#### Primeros pasos

Todos los eventos que superen la selección mencionada tendrán asociado un cluster, cada uno de los cuales tendrá  $N$  celdas y cada celda tendrá una energía  $E_i$ , con  $i = 1, \dots, N$ . Para cada evento, en primer lugar, se obtiene la energía total del cluster  $E$  sumando las energías de cada una de las celdas  $E_i$ . El método de las correcciones de las SSs mediante la corrección de las energías depositadas en el ECAL hace uso de las energías normalizadas en cada celda,  $e_i = E_i/E$ . Estos valores dan a entender qué proporción de la energía total depositada tiene una celda en particular.

El proceso de corrección comienza entonces calculando el valor medio de las distribuciones  $e_i$  (obtenidas una vez que todas los eventos pasan la selección) para la  $i$ -ésima celda, en la simulación MC y en los datos, y la diferencia entre estos valores dan lugar a la corrección  $\Delta_i$  en

dicha celda:

$$\Delta_i = \overline{\left( \frac{E_i^{\text{data}}}{E^{\text{data}}} \right)} - \overline{\left( \frac{E_i^{\text{MC}}}{E^{\text{MC}}} \right)} = \bar{e}_i^{\text{data}} - \bar{e}_i^{\text{MC}}. \quad (5.6)$$

Los valores  $E^{\text{data/MC}}$  son las energías totales del cluster para los datos y MC, respectivamente.

La energía de la celda  $i$ , se corrige entonces como

$$E_i^{\text{MC-RW}} = E_i^{\text{MC}} + \Delta_i E^{\text{MC}}, \quad (5.7)$$

que se traduce en desplazar la energía normalizada de la celda  $e_i^{\text{MC}}$  en una cantidad  $\Delta_i$ , para que los valores medios de las distribuciones de  $e_i$  de datos y MC coincidan.

También es importante notar que, por definición, estos coeficientes de corrección suman 0 en todo el cluster:

$$\sum_i \Delta_i = \sum_i \overline{\left( \frac{E_i^{\text{data}}}{E^{\text{data}}} \right)} - \sum_i \overline{\left( \frac{E_i^{\text{MC}}}{E^{\text{MC}}} \right)} = \overline{\sum_i \frac{E_i^{\text{data}}}{E^{\text{data}}}} - \overline{\sum_i \frac{E_i^{\text{MC}}}{E^{\text{MC}}}} = 1 - 1 = 0,$$

implicando que el cambio de energía total del cluster se mantiene constante:

$$E^{\text{MC-RW}} \equiv \sum_i E_i^{\text{MC-RW}} = \sum_i E_i^{\text{MC}} + \sum_i \Delta_i E^{\text{MC}} = E^{\text{MC}} + E^{\text{MC}} \sum_i \Delta_i = E^{\text{MC}}.$$

Este hecho es de vital importancia, ya que no se desea cambiar la energía total del cluster en la simulación MC, sino que se desea lograr una redistribución de la energía entre las celdas, de forma tal que cada una se asemeje a la de los datos.

Los coeficientes de corrección resultantes para cada celda en clusters de 77 celdas, se pueden visualizar en la Figura 5.9. Como se puede notar de los valores mostrados, la celda central presenta una corrección negativa, mientras que las 8 vecinas a la central tienen correcciones positivas. Esto se puede traducir a que en la simulación, la celda central suele tener más energía, en promedio, que en los datos, mientras que lo opuesto ocurre en las vecinas. Mediante la aplicación de una corrección negativa (implicando un corrimiento negativo de  $e_i$ ), se remueve energía de la celda central que luego es distribuida en las circundantes.

A partir de las energías de las celdas, se pueden calcular las SSs de la segunda capa del ECAL, las cuales son  $R_\eta$ ,  $R_\phi$  y  $w_{\eta 2}$ :

$$\begin{aligned} R_\eta &= \frac{E_{3 \times 7}}{E_{7 \times 7}} \\ R_\phi &= \frac{E_{3 \times 3}}{E_{7 \times 3}} \\ w_{\eta 2} &= \sqrt{\frac{\sum_i E_i \eta_i^2}{\sum_i E_i} - \left( \frac{\sum_i E_i \eta_i}{\sum_i E_i} \right)^2} \end{aligned}$$

donde  $E_{i \times j}$  es la energía de la celda sumada en una región de  $\eta \times \phi = i \times j$  celdas alrededor de la celda central. Se demostró en los estudios anteriores [73] que este método sólo corrige las formas de las variables en promedio, pero las diferencias en la forma de permanecer. Esto se debe al hecho de que este método sólo corrige los valores medios de energía en las celdas. Sin

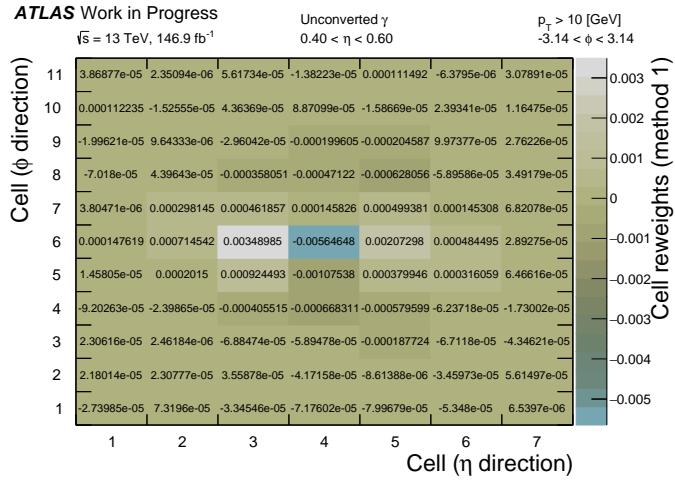


Figura 5.9: Correcciones a las energías de las celdas de la simulación MC utilizando el mismo método diseñado para electrones. [fix luminosity](#)

embargo, estas distribuciones de energía siguen presentando diferencias, especialmente en lo que se refiere a las formas, lo que conduce a una situación muy similar a la observada para los FFs. De este modo, se puede emplear un enfoque muy similar para corregir los valores medios y los anchos de las distribuciones de energía normalizadas.

### Nuevo método de corrección de energías

Este nuevo método pretende corregir tanto el valor medio como la varianza de las distribuciones normalizadas de energía de las celdas, mediante la aplicación de corrimientos (shift) y estiramientos (stretch) de las mismas. De forma similar al enfoque seguido para las SSs utilizando el método de FFs, una primera aproximación a los valores de shift y stretch de las distribuciones de energía consiste en calcular el valor medio y la raíz cuadrática media (RMS) de las mismas en cada celda, respectivamente. Luego, la energía normalizada de la  $i$ -ésima celda se obtiene como:

$$e_i^{\text{MC-RW}} = \underbrace{\frac{\text{RMS}_{e,i}^{\text{data}}}{\text{RMS}_{e,i}^{\text{MC}}} e_i^{\text{MC}}}_{\text{stretch}} + \underbrace{\left( \bar{e}_i^{\text{data}} - \frac{\text{RMS}_{e,i}^{\text{data}}}{\text{RMS}_{e,i}^{\text{MC}}} \bar{e}_i^{\text{MC}} \right)}_{\text{shift}}, \quad (5.8)$$

donde el subíndice  $e$  en los valores de RMS indica que estos se calculan a partir de las distribuciones de energía normalizadas, y el índice  $i$  recorre todas las celdas del cluster. De la expresión anterior se pueden identificar nuevamente un factor de shift, que es una transformación constante de la energía normalizada, y un factor de stretch, lineal en la variable que se requiere corregir.

Dado que la energía normalizada en la celda  $i$  puede calcularse como  $e_i^j = E_i^j / E^j$ , para  $j = \text{MC-RW}, \text{MC}$  y datos, y se requiere tener la misma energía total del cluster luego de aplicar las correcciones ( $E^{\text{MC-RW}} = E^{\text{MC}}$ ), se puede multiplicar la Ec. 5.8 por  $E^{\text{MC-RW}}$  y llegar a una expresión

para  $E_i^{\text{MC-RW}}$ :

$$E_i^{\text{MC-RW}} = \frac{\text{RMS}_{e,i}^{\text{data}}}{\text{RMS}_{e,i}^{\text{MC}}} E_i^{\text{MC}} + \left( \bar{e}_i^{\text{data}} - \frac{\text{RMS}_{e,i}^{\text{data}}}{\text{RMS}_{e,i}^{\text{MC}}} \bar{e}_i^{\text{MC}} \right) E_i^{\text{MC}}. \quad (5.9)$$

Por último, para garantizar que la energía del cluster permanezca constante, las energías de las celdas se reescalan por  $\sum_i E_i^{\text{MC}} / \sum_i E_i^{\text{MC-RW}}$ .

Como el resultado de este procedimiento de corrección de energás involucra una corrección de shift y otra de stretch, se obtienen dos matrices de corrección, y un ejemplo de ellas se presenta en la Figura 5.10. En lo que sigue, este nuevo método se aplica para corregir las energías de las celdas, y se computa de forma inclusiva en  $p_T$  y  $|\eta|$ , sólo separando entre fotones no convertidos y convertidos.

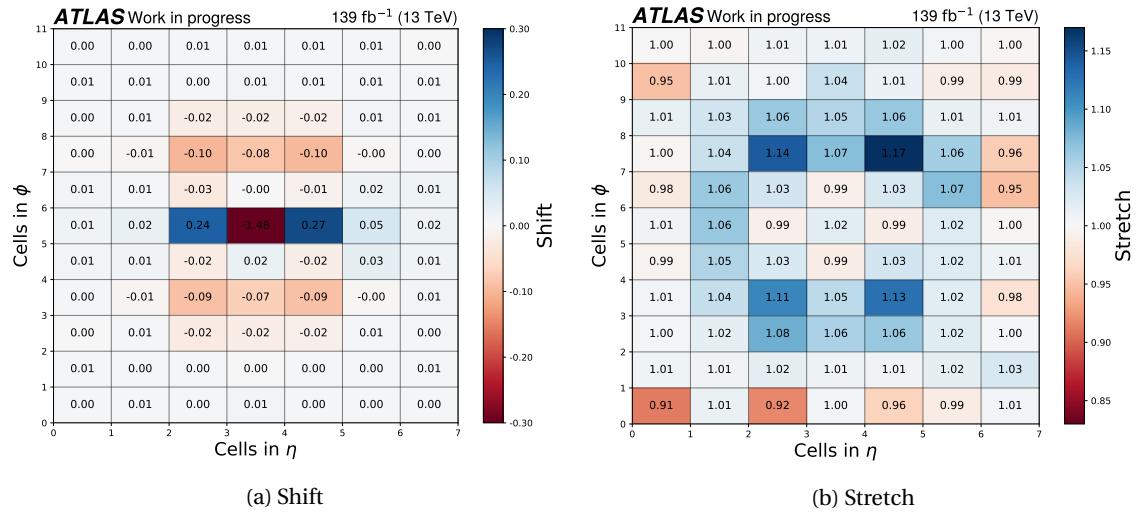


Figura 5.10: Ejemplo de las matrices de corrección de shift (izquierda) y stretch (derecha). Los valores mostrados corresponden al cálculo de las correcciones utilizando fotones no convertidos. Los valores de shift son multiplicados por un factor de 100 para mejorar su visualización.

### 5.2.3. Resultados

La Figura 5.10 muestra las matrices de corrección de shift y stretch obtenidas para fotones no convertidos. Puede observarse que, al igual que en el caso del método anterior, la mayor corrección de shift se realiza en la celda central, donde el shift corresponde a un valor negativo. De la misma forma que en el caso anterior, los shifts de las celdas vecinas en la dirección de  $\eta$  son positivos y grandes, indicando la redistribución de la energía de la celda central en estas dos vecinas. Sin embargo, se puede notar que las segundas celdas vecinas en la dirección de  $\phi$  sufren una gran corrección, quitando energía mediante el shift, pero aumentando también el ancho de la distribución, dado por los estiramientos positivos. Del resto de las celdas del cluster, se nota que no presentan corrimiento significativo, pero presenta un stretch  $< 1$ , indicando que se hacen más angostas, especialmente las celdas de los extremos del cluster.

Utilizando estos factores de corrección para las energías normalizadas de cada celda, en las Figura 5.11 se muestran las distribuciones de energía normalizadas resultantes para las celdas

28, 39 y 50<sup>2</sup>. El nuevo método de corrección consigue grandes mejoras en el acuerdo entre los datos y la simulación. Además, el método logra corregir bien las colas de las distribuciones de todas las celdas, así como los picos de las mismas, lo que puede observarse especialmente en la celda 28.

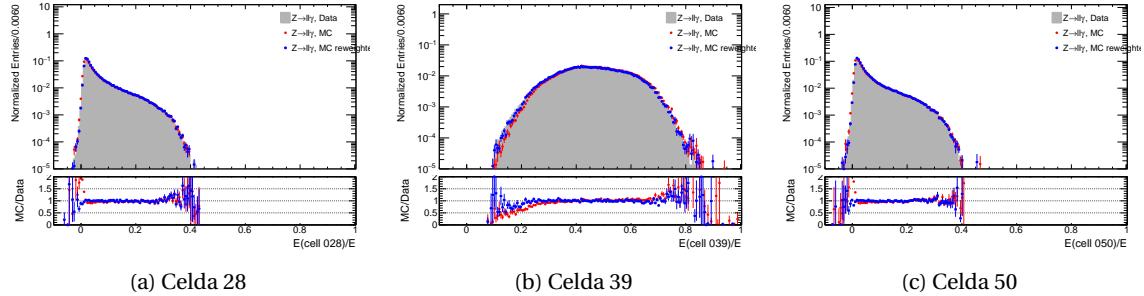


Figura 5.11: Distribuciones de las energías normalizadas de las celdas 28, 39 y 50 de cluster de 77 celdas, para fotones no convertidos. Los puntos azules y rojos corresponden a las distribuciones de la simulación MC con y sin las correcciones, respectivamente, mientras que el histograma gris representa los datos.

Para evaluar el comportamiento del nuevo procedimiento de corrección aplicado a las SSs de la segunda capa ECAL, en la Figura 5.12 se muestra la comparación de los métodos de corrección para las variables  $R_\eta$ ,  $R_\phi$  y  $w_{\eta 2}$ . En los tres casos, se observa una mejora con respecto al MC sin corregir, especialmente para  $R_\phi$  y  $w_{\eta 2}$ . El método de corrección de energía, en el caso de fotones no convertidos, no alcanza el nivel de acuerdo con los datos logrado por el método de FFs, que ha demostrado proporcionar una excelente acuerdo con los datos experimentales. Sin embargo, casi no se observan diferencias entre el método de corrección de energías y el de FFs para fotones convertidos, lo que indica que aún hay margen de mejora en las correcciones.

### 5.3. Conclusiones y trabajo futuro

En el presente capítulo se han estudiado dos métodos para corregir el desacuerdo observado en las Shower Shapes (SSs) entre los datos y la simulación MC.

El método de Fudge Factor (FF) se ha utilizado históricamente en la colaboración, al principio basado únicamente en simples desplazamientos de las distribuciones. A pesar de que las correcciones conducían a buenas mejoras y por tanto a la obtención de mejores SFs, seguían existiendo notables diferencias de forma entre los datos y la simulación. En el contexto de este trabajo, al añadir un término lineal a la transformación de la variable, se logra corregir los anchos de las distribuciones simuladas, lo que conduce a un acuerdo aún mejor con los datos. Este nuevo método de corrección de SSs mediante FFs se denomina método shift+stretch y actualmente se utiliza por toda la colaboración ATLAS.

También se ha desarrollado un método novedoso y de menor nivel de corrección que pretende modificar las energías en las celdas del ECAL. Utilizando las distribuciones de energía en cada celda en clusters alrededor de la celda más energética, es posible corregir todas las SSs en

<sup>2</sup> Como fue mostrado en la Figura 5.8a, la celda número 39 es la central, mientras que las celdas 28 y 50 están a la izquierda y derecha, respectivamente, en la dirección  $\eta$ .

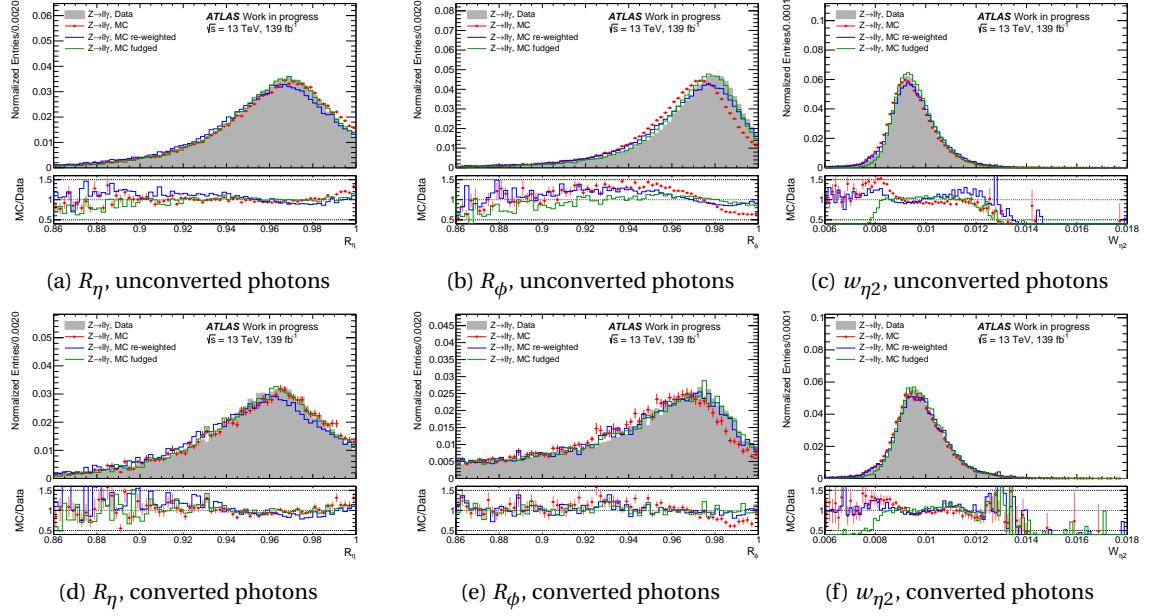


Figura 5.12: Distribuciones de las SSs calculadas en la segunda capa del ECAL para fotones no convertidos (fila superior) y convertidos (fila inferior) con pseudorapidez  $|\eta| < 0.6$ , comparando los diferentes métodos de corrección con los datos. Los datos experimentales están representados por los histogramas grises. La simulación MC sin corregir se muestra con los puntos rojos, la simulación corregida por el método de corrección de energías con la línea azul y la corregida por el método de FFs con la línea verde.

simultáneo. Este método usa la misma estrategia de shift+stretch, pero esta vez aplicado a las distribuciones de energía normalizada en cada celda de la simulación MC, para que coincida con la distribución encontrada en los datos. Aunque el método es nuevo y aún necesita de mejoras, como también extenderlo a las demás capas del ECAL, ha dado resultados prometedores en los que algunas variables se corrigen de la misma manera que con el FFs. El método de corrección de las energías de las celdas muestra un gran potencial en la colaboración, no sólo en el contexto de la identificación de fotones *offline*, sino también a nivel de trigger.

### 5.3.1. Trabajo a futuro

Uno de los enfoques más interesantes y prometedores para corregir el SSs es el método basado en las correcciones de las energías de las celdas. Como se ha mencionado anteriormente, este enfoque podría emplearse en diferentes pasos del proceso de identificación de fotones, como en el nivel de trigger, o de forma *offline* para corregir todos los SSs simultáneamente. Otro uso potencial e importante es utilizar los clusters corregidos para calcular directamente la identificación de fotones, por ejemplo, considerando a los clusters como imágenes y utilizando una red neuronal convolucional (CNN) para realizar la identificación de fotones [73].

Las Shower Shapes (SSs) tienen la gran ventaja de que se pueden interpretar fácilmente en términos físicos. Por esta razón, mantener estas variables sirve para comprender la física subyacente de los procesos. Seguir corrigiendo estas variables es de gran interés y hay varias formas de hacerlo. El método actual de transformar la variable pero utilizando términos de orden superior sigue siendo una tarea difícil, pero aún no explorada. Haciendo uso de las novedosas

técnicas de Machine Learning (ML), es posible obtener factores de corrección para los términos de orden superior en la expansión, corrigiendo además los momentos de orden superior de las distribuciones (asimetría estadística, curtosis, etc.). Otro enfoque interesante es el uso de un re-escaleo Multivariable ([MV](#)), que se exploró en Ref. [74], mostrando resultados muy prometedores.

## **Parte IV**

# **New Physics**

## ANALYSIS MOTIVATION AND STRATEGY

6

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

---

Billie Jean King

yet another template (yat)

# SIGNAL AND BACKGROUND SAMPLES

# 7

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

---

Billie Jean King

yet another template (yat)

## EVENT SELECTION

# 8

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

---

Billie Jean King

yet another template (yat)

## BACKGROUND ESTIMATION

# 9

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

---

Billie Jean King

yet another template (yat)

**BACKGROUND MODELING****10**

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

---

Billie Jean King

yet another template (yat)

## SYSTEMATIC UNCERTAINTIES

11

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

---

Billie Jean King

yet another template (yat)

## STATISTICAL ANALYSIS

# 12

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

---

Billie Jean King

yet another template (yat)

## RESULTS

# 13

*“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

# RELACIONES PARA EL AJUSTE DE UNA ELIPSE

A

Para obtener las incertezas estadísticas en los FFs de shift y stretch se utiliza un ajuste a una elipse, siguiendo los pasos sugeridos en la Ref. [75]. Como salida del ajuste se obtienen un conjunto de parámetros  $\{A, B, C, D, E, F\}$  que parametriza la cónica:

$$F(x, y) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0,$$

con  $B^2 - 4AC < 0$  para el caso de elipses. Las variables  $x, y$  mostradas son generales, pero en el caso de FFs, representan los parámetros de shift y stretch, respectivamente. Este set de parámetros es transformado para obtener la forma canónica de una elipse:

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

de donde se puede extraer el centro de la elipse dado por  $(x_0, y_0)$ , su ángulo de inclinación  $\theta$  y los semiejes mayor y menor,  $a$  and  $b$ , respectivamente. Luego, las incertezas deseadas en  $x$  e  $y$ , junto con su correlación se puede obtener mediante las relaciones (ver Figura A.1):

$$\sigma_x = \sqrt{a^2 \cos^2 \theta + b^2 \sin^2 \theta} \quad (\text{A.1})$$

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

$$\rho = \tan(2\theta) \frac{\sigma_x^2 - \sigma_y^2}{2\sigma_x \sigma_y}. \quad (\text{A.3})$$

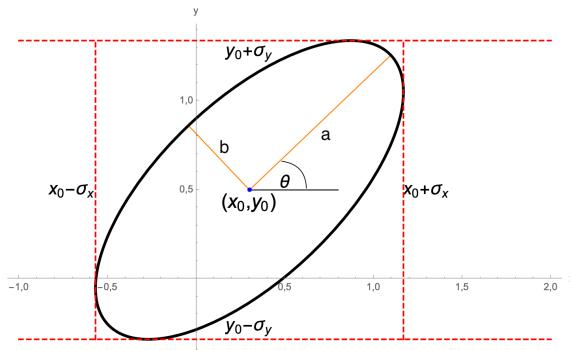


Figura A.1: Parámetros de una elipse.

# GLOSSARY

<b>CERN</b> Organización Europea para la Investigación Nuclear . . . . .	6
<b>ALICE</b> A Large Ion Collider Experiment	
<b>ATLAS</b> A Toroidal LHC ApparatuS . . . . .	ix
<b>CMS</b> Compact Muon Solenoid	
<b>LHC</b> Gran Colisionador de Hadrones . . . . .	ix
<b>LHCb</b> Large Hadron Collider beauty	
<b>LHCf</b> Large Hadron Collider forward	
<b>LS1</b> Long Shut down 1 . . . . .	10
<b>LS2</b> Long Shut down 2 . . . . .	ix
<b>TOTEM</b> TOTal cross section, Elastic scattering and diffraction dissociation Measurement at the <b>LHC</b>	
<b>MoEDAL</b> Monopole & Exotics Detector At the <b>LHC</b>	
<b>PS</b> Proton Synchrotron . . . . .	6
<b>SPS</b> Super Proton Synchrotron . . . . .	6
<b>L1</b> Level-1 . . . . .	16
<b>L1Topo</b> Level-1 Topológico . . . . .	16
<b>LAr</b> Argón Líquido . . . . .	12
<b>ECAL</b> Calorímetro Electromagnético . . . . .	ix
<b>CSC</b> Cathode Strip Chamber . . . . .	15
<b>CTP</b> Central Trigger Processor . . . . .	16
<b>FCAL</b> Calorímetro Forward . . . . .	13
<b>HEC</b> Calorímetro Hadrónico del End-Cap . . . . .	14
<b>HCAL</b> Calorímetro Hadronic . . . . .	ix
<b>HLT</b> High Level Trigger . . . . .	16

<b>IBL</b>	<i>Insertable B-Layer</i>	9
<b>ID</b>	Detector Interno	ix
<b>SCT</b>	Semiconductor de Trazas	9
<b>RPC</b>	<i>Resistive-Plate Chamber</i>	15
<b>TGC</b>	<i>Thin-Gap Chamber</i>	15
<b>TRT</b>	Detector de Radiación de Transición	10
<b>NSW</b>	<i>New Small Wheel</i>	15
<b>MDT</b>	<i>Monitored Drift Tube</i>	15
<b>MS</b>	Espectrómetro de Muones	vi
<b>GRL</b>	<i>Good Runs List</i>	18
<b>LB</b>	Bloques de Luminosidad	17
<b>PSB</b>	Proton Synchrotron Booster	6
<b>BDT</b>	<i>Boosted Decision Tree</i>	31
<b>MC</b>	Monte Carlo	vii
<b>MV</b>	Multivariable	41
<b>KDE</b>	Kernel Density Estimator	x
<b>ROI</b>	Región de interés	16
<b>FF</b>	Fudge Factor	vii
<b>SF</b>	Factor de escala	x
<b>SS</b>	Shower Shape	vii
<b>WP</b>	<i>Working Point</i>	x
<b>PDF</b>	Función de Densidad de Probabilidad	x
<b>EM</b>	electromagnética	35
<b>FSR</b>	Radiación de estado final	x
<b>RZ</b>	$Z$ Radiativo	x
<b>SP</b>	Single Photon	xi
<b>ISR</b>	Radiación de estado inicial	x
<b>LO</b>	Leading Order	41
<b>ME</b>	Matrix Element	26
<b>QCD</b>	Quantum Chromodynamics	46
<b>SM</b>	Modelo Estándar	39
<b>JVT</b>	Jet vertex Tagger	30
<b>JVF</b>	Jet vertex Fraction	30

<b>PV</b>	Vértice Primario . . . . .	22
<b>CB</b>	Muones combinados . . . . .	26
<b>ST</b>	Muones segmentados . . . . .	26
<b>CT</b>	Muones calorimétricos . . . . .	26
<b>ME</b>	Muones extrapolados . . . . .	26
<b>PFlow</b>	Particle Flow . . . . .	vii
<b>FTAG</b>	<i>Flavor Tagging</i> . . . . .	30

## 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, 6, 9, 13, 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] 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>. 9
- [14] 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>. 10
- [15] 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>. 10
- [16] 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>. 10
- [17] 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. 10
- [18] A. Ahmad, Z. Albrechtskirchinger, P. P. Allport, J. Alonso, et al., *The silicon microstrip sensors of the ATLAS semiconductor tracker*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **578** no. 1, (2007) 98–118. 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>. 12

- [22] The ATLAS Collaboration, *ATLAS tile calorimeter: Technical Design Report*, CERN/LHCC96-42(1996) (1996). <https://cds.cern.ch/record/331062>. 14
- [23] *ATLAS muon spectrometer : Technical Design Report*, <https://cds.cern.ch/record/331068>. 15
- [24] 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>. 15
- [25] The ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2010*, The European Physical Journal C **72** no. 1, (2012) 1849. 16
- [26] A. ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2015*, [arXiv:1611.09661](https://arxiv.org/abs/1611.09661). <http://arxiv.org/abs/1611.09661> [http://dx.doi.org/10.1140/epjc/s10052-017-4852-3](https://dx.doi.org/10.1140/epjc/s10052-017-4852-3). 16
- [27] The ATLAS Collaboration, *Operation of the ATLAS trigger system in Run 2*, Journal of Instrumentation **15** no. 10, (2020), [arXiv:2007.12539v2](https://arxiv.org/abs/2007.12539v2). 16
- [28] 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
- [29] 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
- [30] 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
- [31] 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
- [32] 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
- [33] 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
- [34] 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.

- [35] 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, 21
- [36] 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>. 21
- [37] 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>. 21
- [38] 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. 22
- [39] *Vertex Reconstruction Performance of the ATLAS Detector at  $\sqrt{s} = 13$  TeV*, tech. rep., CERN, Geneva, 2015. <https://cds.cern.ch/record/2037717>. 22
- [40] 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, 22, 23, 42, 45
- [41] 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>. 22, 28
- [42] 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. 22, 28
- [43] 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). 22
- [44] 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). 22
- [45] 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>. 22
- [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. 24, 41
- [47] 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). x, 24, 44

- [48] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, Physics Letters B **659** no. 1, (2008) 119–126.  
<https://www.sciencedirect.com/science/article/pii/S0370269307011094>. 25
- [49] 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. 26
- [50] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- $k_t$  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, 27, 28
- [51] The ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, The European Physical Journal C **77** no. 7, (2017) 466. 28, 29
- [52] 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. 30
- [53] *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>. 31
- [54] ATLAS Collaboration, *Measurement of the b-tag Efficiency in a Sample of Jets Containing Muons with  $5\text{ fb}^{-1}$  of Data from the ATLAS Detector*, tech. rep., CERN, Geneva, 2012.  
<https://cds.cern.ch/record/1435197>. 31
- [55] A. Lavrenov, *MV2-algorithm's clones*, CoRR **cs.CR/0208003** (2002).  
<https://arxiv.org/abs/cs/0208003>. 31
- [56] M. Lanfermann, *Deep Learning in Flavour Tagging at the ATLAS experiment*, PoS **EPS-HEP2017** (2018) 764. 31
- [57] 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. 31
- [58] ATLAS Collaboration, *Graph Neural Network Jet Flavour Tagging with the ATLAS Detector*, tech. rep., CERN, Geneva, 2022. <https://cds.cern.ch/record/2811135>. 31
- [59] 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>. 31
- [60] 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.  
<http://arxiv.org/abs/1706.03762>. 31
- [61] Z. Wu, “b-jets calibration in atlas.” ATLAS-CMS Flavour Tagging Workshop, 2024. x, 32
- [62] I. Wingerter-Seez, *Particle Physics Instrumentation*, arXiv:1804.11246.  
<https://cds.cern.ch/record/2315747>. x, 38

- [63] ATLAS Collaboration, *Search for short- and long-lived axion-like particles in  $H \rightarrow aa \rightarrow 4\gamma$  decays with the ATLAS experiment at the LHC*, *The European Physical Journal C* **84** no. 7, (2024) 742. <https://doi.org/10.1140/epjc/s10052-024-12979-0>. 39
- [64] The ATLAS Collaboration, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016*, *The European Physical Journal C* **79** no. 3, (2019) 205. x, 42, 43, 45
- [65] L. Devroye, *Non-Uniform Random Variate Generation*. Springer New York, NY, 1986. 42
- [66] L. Lyons, D. Gibaut, and P. Clifford, *How to Combine Correlated Estimates of a Single Physical Quantity*, *Nucl. Instrum. Meth.* **A270** (1988) 110. 44
- [67] E. Bothmann, G. S. Chahal, S. Höche, J. Krause, F. Krauss, S. Kuttimalai, S. Liebschner, D. Napoletano, M. Schönher, H. Schulz, S. Schumann, and F. Siegert, *Event generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) 034. <https://scipost.org/10.21468/SciPostPhys.7.3.034>. 46
- [68] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, *Computer Physics Communications* **178** no. 11, (2008) 852–867. <https://www.sciencedirect.com/science/article/pii/S0010465508000441>. 46
- [69] S. Agostinelli and J. Allison and K. Amako and J. Apostolakis and H. Araujo and others, *Geant4—a simulation toolkit*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** no. 3, (2003) 250–303. <https://www.sciencedirect.com/science/article/pii/S0168900203013688>. 47
- [70] 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). 47
- [71] N. D. Gagunashvili, *Comparison of weighted and unweighted histograms*, 2006. <https://arxiv.org/abs/physics/0605123>. 48
- [72] M. Khandoga, *Calibration of electron shower shapes, hadronic recoil reconstruction using deep learning algorithms and the measurement of W boson transverse momentum distribution with the ATLAS detector*. PhD thesis, U. Paris-Saclay, 2020. <https://cds.cern.ch/record/2752635>. 54
- [73] M. Belfkir, *Search for Higgs pair production at LHC collider (CERN): The first measurement for Higgs potential and search for new physics*. PhD thesis, Savoie Mont Blanc University, 2021. <https://cds.cern.ch/record/2792268>. 54, 55, 57, 61
- [74] J. L. Spah, *Data-driven corrections to shower shape variables for photon identification at the ATLAS experiment*, 2021. <https://kroeninger-group.physik.tu-dortmund.de/en/research/particle-physics/theses/#c135884>. 62

- [75] R. H. oy and J. Flusser, *Numerically Stable Direct Least Squares Fitting of Ellipses*, 1998.  
<https://api.semanticscholar.org/CorpusID:15772208.73>