

PAPER

## Ring-like calorimeter information for energy calibration in electron trigger at a highly segmented detector using gradient boosted decision trees

To cite this article: Eduardo F. Simas Filho *et al* 2025 *JINST* **20** P06051

### You may also like

- [A water calorimeter for on-site absorbed dose to water calibrations in  \$^{60}\text{Co}\$  and MV-photon beams including MRI incorporated treatment equipment](#)  
Leon de Prez, Jacco de Pooter, Bartel Jansen *et al.*

- [Anomalous magnetic moment of the muon](#)  
I B Logashenko and S I Eidelman

- [The scientific potential and technological challenges of the High-Luminosity Large Hadron Collider program](#)  
Oliver Brüning, Heather Gray, Katja Klein *et al.*

View the [article online](#) for updates and enhancements.



**UNITED THROUGH SCIENCE & TECHNOLOGY**

**ECS** The Electrochemical Society  
Advancing solid state & electrochemical science & technology

**248th ECS Meeting**  
**Chicago, IL**  
October 12-16, 2025  
*Hilton Chicago*

**Science + Technology + YOU!**

**REGISTER NOW**

Register by September 22 to save \$\$

This content was downloaded from IP address 163.116.228.136 on 13/08/2025 at 13:08

RECEIVED: April 10, 2025

REVISED: May 30, 2025

ACCEPTED: June 9, 2025

PUBLISHED: June 26, 2025

# Ring-like calorimeter information for energy calibration in electron trigger at a highly segmented detector using gradient boosted decision trees

Eduardo F. Simas Filho<sup>ID, a,\*</sup> Paulo C. M. A. Farias<sup>ID, a</sup> Edmar E. P. Souza<sup>ID, a</sup>  
Juan L. Marin<sup>ID, b</sup> José M. Seixas<sup>ID, c</sup> Augusto S. Cerqueira<sup>ID, d</sup> and Bertrand Laforge<sup>ID, e</sup>

<sup>a</sup>Electrical and Computer Engineering Department, Federal University of Bahia,  
Rua Aristides Novis, 2, Salvador, Brazil

<sup>b</sup>Federal Institute of Science, Education and Technology of Bahia,  
Avenida Sérgio Vieira Melo, 3150, Vitória da Conquista, Brazil

<sup>c</sup>Federal University of Rio de Janeiro,  
Av Athos da Silveira Ramos, 149, Rio de Janeiro, Brazil

<sup>d</sup>Federal University of Juiz de Fora,  
Rua José Lourenço Kelmer, Juiz de Fora, Brazil

<sup>e</sup>LPNHE, Sorbonne Université,  
4 place Jussieu, Paris, France

E-mail: [eduardo.simas@ufba.br](mailto:eduardo.simas@ufba.br)

**ABSTRACT.** Accurate energy estimation is important for online event selection (trigger) in high-energy physics experiments. In the context of hadron colliders, electrons and positrons play an important role in making precise measurements of different Standard Model predictions or in searching for new physics. At high energy, their detection and characterization rely mainly on calorimeter information (to measure their energies), while the tracking system provides their direction and the collision vertex position. Modern state-of-the-art calorimeters provide high-granularity measurements that resolve the electromagnetic shower development process. This information can be used, on a single shower basis, to correct the stochastic process that drives the electron energy deposition inside the calorimeters but also to mitigate the possible energy loss in the detector dead material. In the context of the ATLAS experiment at the Large Hadron Collider at CERN, where multiple proton-proton collisions produce additional parasitic particles (pileup) that contaminates the electromagnetic shower and lower the electron energy resolution, this correction is still possible but needs to be tuned to exploit the pileup measurement provided every minute by the experiment. This work proposes to use calorimeter information formatted in ring-like features combined with a machine learning regression method based on gradient-boosted decision trees to obtain a calibration factor for electron energy estimation. The proposed calibration method is built using the Lorenzetti Showers framework, a full Geant4 simulation

\*Corresponding author.

of calorimeters in a detector geometry that is representative of the ATLAS experiment at CERN in the present harsh pileup conditions of the LHC. The results indicate a considerable improvement in trigger background rejection efficiency ( $\approx 3.1$  times) even in pileup conditions with respect to the algorithm used during the LHC Run 2. In this study, the ring-like features seem to produce more relevant information for electron energy calibration when compared to traditional shower variables, reducing by 10 percent points the fluctuations after calibration for specific transverse energy ranges.

**KEYWORDS:** Digital signal processing (DSP); Trigger concepts and systems (hardware and software); Data processing methods; Pattern recognition, cluster finding, calibration and fitting methods

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Problem statement and used dataset</b>	<b>3</b>
2.1	The electron energy calibration problem	3
2.2	The used dataset — Lorenzetti showers simulator	5
<b>3</b>	<b>Calorimeter rings as relevant features for electron energy calibration</b>	<b>7</b>
<b>4</b>	<b>The proposed energy calibration method</b>	<b>11</b>
4.1	Boosted decision trees	12
4.2	Evaluating the calibration effects on energy resolution	13
4.3	Evaluating the calibration effects on triggering	13
<b>5</b>	<b>Calibration results</b>	<b>15</b>
5.1	Energy estimation accuracy	15
5.2	Calorimeter resolution estimation	18
5.3	Trigger performance improvement	18
<b>6</b>	<b>Conclusion</b>	<b>21</b>
<b>A</b>	<b>Dataset and code sharing</b>	<b>22</b>

---

## 1 Introduction

In experimental particle physics, accelerators produce collisions between beams of particles that allow the production and measurement of the fundamental characteristics of matter [1, 2]. Unfortunately, the probabilistic nature of quantum physics makes it highly likely that already well-known proton-proton collision processes will happen very often (background noise) while possibly unknown new physics phenomena are barely produced. This situation requires an online filtering system (trigger) providing an efficient selection of the rare events of interest while rejecting with a very high efficiency background events.

Among the high-energy physics (HEP) experiments currently in operation, the ones associated with the Large Hadron Collider [3] (LHC) at CERN are dealing with stringent operational conditions, as the luminosity is increasing to unprecedented levels [4, 5]. As instrumental limitations (e.g., nonlinearities, electronic noise, etc.) and the pileup effect may contribute to decreasing particle characterization accuracy, it requires regular improvement in the trigger information processing chain to maintain high efficiency in signal detection, while rejecting the background events as much as possible.

The energy deposition profiles measured in the calorimeters are fundamental for triggering as they provide discriminant information for particle characterization [6]. Calorimeters produce a fast response and are usually segmented along the direction of incoming particles to form electromagnetic and hadronic sections. Such a multilayered structure reveals the longitudinal shower development

difference between electromagnetic and hadronic showers. The former is dominated by high cross-section phenomena, namely bremsstrahlung and pair creation mechanism sensitive to the electron clouds of atoms in the matter, and the latter by (much lower probability) nuclear reactions of hadrons with the components of detector material nuclei. Calorimeters are designed to measure the energy of particles in a destructive way (by total absorption). These particles lose energy as they interact with the detector material and generate a cascade (shower) of new, less energetic particles.

The trigger systems in HEP are usually divided into sequential levels. The first level receives the full event rate provided by the particle collisions at the detector interaction point, often requiring operation on dedicated hardware. Higher trigger levels are commonly designed in software and may run on distributed computing systems. The trigger system is sensitive to energy measurement errors because it relies on calorimetry information to detect several final-state particles (e.g., electrons, photons, taus, jets, etc.). Additionally, the pileup affects the energy deposition profile and the total amount of transverse momentum of the physical objects to be selected, increasing the selection of low transverse energy background events.

The calorimeter response depends on the type of particle (electrons, photons, or jets of hadrons) that it receives, so that calibration constants need to be determined for each category of objects, leading to possible systematic errors when the identification procedure fails.

Calorimeter response also depends on the energy of the particles, and thus, the calibration procedure must consider this information. Electronic noise and the pileup effect may also degrade the energy estimation performance. Taking into account all these contributions to systematic effects, the calorimeter calibration procedure is usually divided into some complementary procedures as presented in [7].

Unfortunately, as a considerable processing time is required to execute the various correction stages, these complex methods are not suitable for online trigger operation and are more often available in offline reconstruction [8–10]. Calorimeter calibration at the trigger level is usually reduced to a minimum setup, ensuring sub-optimal energy estimation accuracy. However, a more advanced trigger-level energy calibration could help optimize particle detection efficiency.

Calorimeter specialist knowledge has been used to increase trigger performance in modern particle collider experiments. Beyond traditional shower variables often used to feed linear cut-based hypothesis testing [11, 12], different calorimeter-based discriminating features estimated from the energy deposition profile have been proposed. In [13], an image-processing approach was proposed to characterize calorimeter signals. An effective shower ratio parameter was proposed in [14] to optimize electron detection in SPACAL. Zerkine moments and principal component analysis were applied to scintillating-fiber calorimeter information of the RDI Collaboration in [15]. Moving Window Deconvolution (MWD) and Moving Average (MA) filters were applied in [16] for the PANDA experiment. Another successful example is the ring-like features used for triggering in combination with machine learning methods in the ATLAS experiment [17, 18].

This work proposes a fast and accurate machine-learning energy calibration system for a HEP multi-level electron trigger system based on calorimetry. For this, a combination of a gradient-boosted decision tree (GBDT) ensemble and ring-like feature extraction of calorimeter information is proposed. Simulated samples from a general-purpose segmented calorimeter design were generated using the Lorenzetti Showers Framework [19] to validate the proposed method. The pileup of events was also considered in the simulations. The main contributions of this paper are (i) the evaluation of the relevance of

the ring-like features for electron energy calibration, (ii) the proposal of an accurate method for electron energy calibration for triggering purposes in HEP collider experiments, (iii) the evaluation of the benefits of the proposed calibration procedure for trigger chain optimization and energy estimation accuracy.

One of the main purposes of this work was to evaluate the relevance of a new set of features (ring-like calorimeter information) to the energy calibration problem instead of the traditional calorimeter shower descriptors. In this sense, the comparisons were performed in terms of evaluating the pros and cons of the ring-like features in a machine learning algorithm (GBDT) traditionally used in HEP for calibration. Additionally, the application in online trigger systems requires computationally efficient models, such as BDT, to comply with the severe constraints in the time response. In further work under development, we are investigating the replacement of the BDT with deep neural networks, considering aspects such as the calibration efficiency and the time response.

This document is organized as follows. Section 2 presents the energy calibration problem along with the used dataset. Section 3 presents the ring-like features extraction method and its relevance to the energy calibration problem. The proposed calibration method is described in section 4, together with the performance evaluation procedures. The results are detailed in section 5, considering the effects of the proposed calibration method to improve energy resolution and optimize the trigger performance. Conclusions are derived in section 6. In appendix A, there are instructions on how to access the Lorenzetti simulation framework and the used dataset.

## 2 Problem statement and used dataset

Leptons are key signals of electroweak interactions, so the collection of events with such particles in the final state is important to build powerful tests of the SM predictions or search for new physics. Energy calibration of the electromagnetic calorimeter is a primary ingredient in building efficient trigger algorithms to collect events with electrons in the final state. Calibration is of course also very important for offline analysis to use the potential of mass high resolution channels such as  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ^* \rightarrow 4l$  to perform precise measurements of the Higgs boson properties [20, 21] or to measure the W boson mass. This section describes the electron energy calibration problem together with the used dataset.

### 2.1 The electron energy calibration problem

Calibrating a calorimeter system often requires the determination of conversion and correction factors [7]. The former transforms the electronic readings into an energy scale, and the latter corrects the measurements for specific differences between the Monte Carlo (MC) simulation and real detector responses. Considering this, a calorimeter calibration procedure may comprise several steps such as electronic signal to energy conversion, inter-layer calibration, nonlinearities correction, Monte Carlo simulation to experimental data mapping, and specific adjustments for a given type of particle (electron, photon, pions,  $J/\psi$ , jets of particles, etc.). These latest corrections providing calibration adjustments for specific types of particles are often extracted from statistical learning methods on MC simulations [8, 22].

When dealing with the energy readings for each calorimeter sensor cell, in a simplified way, it can be considered that the calorimeter cell ( $k$ ) estimated energy is obtained from eq. (2.1):

$$e_k = \alpha_{\mu A \rightarrow \text{MeV}} \cdot \alpha_{\text{DAC} \rightarrow \mu A} \cdot \sum_{j=1}^N a_j (s_j - p) \quad (2.1)$$

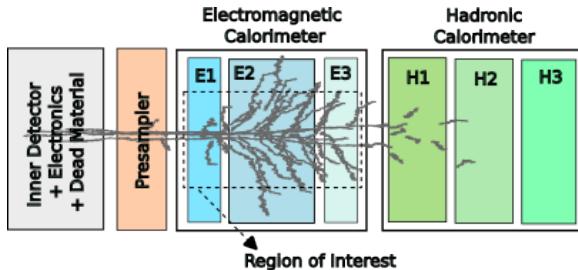
where:

- $\alpha_{\mu A \rightarrow \text{MeV}}$  and  $\alpha_{\rightarrow \mu A}$  are conversion factors. The first one is usually estimated with simulations and test beam results. The last one is determined from experimental calibration procedures [8];
- $a_j$  and  $N$  are the weights and order of the matched filter used to measure pulse characteristics in the calorimeter [8];
- $s_j$  are the samples ADC counts measured in the calorimeter at time  $t_j$  ( $j$  corresponds to a sample number, which, for example, at the LHC are taken every 25 ns);
- $p$  is the pedestal (average level expected in the calorimeter sensors).

Considering an event with an electromagnetic particle candidate, the total particle energy can be estimated by summing the calorimeter cells' readings ( $e_k$ ) within a given 3D region of interest (RoI) defined in the electromagnetic calorimeter section:

$$\hat{E} = \sum_{k \in \text{RoI}} e_k. \quad (2.2)$$

As illustrated in figure 1, during particle interaction with the detector components, energy losses can occur in some ways: (i) laterally beyond the RoI around a lepton candidate provided to the trigger system, (ii) longitudinally (in the hadronic layers — H1, H2 and H3), and (iii) due to the particle's interaction with the detector's material before reaching the calorimeter (upstream).



**Figure 1.** Typical electron interaction with the experiments subdetectors. The trigger Region of Interest (RoI) for an electron candidate is marked in a dashed rectangle. The possible energy deposition profile is drawn in dark gray lines showing that energy might be lost outside the RoI.

Additionally, limitations of the high-energy particle experiments instrumentation systems, such as electronic noise, adjacent sensors crosstalk, and the pileup of different events in the same calorimeter region, may degrade the energy estimation accuracy achieved from the matched filter approach (as presented in eq. (2.1)).

In this context, the calibration procedure tries to mitigate these effects. For this, the ratio between the truth ( $E$ ) and the estimated ( $\hat{E}$ ) energies is used to train different correction algorithms:

$$\alpha = \frac{E}{\hat{E}} \quad (2.3)$$

The response of a calorimeter among its different cells has different characteristics depending on the energy of the particle  $E$ , as the deposition profile changes as a function of  $E$ . There is also

a variable behavior in terms of angular distance of the detected particle with respect to the beam direction, measured from the pseudo-rapidity  $\eta$  ( $|\eta| = 0$  indicates a position at  $90^\circ$  relative to the beam and  $|\eta| \rightarrow \infty$  indicates the beam direction, i.e., making an angle with the beam that  $\rightarrow 0^\circ$ ). The  $\eta$  dependency occurs mainly due to the sensor granularity variation for different detector regions.

## 2.2 The used dataset — Lorenzetti showers simulator

This work used the default segmented calorimeter structure of the Lorenzetti Showers framework [19], which mimics the ATLAS experiment granularity [23], to generate simulated events for the calibration studies. Such a segmented sampling calorimeter comprises six layers, three electromagnetic (EM1, EM2, and EM3) and three hadronic (HD1, HD2, and HD3), preceded by a pre-sampler (PS). The electromagnetic layers comprise alternate volumes of liquid argon (LAr) as sensitive material and lead (Pb) as absorber material. In turn, the hadronic layers are considered to be made of plastic scintillator and iron (Fe) as sensitive and absorber materials, respectively. Table 1 illustrates the cell granularity in  $\Delta\eta \times \Delta\phi$  for each calorimeter layer<sup>1</sup>. An illustration of the used calorimeter design is presented in figure 2, in which it is possible to see (in a longitudinal section view) the electromagnetic and hadronic sections. It is interesting to note that the used calorimeter is uniform in  $\phi$  and symmetric with respect to  $\eta = 0$ .

**Table 1.** Main characteristics of the used calorimeter system for the considered region ( $|\eta| < 1.3$ ).

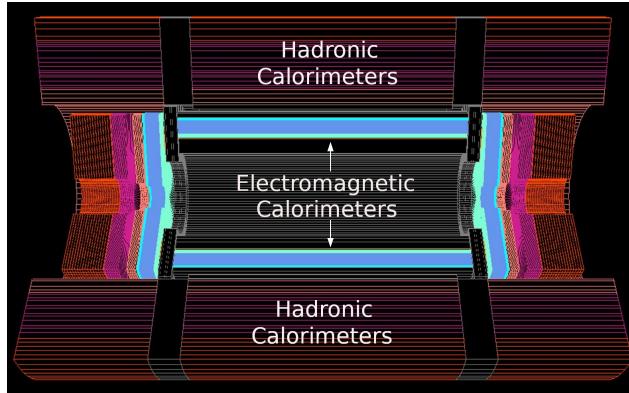
Layer	Granularity ( $\Delta\eta \times \Delta\phi$ )	Material
PS	$0.025 \times 0.1$	LAr+Pb
EM1	$0.003 \times 0.1$	LAr+Pb
EM2	$0.025 \times 0.025$	LAr+Pb
EM3	$0.05 \times 0.1$	LAr+Pb
HD1	$0.1 \times 0.1$	Plastic Scintillator+Fe
HD2	$0.1 \times 0.1$	Plastic Scintillator+Fe
HD2	$0.2 \times 0.1$	Plastic Scintillator+Fe

For the studies performed in this work, approximately 500000 simulated single electron samples ( $e^-$  and  $e^+$ ) were simulated with a transverse energy ( $E_t$ )<sup>2</sup> spanning from 10 to 500 GeV and pseudo-rapidity ( $\eta$ ) from  $-1.3$  to  $1.3$  (covering the barrel region from the considered calorimeter). Figure 3 illustrates the distribution in  $E_t$  and  $\eta$  of the used samples, showing an approximate uniform distribution in  $\eta$  and an energy profile more concentrated between 10 and 180 GeV.

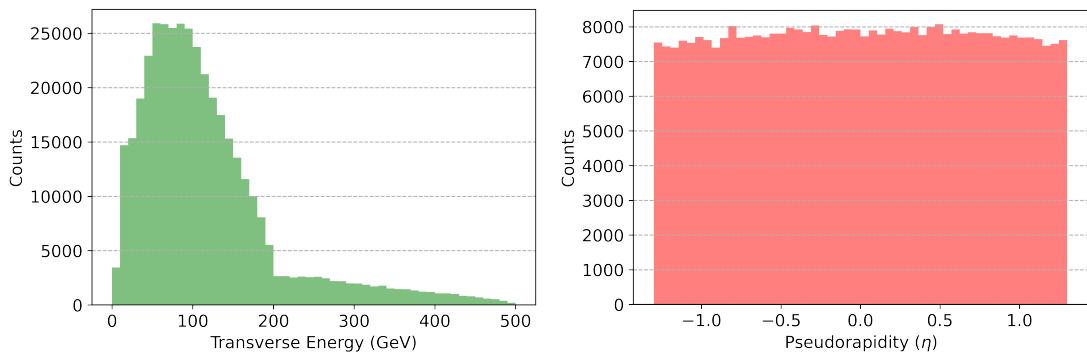
This work focused on the use of simulated data from a general-purpose segmented calorimeter inspired by the ATLAS experiment. The aim is to validate the proposed solution in a controlled case prior to using it for experimental purposes. In this way, it is also possible in the future to easily port the calibration method for different experiments.

<sup>1</sup>We consider a coordinating system comprising: the azimuthal angle of the particle direction in the plane orthogonal to the beam axis ( $\phi$ ) and the pseudorapidity ( $\eta = -\ln(\tan \frac{\theta}{2})$ ), where  $\theta$  is the polar angle between the particle propagation direction and the beam axis.

<sup>2</sup>In experimental HEP,  $E_t = \frac{E}{\cosh \eta}$  is commonly used for energy estimation as it represents the energy measured in a transverse plane.



**Figure 2.** Illustration of the used calorimeter structure showing the electromagnetic and hadronic sections.



**Figure 3.** Histograms of  $E_t$  (left) and  $\eta$  (right) distributions of the simulated single electron samples used in this work. The distribution is also flat in  $\phi$  (not shown).

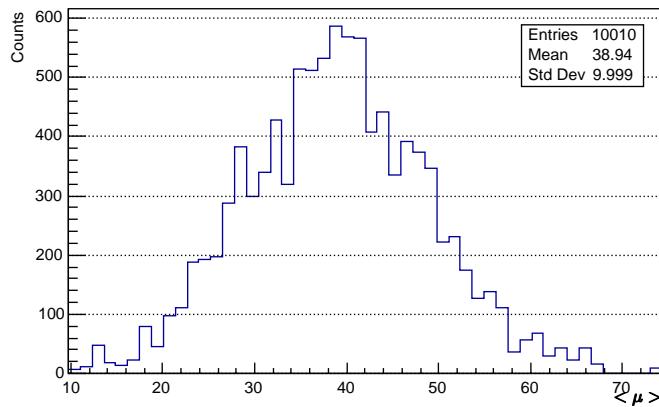
The region between  $0 < \eta < 1.3$  covers a polar angle interval of  $\Delta\theta \approx 60^\circ$ , so to properly account for the differences in the energy deposition profile due to the  $\theta$  (or  $\eta$ ) dependence,  $\eta$  segmentation is required. Exploratory tests were done using two, three, four and five bins, and we achieved better results with four. Intervals of  $\Delta\eta = 0.3$  were used, except in the central region, where  $\Delta\eta = 0.4$ . In a similar way, nine bins were defined for  $E_t$  to deal with the expected variations in the calorimeter profile due to the energy dependence. In this case, smaller intervals of 10 GeV were used in the lower energy ranges, where it is expected larger changes in the deposition profiles. Beyond 50 GeV, the intervals increase gradually. Considering, this, the proposed calibration procedure is designed for different intervals of  $|\eta|$  and  $E_t$  with the following segmentation:

- $|\eta| — [0, 0.4, 0.7, 1.0, 1.3];$
- $E_t — [10, 20, 30, 40, 50, 70, 100, 150, 200, 500] \text{ GeV}.$

As four intervals in  $|\eta|$  and nine in  $E_t$  were used, the 2D binning of the  $|\eta| \times E_t$  phase-space provides 36 independent regions. This means that the proposed calibration strategy comprises an ensemble of 36 specialized machine learning models.

In the Lorenzetti Simulation framework, as shown in equation (2.1), the calorimeter cell energy is estimated using optimal filter and conversion factors. As described in [19], such factors were defined based on minimum-bias simulation to improve energy resolution by reducing the signal-to-noise ratio.

Simulations also include multiple simultaneous pp collisions at each bunch crossing, resulting in multiple interactions in the calorimeter (pileup of events). Using the Lorenzetti framework, minimum bias events were generated and merged with the single electron events to simulate in-time (coming from the same bunch-crossing) and out-of-time (originating from different bunch-crossings) pileup. Figure 4 illustrates the pileup ( $\mu$ ) distribution<sup>3</sup> in a given bunch-crossing, used for this study with an average pileup spanning in a range similar to the 2022 LHC conditions ( $\mu = 40$ ) [24]. Instructions to access the used dataset (.root files) and a Python notebook code to open it are available at this [link](#).



**Figure 4.** Pileup distribution used to merge with the single electron samples. The average  $\mu$  is around 40, simulating the 2022 LHC conditions.

### 3 Calorimeter rings as relevant features for electron energy calibration

Historically, the ATLAS experiment at the LHC is using physics-driven variables (known as shower shapes) computed from calorimeter cells' energies and geometric positions for the purposes of particle identification and calibration [8, 11]. Those shower shapes can be viewed as a kind of momentum of the EM shower. Such variables are traditionally used for particle characterization in trigger or offline analysis and may describe, for example, longitudinal shower properties, shower width, and hadronic leakage.

A different feature extraction strategy has been proposed to develop a more efficient trigger for electrons inside the ATLAS experiment using the natural shape of an EM shower to compute transverse energy ring-sums [17, 18] around the highest energy cell of a possible electron candidate energy deposit. Such features provide a detailed description of the energy deposition profile in a compact representation, and their combination with machine learning methods has been shown to produce high efficiency for electron discrimination against QCD jets background.

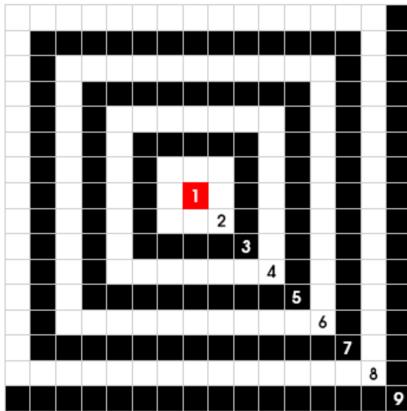
This work proposes to extend the use of calorimeter ring-sums for electron energy calibration in an online trigger context. The ring feature extraction exploits the approximately concentric nature of electron energy depositions in calorimeters. As illustrated in figure 5, starting from the most energetic (hottest) cell, which is defined as the central (first) ring, the sequential rings are defined within a

<sup>3</sup> $\mu$  represents the number of simultaneous interactions.

squared (or rectangular) region of interest (RoI) in the  $\eta \times \phi$  plane. After that, the cells  $i$  transverse energies ( $e_{k,i}$ ) are summed within each ring  $k$  to compute the ring transverse energy ( $R_k$ ):

$$R_k = \sum_{i \in \text{Ring } k} e_{k,i} \quad (3.1)$$

The rings are computed over all the calorimeter longitudinal segmentation by repeating the same procedure layer-wise. By doing this, it is possible to improve the signal-to-noise ratio of the individual cells using ring sums, especially in the tail of the energy distributions along the calorimeter layers.



**Figure 5.** Illustration of the ring-sums construction around the hottest cell (marked in red). The index  $k$  is annotated in each ring (from 1 to 9).

In this work, the rings were computed over a  $0.2 \times 0.2$  RoI in  $\eta \times \phi$ , which is expected to contain most of the valuable information for electron characterization, considering the detector granularity and layer segmentation described previously for the Lorenzetti framework. As shown in table 2, considering the different sensor granularity from each calorimeter layer, the number of rings may vary from 2 (in the hadronic layers) to 32 (in EM1). For layers with “squared” cells in  $\eta \times \phi$ , such as EM2 (which presents cells granularity of  $0.025 \times 0.025$  in  $\eta \times \phi$ , information previously shown in table 1), the rings are forming closed paths, such as described previously in figure 5. In this case, as each ring (except the central one) comprises two rows (and columns) of cells, a  $0.2 \times 0.2$  RoI accommodates four rings. For the calorimeter layers where the sensor cells are not “squares” in  $\eta \times \phi$ , the rings do not generate a closed path (and look more like parallel strips), this is the case for EM1, which has a thinner segmentation in  $\eta$  (its cell granularity is  $0.003 \times 0.1$  in  $\eta \times \phi$ ). For such a layer, the number of rings is determined by the direction thinly segmented (in this case  $\eta$ ). Taking into account all layers, a total of 50 rings are generated.

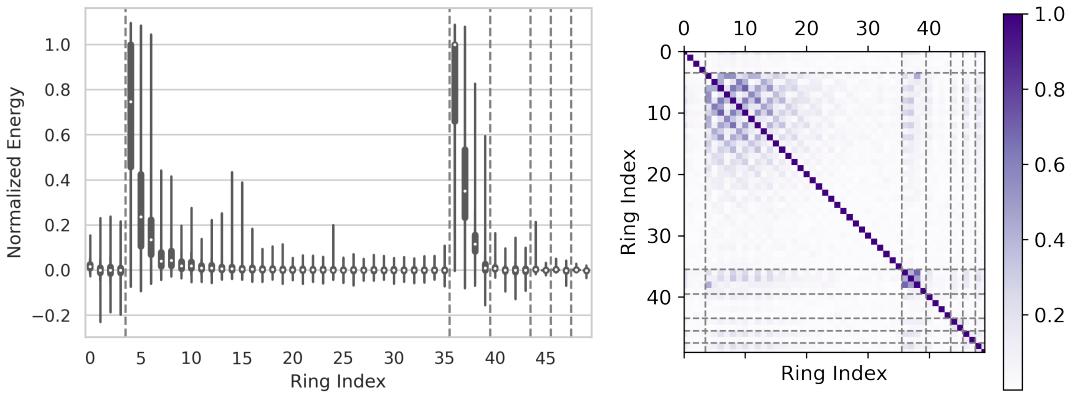
**Table 2.** Number of rings produced along the Lorenzetti calorimeter segmentation.

Layer	PS	EM1	EM2	EM3	HD1	HD2	HD3
Rings	4	32	4	4	2	2	2

The normalized ring boxplot profile for electrons with truth energy between 30 and 50 GeV is shown in figure 6 (left). It is possible to observe that the first rings of EM1 and EM2 (ring indexes 5 and 36, respectively) collect most of the energy of EM particles, while hadronic rings have, in that case,

very low energy deposition fractions. In figure 6 (right), the calorimeter rings auto-correlation matrix is presented, showing that higher correlations appear for adjacent rings within the same calorimeter layer. In the EM1 layer (ring indexes between 5 and 35), the very thin cell granularity (which generates non-closed rings) seems to produce a small autocorrelation between some specific rings, such as the second EM1 ring. So, this may be an effect produced by the lateral decay profile of electrons associated with the irregular rings of this layer. This pattern is not observed for other layers.

A similar analysis is repeated in figure 7 when the pileup of events is considered. In this case (left plot) electron's energy is spread beyond the center of the clusters, increasing the average energy on the second ring of EM1 and EM2 (ring indexes 6 and 37, respectively). An increase in the fluctuation of the hadronic rings is also observed due to pileup in those layers. The autocorrelation plot presented in the right panel of figure 7 highlights the fact that correlation becomes more intense around the center of the cluster for EM1 and EM2. There is also a slight increase in the inter-layer correlation, especially considering EM1 vs EM2 and EM1 vs EM3. For the pileup case, where several final-state particles are together in the same event, a more uniform autocorrelation behaviour is produced if compared to the no-pileup case.



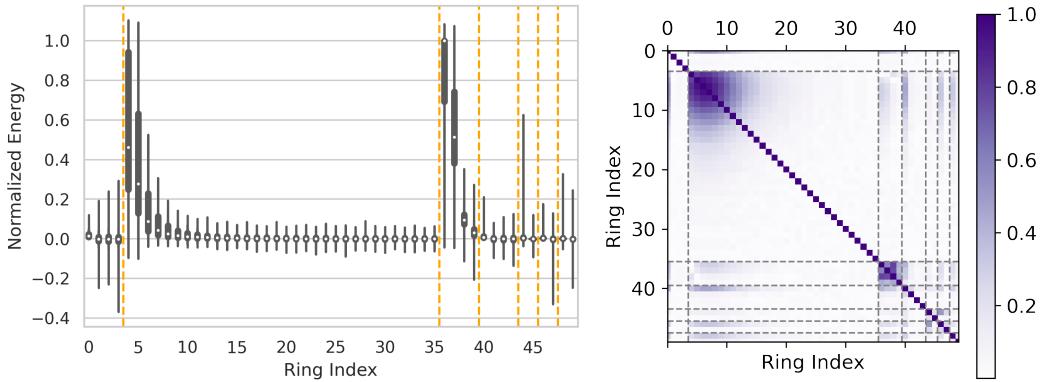
**Figure 6.** Calorimeter ring profiles for electrons (left) and their auto-correlation (right), the calorimeter longitudinal segmentation is marked using vertical grey dashed lines that separate PS, EM1, EM2, EM3, H1, H2, and H3 layers (from rings 0 to 49).

The current ATLAS experiment's shower-shape-based calibration procedure is used in this work as a baseline to compare the proposed method with the state-of-the-art calibration methodology. As described in table 3, this baseline method uses the energy ratio deposited in sequential layers, the total electromagnetic energy, the hadronic energy ratio, and the shower center position as calorimeter shower descriptors.

To estimate the relevance of the calorimeter rings to the calibration problem, the correlation coefficient ( $\Delta_E$ ) and the mutual information ( $I$ ) between each ring  $R_k$  and the calibration factor  $\alpha$  were estimated from eqs. (3.2) and (3.3):

$$\rho_{R_k, \alpha} = \frac{E[(R_k - \bar{R}_k)(\alpha - \bar{\alpha})]}{\sigma_{R_k} \sigma_\alpha}, \quad (3.2)$$

$$I(R_k; \alpha) = \sum_{y \in \alpha} \sum_{x \in R_k} p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right), \quad (3.3)$$



**Figure 7.** Calorimeter ring profiles for electrons (left) and their auto-correlation (right) when pileup is considered. The calorimeter longitudinal segmentation is marked using vertical grey dashed lines that separate PS, EM1, EM2, EM3, H1, H2, and H3 layers (from rings 0 to 49).

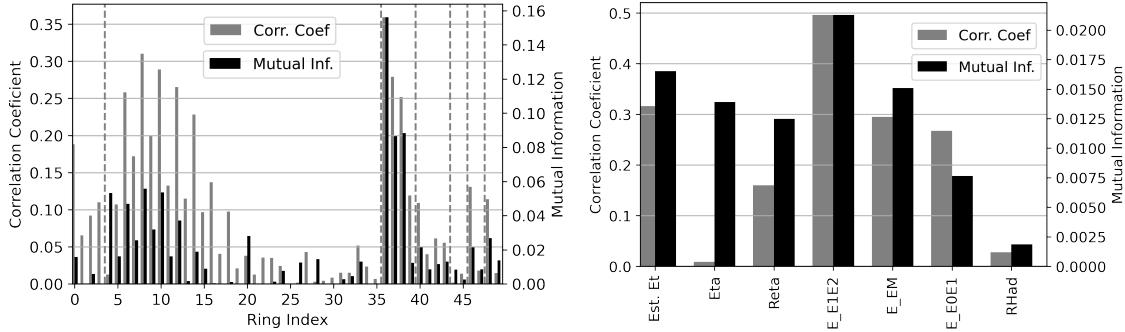
**Table 3.** Shower shapes used in this work for energy calibration as a baseline comparison to the ring-sums based calibration.

Shower-shape	Description
$E_T$ cluster	Estimated cluster energy
$\eta_{\text{cluster}}$	Estimated position of the central cell
$\frac{E_{1\text{raw}}}{E_{2\text{raw}}}$	Ratio between the energy estimated at EM1 and EM2 calorimeter layers
$E_{\text{EM}} = E_{1\text{raw}} + E_{2\text{raw}} + E_{3\text{raw}}$	EM energy sum
$\frac{E_{0\text{raw}}}{E_{1\text{raw}}}$	Ratio between the energy estimated at PS and EM1
$\frac{E_{\text{Had}}}{E_{\text{EM}}}$	Ratio between HAD and EM energies

where  $E[\cdot]$  is the expectation operator,  $\bar{X}$  and  $\sigma_x$  are the mean and standard deviation of the random variable  $x$ ; and  $p(\cdot)$  is the probability density function (pdf). For the mutual information estimation, a method based on entropy estimates from  $k$ -nearest neighbor distances [25] was applied. The same procedure was adopted for the shower shapes for comparison. These quantities provide complementary information as  $\rho$  estimates the linear relation between two random variables while  $I$  accounts for higher-order (nonlinear) dependence. Considering both parameters together leads to a better understanding of the variables' relations.

As shown in figure 8 (left), for single electron events with no pileup, it is possible to observe that some rings exhibit considerable potential to be used as inputs in a machine-learning regressor to estimate the correction factor. Notably, the initial rings from EM1 and EM2 are highly correlated to  $\alpha$  (as defined in eq. (2.3)), and these EM2 rings also present high mutual information with  $\alpha$ .

Considering the shower shapes, figure 8 (right) shows that their correlations with  $\alpha$  are comparable to the ones exhibited by rings, while their mutual information values are much smaller, reinforcing the claim that ring-sums should add relevant information for energy calibration. Similar profiles were observed for the electron events with pileup. These results motivate using the rings set as input for the proposed machine learning calibration method (instead of the shower-shape variables), as the nonlinear processing provided by such regressors can explore the natural nonlinear correlations between the rings and  $\alpha$ .

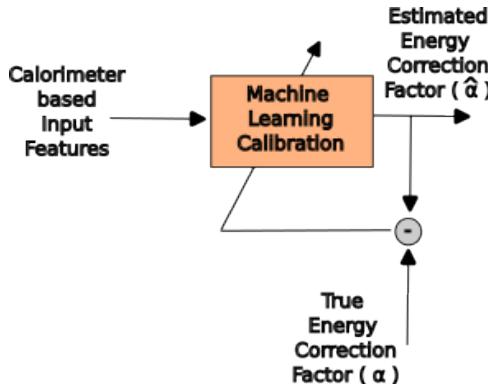


**Figure 8.** Correlation/mutual information computed with respect to the energy calibration factor ( $\alpha$ ) and the calorimeter rings (left), and shower variables (right). For the left plot, the dashed grey lines indicate the boundaries of the calorimeter layers: PS, EM1, EM2, EM3, H1, H2, H3.

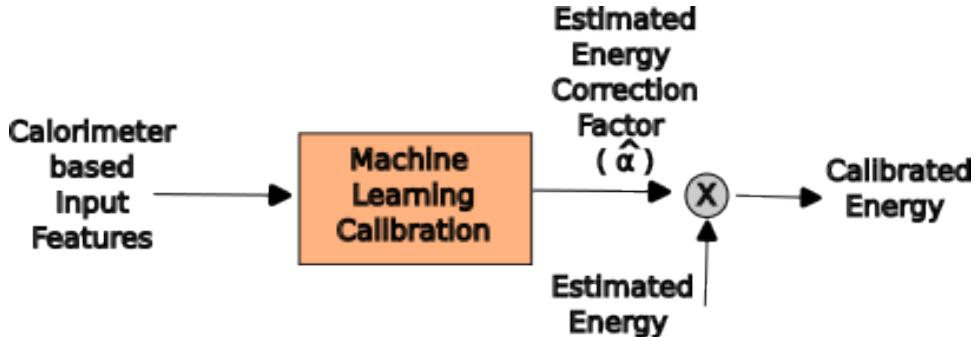
#### 4 The proposed energy calibration method

As illustrated in figure 9, this work proposes using a machine learning (ML) method for energy calibration, which is trained as a regressor to estimate an approximation  $\hat{\alpha} \approx \alpha$  of the optimal calibration factor (see eq. (2.3)). In the operation stage, the correction factor estimated by the regression machine is used to obtain the calibrated energy, as shown in eq. (4.1) and illustrated in figure 10.

$$E_{\text{Calibrated}} = \hat{\alpha} \times E_{\text{Estimated}} \quad (4.1)$$



**Figure 9.** Training phase diagram for the proposed machine learning calibration method showing the true ( $\alpha$ ) and estimated ( $\hat{\alpha}$ ) correction factor.



**Figure 10.** Operating phase diagram for the proposed machine learning calibration methods.

A similar calibration strategy has been adopted in [8] using the shower shapes as input features of a boosted decision tree (BDT) regression algorithm. This work proposes using calorimeter ring sums to replace these variables and achieve more accurate energy estimation results. To allow a comparative baseline for the ring-sums-based calibration, a shower-shape counterpart calibration was developed for each  $|\eta| \times E_t$  phase-space bin using the variables described in table 3 as inputs for the machine-learning regressor, which has also been chosen to be a BDT to better compare the impact of changing shower shapes by ring-sums.

#### 4.1 Boosted decision trees

Decision trees (DT) [26] are a popular machine-learning method for knowledge representation and discovery. The results are usually competitive with state-of-the-art algorithms, with the advantage of presenting easily understandable learning and predicting processes, especially compared to more complex approaches such as deep neural networks [27].

Boosted decision trees (BDT) consist of an ensemble learning model [28] that aggregates the predictions of several DTs, which individually perform only slightly better than random guessing. The results of these multiple simple models (*weak learners*) are combined to produce a more powerful representation that tends to perform better. In ensemble learning, bootstrapping (training individual base models based on randomly selected data subsets) and boosting (combining base models to create strong classifiers or regressors) produce enhanced models. In this context, gradient boosting [29] is an extension incorporating a gradient descent optimization strategy into the boosting process to minimize the loss function when adding new models to the ensemble.

In HEP, DT has been widely applied for tasks such as level-1 muon trigger at the CMS experiment [30], particle identification at the CALICE hadronic calorimeter [31], and multi-jet topologies classification on ALEPH [32]. A modified boosted decision tree (BDT) was proposed for high-level triggering in [33], and an FPGA implementation library for decision trees in HEP applications allows for their efficient use at the hardware level [34], especially in trigger applications.

This work used the LightGBM framework [35] to train the gradient-boosted decision trees (GBDT) models for the electron energy calibration task. This framework provides a straightforward configuration of the model architecture and training hyperparameters and is also optimized for large datasets, allowing a fast training process.

## 4.2 Evaluating the calibration effects on energy resolution

For evaluating the results achieved after calibration, the normalized energy estimation error was computed using eq. ((4.2)):

$$Er_N = \frac{Et - \hat{Et}}{Et}, \quad (4.2)$$

where  $E$  and  $\hat{E}$  are the true and the estimated energy values, respectively. To have more precise energy measurements after calibration, the spread of the distribution of  $Er_N$  must be reduced, and its central value should be moved toward zero.

An alternative way to quantify the calibration results is to evaluate the interquartile distance (IQR) of the ratio of reconstructed and truth energies ( $\alpha$  factor) before and after calibration. The IQR measures the statistical distribution width (spread), which is useful for characterizing symmetric and asymmetric variables. After calibration, the  $\alpha$  distribution is expected to reduce the deviation from its central value, becoming narrower, as the differences between reconstructed and truth energies should diminish.

Another important aspect is estimating the energy resolution ( $\frac{\sigma_E}{E}$ ) as a function of the true particle energy. In calorimeters, the energy resolution is usually modeled as a function of  $E$  using the following expression in eq. ((4.3)):

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c, \quad (4.3)$$

where  $\oplus$  stands for the quadratic sum operator, and the constants  $a$  (sampling, or stochastic term),  $b$  (noise term integrating both electronic and pileup noises), and  $c$  (constant term) are related to fluctuations in the shower development, noise, and instrumentation issues (such as shower leakage, saturation, non-linearity, and all the geometric inhomogeneities in the calorimeter structure, i.e. dead materials, cables, etc), respectively.

In this work, the Lorenzetti calorimeter resolution is estimated using both raw (uncalibrated) and calibrated measurements. These results are compared to observe the effects of calibration on the calorimeter performance.

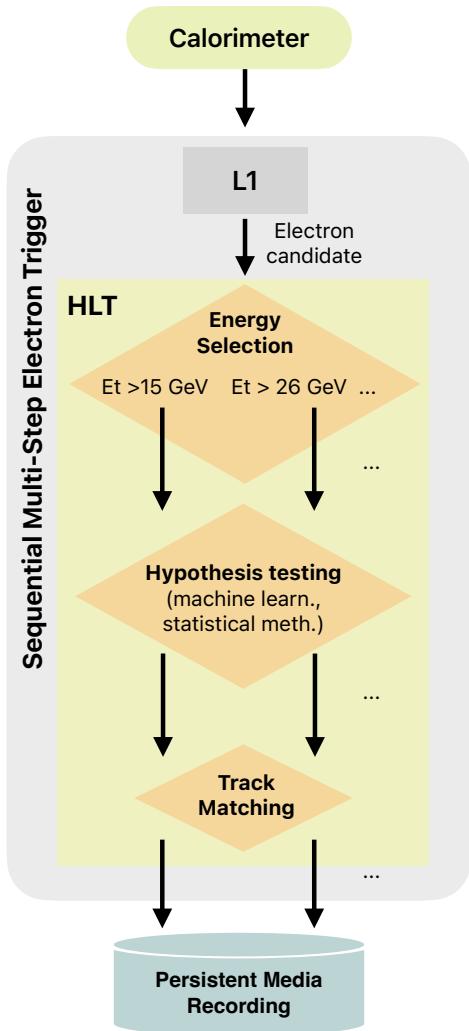
## 4.3 Evaluating the calibration effects on triggering

Electron triggers often use a collection of multi-step selection chains combining hypothesis testing based on calorimeter information and track matching (to distinguish electrons from photons). In the initial step, the candidate events are selected based on their raw estimated energy [36–38]. Figure 11 illustrates such a general electron trigger chain, which has as a first step the energy selection block. The energy estimation error may affect the overall trigger efficiency as the chain selector may assign particle candidates with incorrect energy levels to a given trigger chain.

Considering an electron trigger chain ( $EX_E$ ) that should operate over particle candidates with total (true) electromagnetic energy above  $X_E$  GeV, an energy threshold ( $\lambda$ ) is applied to the estimated electromagnetic energy. Considering that typically, there is an underestimation concerning the true transverse energy value, one can define a particle candidate selection such as:

$$\text{Accept a candidate if } \hat{Et} > \lambda \quad (4.4)$$

with  $\lambda = X_E - \Delta_E$ , where  $X_E$  is the desired truth energy threshold and  $\Delta_E > 0$  is a bias correction constant.



**Figure 11.** Diagram of a generic multi-step electron trigger, showing the level 1 (L1), usually implemented in dedicated hardware and the high-level trigger (HLT), implemented in software. The HLT may also be split into intermediate steps such as energy cuts, hypothesis testing based on calorimeter information and track matching algorithms.

The constant  $\Delta_E$  may be reduced after calibration as the expected energy estimation accuracy improves (as  $\hat{E}$  will be closer to  $E$ ). Considering this, it is possible to optimize the trigger efficiency by reducing the number of candidates that do not have the truth energy above the desired threshold (fakes) and are incorrectly selected by applying eq. (4.4) (due to energy resolution effects).

To evaluate such effects, a Receiver Operating Characteristic (ROC) analysis is used [39]. This is done by plotting the probability of fake acceptance (PF) versus the probability of detecting the proper candidates (PD) for different values of the threshold correction term ( $\Delta_E$ ). The menu optimization can also be evaluated by comparing the non-calibrated and the calibrated cases for a specific detection efficiency.

## 5 Calibration results

After exhaustive (grid-search) testing using different combinations of hyperparameters for the BDTs, accurate results were achieved for the different input features when using the configuration shown in table 4.

**Table 4.** Best GBDT hyperparameter set after exhaustive (grid-search) procedure for calibration using rings and shower variables.

Hyperparameter	Rings	Shower Variables
<b>Regularization</b>	L1	L1
<b>Leafs per tree</b>	60	45
<b>Boosting rounds</b>	2000	2000
<b>Learning rate</b>	0.05	0.09
<b>Bagging rate</b>	0.8	0.9

This section is organized as follows. The calibration effects on energy estimation accuracy are presented, followed by the calibration effects for calorimeter resolution and triggering.

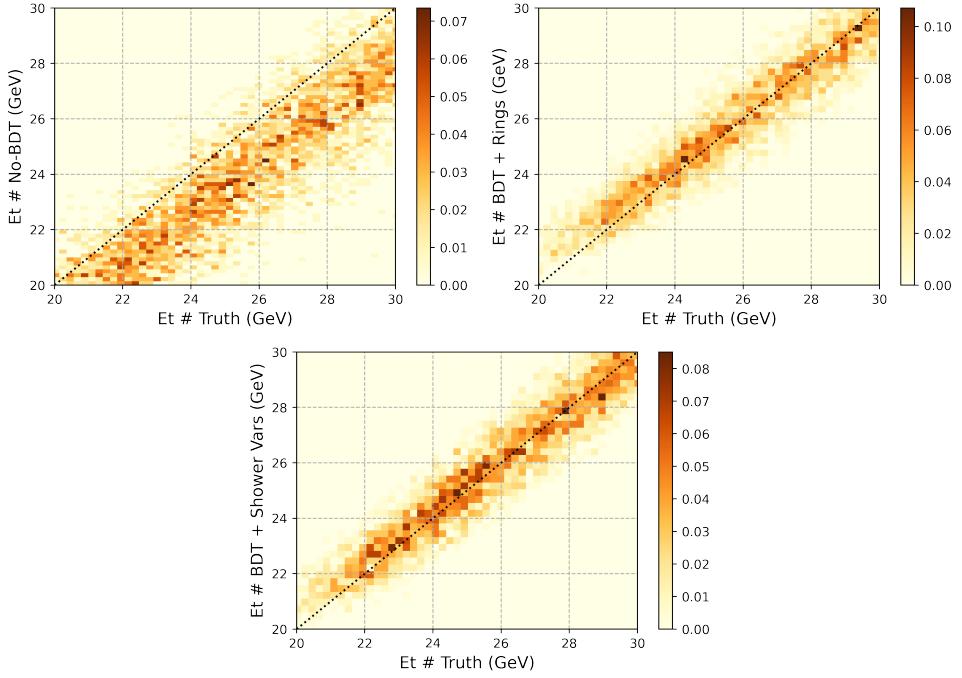
### 5.1 Energy estimation accuracy

As indicated previously, the calibration process (BDT training and testing) was performed using a  $\eta \times Et$  phase-space binning for 36 independent bins. As illustrated in the scatter-plots from figure 12 (top-left) for electrons between 20 to 30 GeV, there is a considerable mismatch between the truth (MC) and the estimated energy in the electron samples. Note that the ideal case would be if all the events were aligned on the dashed line, where  $E = \hat{E}$ . After the proposed calibration process, as we can see from figure 12 (top-right) using rings as input and figure 12 (bottom) using shower variables, the scatter plots are shifted and towards the ideal case (dashed line), and the spread is considerably reduced. It is worth noting that the calibration using rings produces a slight high concentration of the number of counts (note that the color scale indicates 0.02 points higher probability).

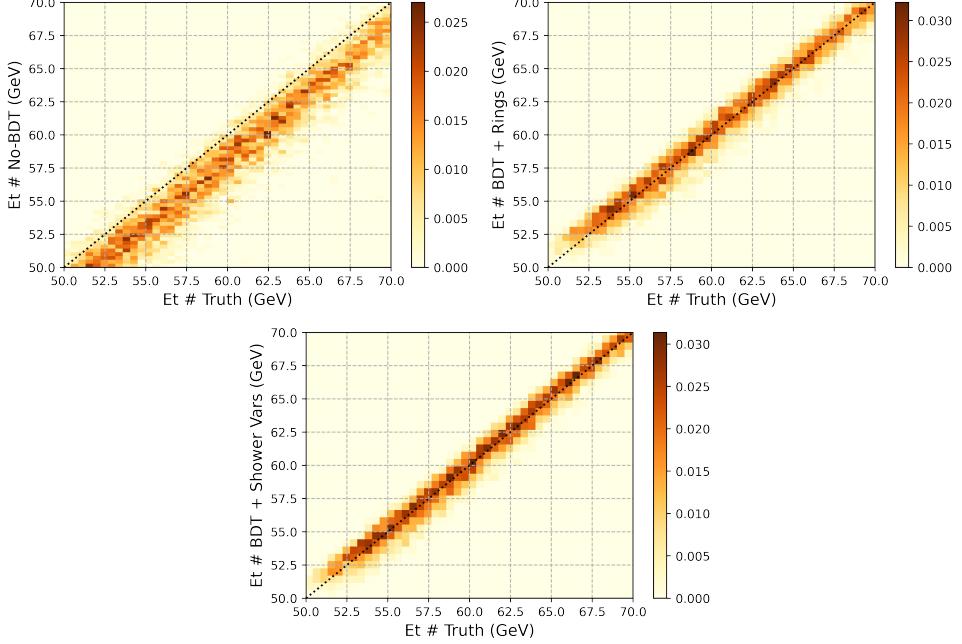
Similar behavior is observed in figure 13 for electrons between 50 to 70 GeV. It is also interesting to note that the estimation mismatch decreases as the electron energy range increases from 20 to 30 GeV (figure 12) up to 50 to 70 GeV (figure 13), showing the interest of this method to lower transverse energy trigger thresholds. When considering the pileup of underlying events, the scatter-plot analysis produces similar results and was omitted for simplicity.

A broader view of the effects of calibration for energy estimation accuracy for the different  $Et \times \eta$  phase-space bins can be obtained from the IQR plots. As can be observed in figure 14 (left), integrating the events into all  $|\eta|$  bins after calibration, the spread (IQR) of the correction factor ( $\alpha$ ) is reduced for all ranges of  $Et$ . The ratio plots at the bottom highlight the benefits of using rings compared to shower variables. Considering, for example, electron candidates with  $Et$  around 90 GeV, it is possible to reduce the IQR in  $\approx 35\%$  using rings, which represents more than 10 percent points reduction if compared to the calibration with shower variables.

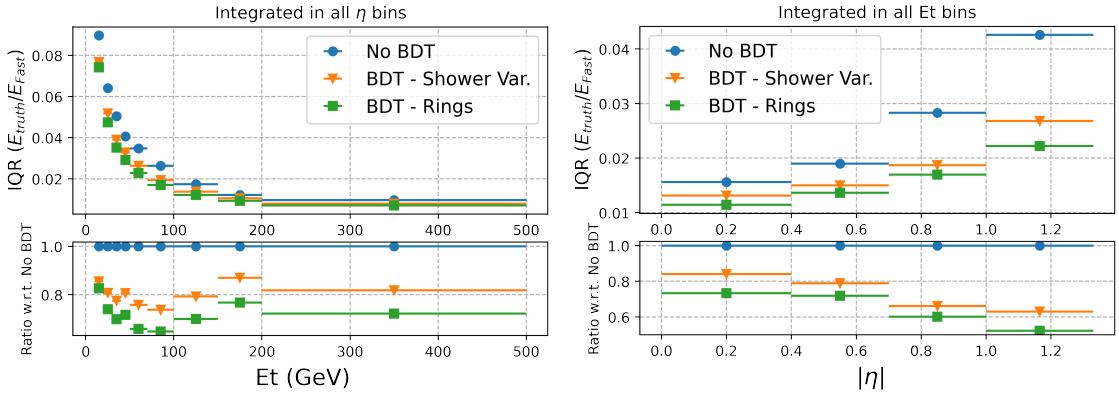
Similar results were observed in figure 14 (right), as the IQR reduction is more expressive for the calibration using ring-sums information for all  $|\eta|$  ranges (integrating the events into all  $Et$  bins). It is interesting to point out, for example, that the proposed approach reduces the IQR by more than 45% for  $|\eta| \approx 1.2$ .



**Figure 12.** 2D Histograms for the estimated  $E_t$  versus the truth  $E_t$  for 20 to 30 GeV before calibration (top-left), after calibration using BDT+rings (top-right), and after calibration using BDT+shower variables (bottom). It is possible to see the benefits of the proposed calibration methods as the histogram becomes concentrated around  $y = x$  (dotted line).

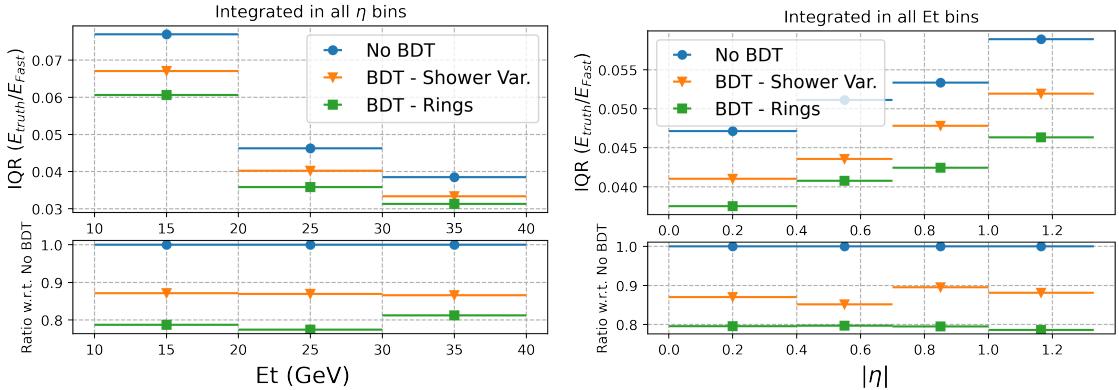


**Figure 13.** 2D Histograms for the estimated  $E_t$  versus the truth  $E_t$  for 50 to 70 GeV before calibration (top-left), after calibration using BDT+rings (top-right), and after calibration using BDT+shower variables (bottom). It is possible to see the benefits of the proposed calibration methods as the histogram becomes concentrated around  $y = x$  (dotted line).



**Figure 14.** Calibration factor ( $\alpha$ ) IQR for electrons without pileup integrated in the  $|\eta|$  bins for all  $E_t$  interval (left); and integrated in the  $E_t$  bins for all  $|\eta|$  interval (right). The ratio plots (at the bottom of each subplot) show a significant IQR reduction by using rings information for calibration.

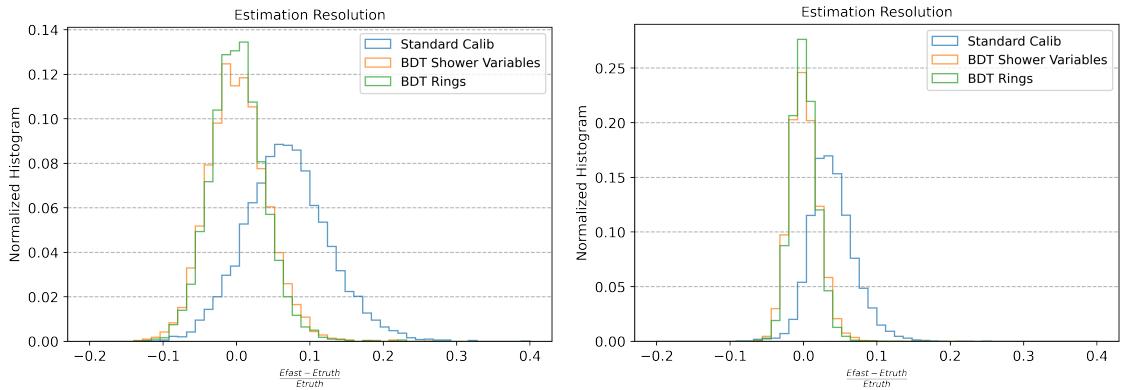
In the IQR analysis, the pileup simulation was performed for electrons with  $10 < E_t < 40$  GeV, which is the energy range where the calibration effects are more important for trigger purposes. We can observe from figure 15 that calibration using rings information also provides better results than shower variables for this case, reinforcing the benefits of the proposed calibration features for the operational conditions observed in state-of-the-art particle collider experiments.



**Figure 15.** Calibration factor ( $\alpha$ ) IQR for electrons with pileup integrated in the  $|\eta|$  bins for all  $E_t$  interval (left); and integrated in the  $E_t$  bins for all  $|\eta|$  interval (right). The ratio plots (at the bottom of each subplot) show a significant IQR reduction by using rings information for calibration.

The normalized energy estimation error histograms are shown in figure 16 for electrons with transverse energy between 20 and 30 GeV (left) and 50 and 70 GeV (right), integrating over all  $|\eta|$  bins. It is observed in both cases that the estimation error distribution peak (mode) shifts towards zero after calibration. Additionally, a considerable increase in the concentration around the mode was observed. For energy between 20 to 30 GeV (figure 16-left), it rises from  $\approx 9\%$  of the events in the mode interval before calibration to  $\approx 12.5\%$  and  $\approx 13.5\%$  after calibration using shower shapes and rings, respectively. Considering electrons with transverse energy between 50 to 70 GeV (figure 16-right) it raises from  $\approx 17\%$  to  $\approx 25\%$  (BDT+shower-shapes) and  $\approx 27\%$  (BDT+rings). Similar results were achieved for the other energy ranges and for the electron events contaminated

with pileup, and the plots were omitted here for simplicity. These results reinforce the benefits of using ring-sum information for electron energy calibration.



**Figure 16.** Histograms for the normalized energy estimation error ( $Er_N$ ) considering different energy ranges: 20 to 30 GeV (left) and 50 to 70 GeV (right). After calibration the histograms become narrower and have their peaks shifted towards zero, indicating a reduction in the estimation error.

## 5.2 Calorimeter resolution estimation

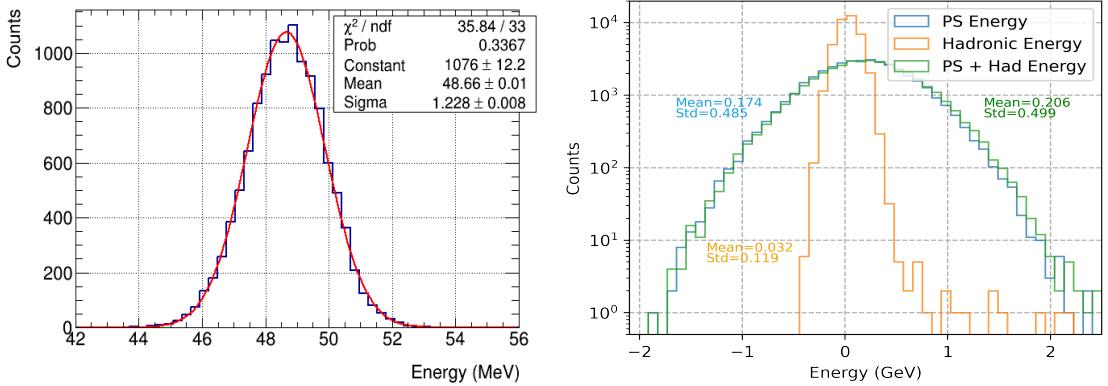
A specific dataset comprising 500000 simulated electron events with truth energies of 10, 50, 100, 200, and 300 GeV (100000 events for each energy value) was produced to estimate the Lorenzetti calorimeter resolution energy dependence using eq. (4.3). For each particle truth energy, the reconstructed energy was computed for the complete set of events, and a Gaussian distribution was fitted to the observed histogram to compute the mean and standard deviation, which were used as estimates of  $E$  and  $\sigma_E$ , respectively, for each of the five energy values. Figure 17 (left) illustrates this data processing for the 50 GeV electrons. Similar results were achieved for the other energy values.

Using such electron samples, we can observe the amount of energy lost by these particles outside the EM layers. As shown in figure 17 (right), most of the energy loss for 50 GeV electrons is in the pre-sampler, as the shower development starts before the EM calorimeter. There is also a contribution from the hadronic layers, but it tends to be smaller, as the few electron showers that pass through the EM calorimeter have a small energy fraction left. It is interesting to note that sometimes the estimated energy may be negative for some cells, especially when the sampled signal is smaller than the pedestal ( $p$ ) in eq. (2.1).

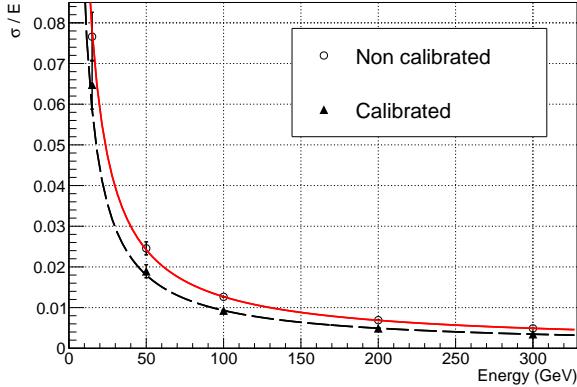
To estimate the Lorenzetti calorimeter resolution eq. (4.3) was fitted to the five  $\frac{\sigma_E}{E}$  points to estimate the  $a$  (stochastic),  $b$  (noise) and  $c$  (constant) terms. As illustrated in figure 18, the resolution plots were compared before and after calibration. It can be seen that the calibrated resolution curve is below the non-calibrated one, and as a result, both the stochastic, and electronic and pileup noise terms were reduced after calibration:  $a : 5.316 \rightarrow 3.278$  ( $\approx 62\%$  reduction) and  $b : 1150 \rightarrow 869.5$  ( $\approx 76\%$  reduction). These results reinforce the efficiency of the proposed calibration procedure.

## 5.3 Trigger performance improvement

The online event selection (trigger) system benefits from the calibration as it allows for a more accurate selection of the electron candidates at the (usually) first stage of the trigger menu, and this



**Figure 17.** Gaussian fit applied to the non-calibrated estimated energy of single electron events with truth energy equal to 50 GeV (left), and transverse energy measured in the PS and HAD layers for the electron events with truth energy of 50 GeV (right). It is possible to observe the underestimation tendency on the left plot (mean value approximately 48.66 GeV) as a consequence of the energy losses outside the EM RoI, shown in the right plot.

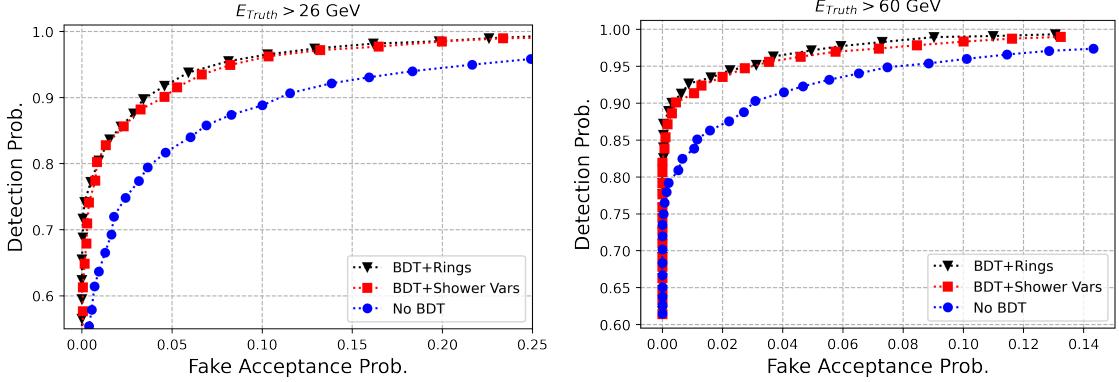


**Figure 18.** Lorenzetti calorimeter resolution estimation before and after the proposed calibration strategy.

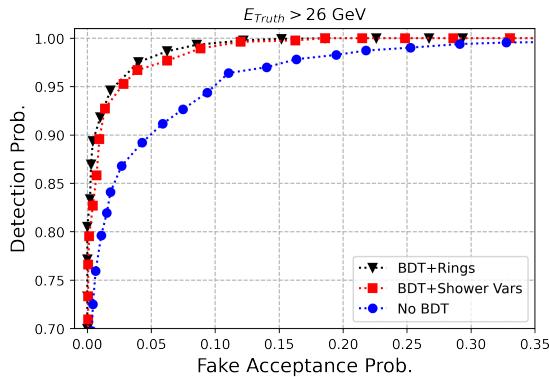
allows for a lower trigger  $E_t$  threshold while keeping the same background rate. To estimate the effects on a multi-chain trigger system, we defined four  $E_t$  cuts: E26, E60, and E120 so that an ideal selection would approve for those chains only electron candidates with truth  $E_t$  higher than 26, 60, and 120 GeV, respectively.

Aiming to accommodate energy estimation errors often present at early trigger stages, a conservative threshold  $\lambda = X_E - \Delta_E$  was applied (where  $X_E$  is the desired truth energy cut). Figure 19 shows the ROC curves achieved by varying  $\Delta_E$  from 3 to zero GeV. It is possible to observe that when applying the proposed calibration method (BDT+Rings), the fake acceptance probability is reduced considerably for all PD values. Compared with the BDT+Shower Variables case, one can see a slight improvement in the ROC performance. It is also worth noting that it is possible to achieve high PD values after calibration with BDT+Rings (approximately 97% for E26 and 95% for E60), which are not reached for the non-calibrated case. For simplicity, only the E26 and E60 ROCs were presented; similar results have been obtained for E120. Figure 20 shows the ROC curve for the E26 trigger chain when considering the pileup of events. In this case, the ROC curves after calibration also

show a higher detection efficiency since pileup brings on average a bias towards positive transverse energies. Once again, the calibration with rings as input produces a slight improvement with respect to the shower variables case. Only this trigger chain could be considered for the pileup case, as we used single electrons with energy up to 40 GeV.



**Figure 19.** ROC curves for E26 (left), and E60 (right) chains. A clear improvement is observed for the calibrated case, with slightly better performance when the rings information is used as input.



**Figure 20.** ROC curves for E26 for the pileup case. A clear improvement is observed for the calibrated case, with slightly better performance when the rings information is used as input.

Table 5 presents for each trigger chain the probability of accepting electrons with the correct energy ranges (PD) and the respective probability of accepting candidates with truth transverse energy lower than the threshold (PF) when using  $\Delta_E = 3$  GeV for the non-calibrated case. By adjusting the values of  $\Delta_E$  to maintain the same PD, we can see a considerable PF reduction for the calibrated case. Considering, for example, the E26 chain, PF was reduced from 16.2% to 7.7% ( $\approx 8.5$  percent points reduction), which represents a reduction factor of 2.1 of the background contribution. For the E120 trigger chain, the false detection rate was reduced by a factor of 4.3. From table 6, we can observe similar results for the electron events with pileup. In this case, the PF was reduced by a factor of 3.1 (from 48.5% to 15.9%).

**Table 5.** Probability of detection (PD) and of fake acceptance (PF), in %, and the energy selection threshold ( $\lambda$ ) for electron candidates considering four energy cuts: E15, E26, E60, and E120. PF is computed before and after the proposed calibration method (BDT+Rings).

	<b>PD</b> (%)	<b>PF No Calib</b> ( $\Delta_E = 3$ GeV) (%)	<b>PF Calib</b> (%)	<b>PF Reduc.</b> <b>factor</b>	<b>Calib</b> $\Delta_E$ (GeV)
<b>E26</b>	93.9	16.2	7.7	2.1	2.3
<b>E60</b>	92.7	6.6	2.8	2.4	1.5
<b>E120</b>	95.6	1.3	0.3	4.3	1.0

**Table 6.** Probability of detection (PD) and of fake acceptance (PF), in %, and the energy selection threshold ( $\lambda$ ) for electron candidates + pileup, considering the E26 chain. PF is computed before and after the proposed calibration method (BDT+Rings).

	<b>PD</b> (%)	<b>PF No Calib</b> ( $\Delta_E = 3$ GeV) (%)	<b>PF Calib</b> (%)	<b>PF Reduc.</b> <b>factor</b>	<b>Calib</b> $\Delta_E$ (GeV)
<b>E26</b>	99.1	48.5	15.9	3.1	1.1

## 6 Conclusion

Calorimeter energy measurements with high accuracy are essential for achieving good electron trigger performance in high-energy physics collider experiments. Calibration of calorimeters usually requires several complementary steps to obtain the conversion and correction factors. This work proposes using calorimeter information arranged into ring-shaped features for electron energy calibration in early trigger levels. Simulated electrons are used to train a boosted decision tree ensemble as a regressor for the correction factors. The results indicate that, by using calorimeter ring-sums information, it is possible to achieve a better energy estimation if compared to the traditional shower-shape variables, reducing the IQR by more than 45% for specific  $|\eta|$  ranges. Using the proposed method, improving the calorimeter resolution profile and optimizing the trigger efficiency is possible. For example, considering the E26 trigger chain, the maximum detection efficiency was increased from  $\sim 94\%$  up to  $\sim 97\%$ , and for the E120 trigger chain, the false detection rate was reduced by a factor of 4.3. The proposed method also showed good results when there is a pileup of events in the calorimeter, which is usual in LHC experiments, reinforcing its benefits for such scenarios. In this case, the rate of fake acceptance was reduced by a factor of 3.1 after calibration. The next steps include the use of different machine learning methods for the energy calibration task. A version of the proposed system is under evaluation in the ATLAS experiment trigger system to possibly replace the current constant calibration method. Preliminary studies indicate that no significant increase in the latency time was observed while achieving a reduction in the fake candidates. These results are being compiled for a future publication.

## Acknowledgments

The authors thank CNPQ, CAPES, FAPESB, FAPEMIG and FAPERJ (Brazil), COFECUB and CNRS (France) for the financial support.

## A Dataset and code sharing

The used dataset was produced with the Lorenzetti Showers framework [19]. Instructions for download, installation, and running the simulation framework are available at <https://github.com/lorenzetti-hep/lorenzetti/blob/master/README.md>.

Additionally, Instructions to access the used dataset (.root files) and a Python notebook code to open it are available at [https://github.com/EduardoSimas/electron\\_calibration](https://github.com/EduardoSimas/electron_calibration).

**Data Availability Statement.** This article has associated data in a data repository. Available at <https://github.com/lorenzetti-hep/lorenzetti/blob/master/README.md>.

**Code Availability Statement.** This article has associated code in a code repository. Available at [https://github.com/EduardoSimas/electron\\_calibration](https://github.com/EduardoSimas/electron_calibration).

## References

- [1] H.M. Gray, *Future colliders for the high-energy frontier*, *Rev. Phys.* **6** (2021) 100053.
- [2] D.J. Griffiths, *Introduction to elementary particles*, 2nd revised edition, Physics textbook, Wiley, New York, NY, U.S.A. (2008).
- [3] L. Evans and P. Bryant, *LHC Machine*, [2008 JINST \*\*3\*\* S08001](https://doi.org/10.1088/1748-0221/3/08/S08001).
- [4] ATLAS and CMS collaborations, *Higgs Physics at HL-LHC*, in the proceedings of the *30th International Workshop on Deep-Inelastic Scattering and Related Subjects*, East Lansing, U.S.A., March 27–31 (2023) [[arXiv:2307.07772](https://arxiv.org/abs/2307.07772)].
- [5] ATLAS collaboration, *Higgs measurements at the LHC and HL-LHC*, [ATL-PHYS-SLIDE-2024-233](https://cds.cern.ch/record/2850000) (2024).
- [6] R. Wigmans, *Calorimetry: Energy measurement in particle physics*, Clarendon Press (2017) [[DOI:10.1093/oso/9780198786351.001.0001](https://doi.org/10.1093/oso/9780198786351.001.0001)].
- [7] M. Livan and R. Wigmans, *Calorimetry for Collider Physics, an Introduction*, Springer (2019) [[DOI:10.1007/978-3-030-23653-3](https://doi.org/10.1007/978-3-030-23653-3)].
- [8] ATLAS collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 2 data*, [2024 JINST \*\*19\*\* P02009](https://doi.org/10.1088/1748-0221/19/02/P02009) [[arXiv:2309.05471](https://arxiv.org/abs/2309.05471)].
- [9] CMS collaboration, *Obtaining the ultimate calibration and performance of the CMS Electromagnetic Calorimeter in LHC Run 2*, *PoS EPS-HEP2023* (2024) 533.
- [10] CMS collaboration, *Performance of the CMS electromagnetic calorimeter in pp collisions at  $\sqrt{s} = 13 \text{ TeV}$* , [2024 JINST \*\*19\*\* P09004](https://doi.org/10.1088/1748-0221/19/09/P09004) [[arXiv:2403.15518](https://arxiv.org/abs/2403.15518)].
- [11] ATLAS collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at  $\sqrt{s} = 13 \text{ TeV}$* , *Eur. Phys. J. C* **79** (2019) 639 [[arXiv:1902.04655](https://arxiv.org/abs/1902.04655)].
- [12] ATLAS collaboration, *Reconstruction, identification, and calibration of electrons and photons at ATLAS and CMS*, [ATL-PHYS-SLIDE-2024-155](https://cds.cern.ch/record/2850000) (2024).
- [13] L. de Oliveira, B. Nachman and M. Paganini, *Electromagnetic Showers Beyond Shower Shapes*, *Nucl. Instrum. Meth. A* **951** (2020) 162879 [[arXiv:1806.05667](https://arxiv.org/abs/1806.05667)].
- [14] D. Acosta et al., *The performance of a lead / scintillating fiber calorimeter at LHC / SSC compatible gate widths*, *Nucl. Instrum. Meth. A* **314** (1992) 431.

- [15] G. Stimpfl-Abele, *Feature extraction and classification of calorimeter data with neural nets*, *Nucl. Instrum. Meth. A* **373** (1996) 102.
- [16] M. Kavatsyuk et al., *Front-end electronics and feature-extraction algorithm for the PANDA Electromagnetic Calorimeter*, *J. Phys. Conf. Ser.* **293** (2011) 012020.
- [17] ATLAS collaboration, *Performance of electron and photon triggers in ATLAS during LHC Run 2*, *Eur. Phys. J. C* **80** (2020) 47 [[arXiv:1909.00761](#)].
- [18] ATLAS collaboration, *Ring-shaped Calorimetry Information for a Neural Egamma Identification with ATLAS Detector*, *J. Phys. Conf. Ser.* **762** (2016) 012049.
- [19] M.V. Araújo et al., *Lorenzetti Showers - A general-purpose framework for supporting signal reconstruction and triggering with calorimeters*, *Comput. Phys. Commun.* **286** (2023) 108671.
- [20] A. Gabrielli, *Measurement of the Higgs boson mass from the  $H \rightarrow ZZ^* \rightarrow 4\ell$  and  $H \rightarrow \gamma\gamma$  channels with the ATLAS detector at the LHC*, *Nucl. Part. Phys. Proc.* **273-275** (2016) 2479.
- [21] ATLAS collaboration, *Measurement of the  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ^* \rightarrow 4\ell$  cross-sections in pp collisions at  $\sqrt{s} = 13.6$  TeV with the ATLAS detector*, *Eur. Phys. J. C* **84** (2024) 78 [[arXiv:2306.11379](#)].
- [22] CMS collaboration, *Calibration and Performance of the CMS Electromagnetic Calorimeter in LHC Run 2*, *PoS ICHEP2020* (2021) 804.
- [23] ATLAS collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *2008 JINST* **3** S08003.
- [24] ATLAS collaboration, *Software Performance of the ATLAS Track Reconstruction for LHC Run 3*, *Comput. Softw. Big Sci.* **8** (2024) 9 [[arXiv:2308.09471](#)].
- [25] A. Kraskov, H. Stögbauer and P. Grassberger, *Estimating mutual information*, *Phys. Rev. E* **69** (2004) 066138.
- [26] L. Rokach and O. Maimon, *Data Mining with Decision Trees: Theory and Applications*, World Scientific (2013) [[DOI:10.1142/9097](#)].
- [27] M. Kretowski, *Evolutionary Decision Trees in Large-Scale Data Mining*, Springer International Publishing (2019) [[DOI:10.1007/978-3-030-21851-5](#)].
- [28] C. Zhang and Y. Ma, eds., *Ensemble Machine Learning: Methods and Applications*, Springer New York (2012) [[DOI:10.1007/978-1-4419-9326-7](#)].
- [29] J.H. Friedman, *Greedy function approximation: A gradient boosting machine*, *Annals Statist.* **29** (2001) 1189.
- [30] CMS collaboration, *Boosted Decision Trees in the Level-1 Muon Endcap Trigger at CMS*, *J. Phys. Conf. Ser.* **1085** (2018) 042042.
- [31] CALICE collaboration, *Particle identification using Boosted Decision Trees in the Semi-Digital Hadronic Calorimeter prototype*, *2020 JINST* **15** P10009 [[arXiv:2004.02972](#)].
- [32] M. Mjahed, *Multivariate decision tree designing for the classification of multi-jet topologies in  $e^+e^-$  collisions*, *Nucl. Instrum. Meth. A* **481** (2002) 601.
- [33] V.V. Gligorov and M. Williams, *Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree*, *2013 JINST* **8** P02013 [[arXiv:1210.6861](#)].
- [34] S. Summers et al., *Fast inference of Boosted Decision Trees in FPGAs for particle physics*, *2020 JINST* **15** P05026 [[arXiv:2002.02534](#)].
- [35] G. Ke et al., *LightGBM: A Highly Efficient Gradient Boosting Decision Tree*.

- [36] LHCb collaboration, *The LHCb Trigger System: Present and Future*, *J. Phys. Conf. Ser.* **623** (2015) 012003.
- [37] CMS collaboration, *The CMS Trigger System*, *J. Phys. Conf. Ser.* **2375** (2022) 012003.
- [38] ATLAS collaboration, *The ATLAS trigger system for LHC Run 3 and trigger performance in 2022*, *2024 JINST* **19** P06029 [[arXiv:2401.06630](https://arxiv.org/abs/2401.06630)].
- [39] H.L.V. Trees, *Classical detection and estimation theory*, in *Detection, Estimation, and Modulation Theory, Part I*, John Wiley & Sons, Inc. (2001), p. 19–165 [[DOI:10.1002/0471221082.ch2](https://doi.org/10.1002/0471221082.ch2)].