



Article

Advanced Big Data Solutions for Detector Calibrations for High-Energy Physics

Abdulameer Nour Jalal, Stefan Oniga and Balazs Ujvari



Article

Advanced Big Data Solutions for Detector Calibrations for High-Energy Physics

Abdulameer Nour Jalal ^{1,2} , Stefan Oniga ^{3,4,*}  and Balazs Ujvari ^{5,6,*} ¹ Doctoral School of Physics, University of Debrecen, 4032 Debrecen, Hungary; nour.abdulameer@wigner.hu² NAPLIFE-WIGNER Institute, 1121 Budapest, Hungary³ Department of IT Systems and Networks, Faculty of Informatics, University of Debrecen, 4032 Debrecen, Hungary⁴ Department of Electric, Electronic and Computer Engineering, North University Center of Baia Mare, Technical University of Cluj-Napoca, 430083 Baia Mare, Romania⁵ Department of Data Science and Visualization, Faculty of Informatics, University of Debrecen, 4032 Debrecen, Hungary⁶ HUN-REN Institute for Nuclear Research, 4026 Debrecen, Hungary

* Correspondence: oniga.istvan@inf.unideb.hu (S.O.); balazs.ujvari@science.unideb.hu (B.U.)

Abstract: This investigation examines the Dead Hot Map (DHM) method and timing calibration for Run 14 Au+Au collisions in the PHENIX experiment. The DHM method guarantees data integrity by identifying and omitting defective detector towers (nonfunctional, hot, and very hot towers) via a set of criteria and statistical evaluations. This procedure entails hit distribution analysis, pseudorapidity adjustments, and normalization, resulting in an enhanced map of functional detector components. Timing calibration mitigates the issues associated with time-of-flight measurement inaccuracies, such as slewing effects and inter-sector timing differences. Numerous corrections are implemented, encompassing slewing, tower-specific offsets, and sector-by-sector adjustments, resulting in a final resolution of 500 picoseconds for the electromagnetic calorimeter. These calibrations improve the accuracy of photon and π^0 measurements, essential for investigating quark–gluon plasma in high-energy nuclear collisions.

Keywords: big data analytics; high-energy physics (HEP); detector calibration; data-processing pipelines; dead hot map (DHM); timing calibration; statistical analysis tools



Academic Editor: Peter Baumann

Received: 29 March 2025

Revised: 30 April 2025

Accepted: 30 April 2025

Published: 21 May 2025

Citation: Jalal, A.N.; Oniga, S.; Ujvari, B. Advanced Big Data Solutions for Detector Calibrations for High-Energy Physics. *Electronics* **2025**, *14*, 2088. <https://doi.org/10.3390/electronics14102088>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Quantum Chromodynamics (QCD) is a non-Abelian theory of quantum fields that describes the strong interactions between particles. In the field of ultra-relativistic heavy ion (URHI) physics, one of the main goals is to study non-perturbative phenomena in QCD. When URHI collide, a state of matter called quark–gluon plasma (QGP) forms [1]. This phenomenon makes it possible to study the QCD phase transition. QGP describes an environment characterized by the absence of binding between quarks and gluons. Such a situation was present at the very beginning of the universe, an extremely quick period after the Big Bang. As the universe underwent expansion and cooling, quarks and gluons, which are constituents of subatomic particles known as partons, became more widely dispersed [2]. Ultimately, a phase change occurred. Particles combined and became confined within colorless hadrons due to an increased binding force between them [3].

High-energy physics experiments, such as those conducted at RHIC and LHC, require precise detector calibration to ensure accurate measurements [4]. The PHENIX experiment's

light meson and photon group studies the quark–gluon plasma by carefully measuring photons and pions (π^0).

The study of QGP in high-energy nuclear collisions relies on precise measurements from the PHENIX electromagnetic calorimeter (EMCal). However, calibrating the highly segmented PHENIX calorimeter presents unique challenges due to its complex data and specific architecture.

- **Azimuthal Segmentation:** The PHENIX EMCal is divided into sectors, each with variations in response due to geometric factors, timing offsets, noise, and radiation exposure. These require careful corrections to ensure uniformity across the detector for accurate particle energy measurements.
- **Beam Luminosity-Dependent Noise Patterns:** PHENIX experiences noise that varies with beam luminosity, particularly affecting photon and pion measurements. Expanded noise-filtering techniques, such as the Dead Hot Map (DHM) and timing calibration, are used to isolate valid data from background noise.

In contrast to PHENIX, the ATLAS and CMS experiments at the LHC use calorimeter systems with different geometries, each posing unique calibration challenges. ATLAS has a highly segmented, barrel-shaped detector, while CMS features a cylindrical section with endcaps. Both detectors face the following:

- **Azimuthal Segmentation:** ATLAS and CMS also experience calibration challenges due to segmentation, though their more symmetrical designs make them less sensitive to sector-dependent effects than PHENIX. Still, they require detailed calibration to address energy response variations across sectors.
- **Beam-Luminosity Noise:** Both detectors experience noise from beam-induced background and luminosity fluctuations, which is mitigated through advanced filtering and noise-reduction techniques, such as the CMS machine learning-based anomaly detection system.

After finding the challenges discussed above it is important to nonfunctional towers in electromagnetic calorimeters (EMCal) and making timing resolution better, to distinguish particle types accurately is the main challenge in detector calibration. Previous studies in ALICE, STAR, and CMS have addressed similar calibration issues, but their approaches lacked the iterative optimization seen in PHENIX [5,6]. This research shows a better way to make a more effective DHM and more precise timing, which helps to investigate the formation of neutral pions and direct photons in high-energy collisions between gold nuclei (Au + Au) at a center-of-mass energy of $\sqrt{s_{NN}} = 200$ GeV. The PHENIX electromagnetic calorimeter has eight sectors, and each sector has many towers (PHENIX at RHIC, Figure 1). These towers may not work properly, which could lead to errors in the analysis [7].

The DHM finds that “bad” towers—either “dead” towers that do not respond or “hot” towers that send signals without any energy deposits—need to be removed from the analysis to keep the data correct. The identification of these errors depends on a statistical approach. Considering that energy deposits among towers must be uniformly allocated for numerous events, substantial discrepancies suggest tower malfunctions. Determining what is “statistically significant” is complex due to rapidly declining energy spectra and the inhomogeneity of failures. The researchers iteratively modify settings to optimize the DHM, striving to preserve data quality while reducing unnecessary exclusions. By checking the stability of spectral shapes across different DHM configurations, it was possible to make sure that only the broken towers were left out, the acceptance was kept, and data loss was cut down [8].

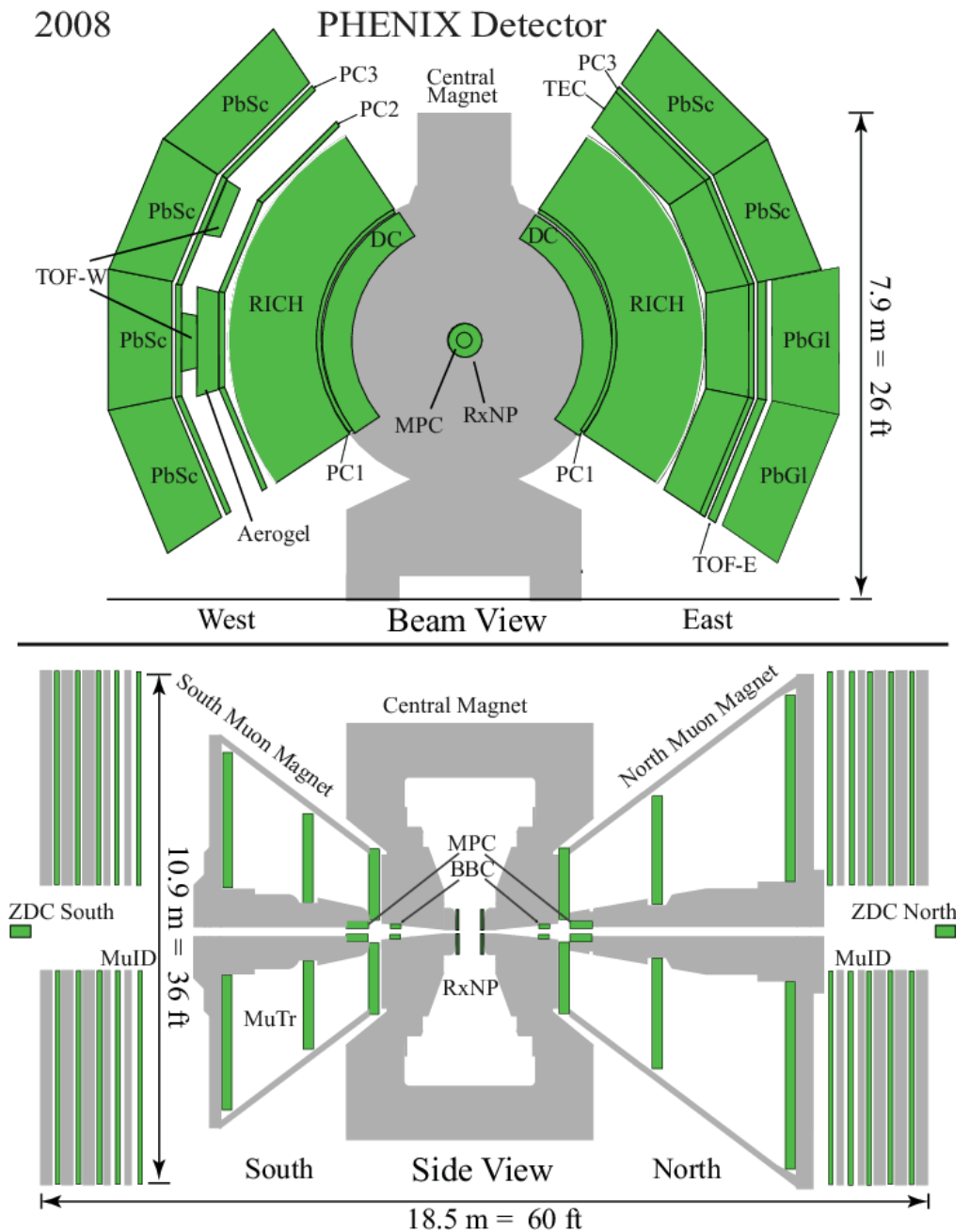


Figure 1. The PHENIX spectrometer setup was viewed both from the beam direction and the side [9].

The next study is about calibrating the timing in the PHENIX experiment. This is very important for correctly converting from the Time-to-Digital Converter (TDC) counts by the electromagnetic calorimeter to the particles' physical time of flight (ToF) measured in nanoseconds. Accurate timing is essential for distinguishing particle types, such as photons, which, being the fastest particles, serve as a reference for slewing correction. About half of the particles reaching the EMCal are photons, making them a reliable baseline for timing adjustments. We apply several corrections, such as the slewing (or walk) correction, sector-specific offsets, and high-energy timing adjustments, to address timing inaccuracies. The slewing correction can be calculated from the relationship between the photon arrival time and the ADC signal for each tower, while additional terms, such as the sector offset, are analyzed by the sector-level (several thousand towers together) data to refine the timing measurements. If these changes are not made, sector-specific differences and high-energy errors in measurements could make it harder to identify particles and resolve their

energies. Proper timing calibration ultimately enhances the accuracy of photon and π^0 measurements [10].

ATLAS and CMS benefit from more uniform detector designs and greater data-taking stability. However, their large-scale, high-resolution calorimeters still face challenges regarding energy-dependent calibration, timing precision, and sector-specific variations that are similar to those in PHENIX. The methodological innovations developed in PHENIX, such as the DHM and five-parameter slewing fit, provide significant advantages in handling sector-specific noise patterns and calibrating timing resolutions.

2. Related Work

The handling of malfunctioning detector towers is crucial for data quality assurance in high-energy physics experiments. PHINEX, CMS, and ALICE experiments have developed strategies to monitor and manage these anomalies. CMS uses a machine learning-based anomaly detection system to monitor its electromagnetic calorimeter (ECAL) in real time, detecting and localizing anomalies by analyzing spatial and temporal deviations in detector responses [5]. ALICE's ECAL has a multi-step calibration and monitoring process for fixing broken towers. These steps include data quality monitoring (DQM), bad channel masking, and cross-talk emulation. These procedures ensure high-quality data integrity while minimizing data loss due to faulty detector components [6]. PHENIX uses a DHM method [11] over 39 different energy ranges (0 to 30 GeV) along with a data-driven method to handle calorimeter towers that do not work, putting them into three groups: dead towers, hot towers, and extra-hot towers. These techniques ensure that high-energy physics experiments continue to produce reliable and precise data for fundamental research. Traditional threshold-based monitoring techniques have previously missed anomalies that PHINEX techniques have successfully detected. Table 1 provides a comparative summary of these approaches.

Table 1. Comparison of methods for handling malfunctioning towers.

Experiment	Method Used	Key Features	Strengths
PHENIX	DHM parallel with Data-Driven Dead Hot Map	Statistical hit distribution, bad tower exclusion, timing calibration	Precise data filtering, minimal data loss
CMS	Autoencoder-based ML	Real-time anomaly detection, spatial and temporal analysis	High efficiency, detects subtle anomalies
ALICE	Data quality monitoring	Energy calibration, bad channel masking, cross-talk emulation	Extensive calibration, robust quality assurance

3. DHM Method

The DHM is a critical method used in high-energy physics experiments, such as the PHENIX experiment during the Run 14 Au+Au collision. It ensures data integrity by identifying and excluding malfunctioning detector towers while preserving valid data in the electromagnetic calorimeter (EMCal). These towers are categorized into three distinct types based on their response behavior:

- **Dead Towers** : Towers that are either minimally operational or completely nonfunctional.
- **Hot Towers**: Towers that emit signals without corresponding energy deposits, often due to excessive noise.
- **Extra-Hot Towers**: Towers that consistently register excessive signals due to hardware issues, such as ADC or front-end faults, leading to an abnormally high number of hits.

The DHM procedure includes a multi-step methodology:

1. Visualizing the Raw Hit Map: Analyzing the detector's initial response before applying any corrections or conditions.
2. Run Selection and Quality Control: Filtering out unreliable runs based on event count, detector stability, and cluster distributions.
3. Hit Map Construction: Creating detailed spatial distributions of recorded hits per tower for further classification.
4. Pseudo-Rapidity (η) Correction: Normalizing hit distributions across detector regions to compensate for geometric asymmetries.
5. Statistical Analysis of Hit Distributions: Applying Gaussian, Poisson, and binomial fits to define thresholds for dead, hot, and extra-hot towers.
6. DHM Generation: Merging tower classification results across multiple energy bins to produce a stable, globally valid DHM.

Each of these steps is detailed in the following sections.

3.1. Step 1: Raw Hit Map Before Applying DHM Conditions

Before applying any corrections or selection criteria, the Raw Hit Map (Figure 2) provides a direct visualization of the detector's initial response. This map displays all recorded hits in a given sector, without any event filtering, quality cuts, or pseudo-rapidity corrections.

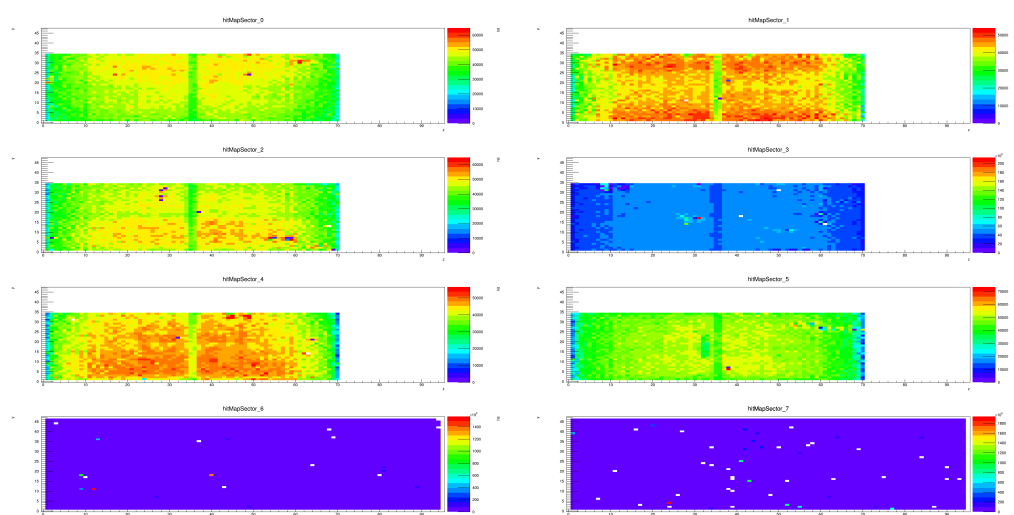


Figure 2. Raw Hit Map before applying DHM conditions. The color scale represents the number of hits per tower in a given sector. Green indicates average tower strikes (hot tower later). Yellow to red indicate more strikes per tower (extra-hot tower later). Blue and purple indicate low-hit areas (dead tower later).

Observations from the Raw Hit Map:

- High-density regions: Some towers register anomalously high hit counts, which may indicate hot or extra-hot towers due to hardware noise.
- Zero-hit regions: Certain towers show no recorded activity, potentially representing dead towers or inactive detector regions.
- Non-uniform distributions: Variations across the sector suggest geometric effects, such as pseudo-rapidity (η) dependence.

This uncorrected hit map serves as a baseline, helping to understand the necessity of event filtering, quality cuts, and statistical classification, which are applied in the following sections.

3.2. Step 2: Run Selection and Quality Control

To ensure only high-quality runs contribute to DHM calibration, strict selection criteria are applied.

Criteria for Run Selection:

- Runs with fewer than 1 million recorded events are discarded (Figure 3).
- Runs affected by hardware malfunctions, beam instabilities, or missing detector sectors are excluded.
- Only “good long runs” with stable detector operation are selected.

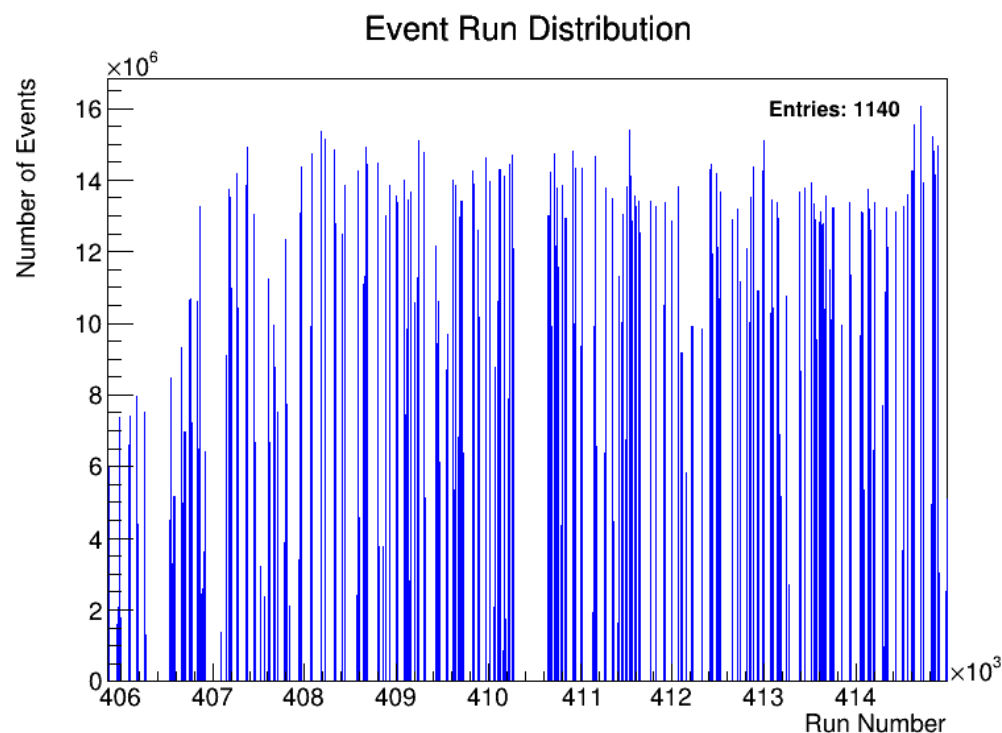


Figure 3. Number of recorded events per run, illustrating the selection of long, stable runs for DHM method.

Event-Level Quality Filtering

Beyond selecting stable runs, event-level quality checks ensure that unreliable data do not affect the final DHM. Two key parameters are monitored:

1. Zero-Hit Towers: Runs are discarded if any sector contains more than 100 zero-hit towers, as this suggests large-scale detector failure.
2. Cluster-Per-Event Distribution: Runs where any sector registers fewer than 10 clusters per event are removed to prevent data corruption from nonfunctional detector elements.

Figure 4 for sectors 0–5 and Figure 5 for sectors 6–7 represent the cluster-per-event distribution of the PHENIX detector. They illustrate how many clusters (groups of hits) were detected per event across different sectors. The distributions highlight any irregularities in the detector, such as low numbers of clusters per event, which could indicate detector malfunctions or nonfunctional sectors. These distributions are essential for quality control and ensuring that the data are reliable before proceeding with the Dead Hot Map (DHM) methodology.

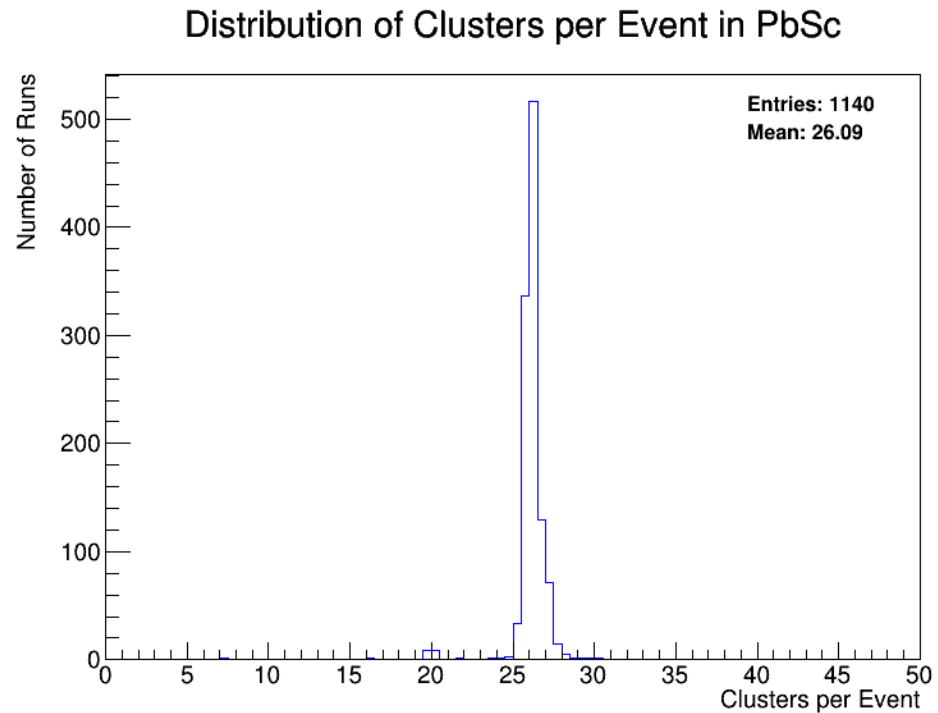


Figure 4. Cluster-per-event distribution for PbSc sectors (0–5).

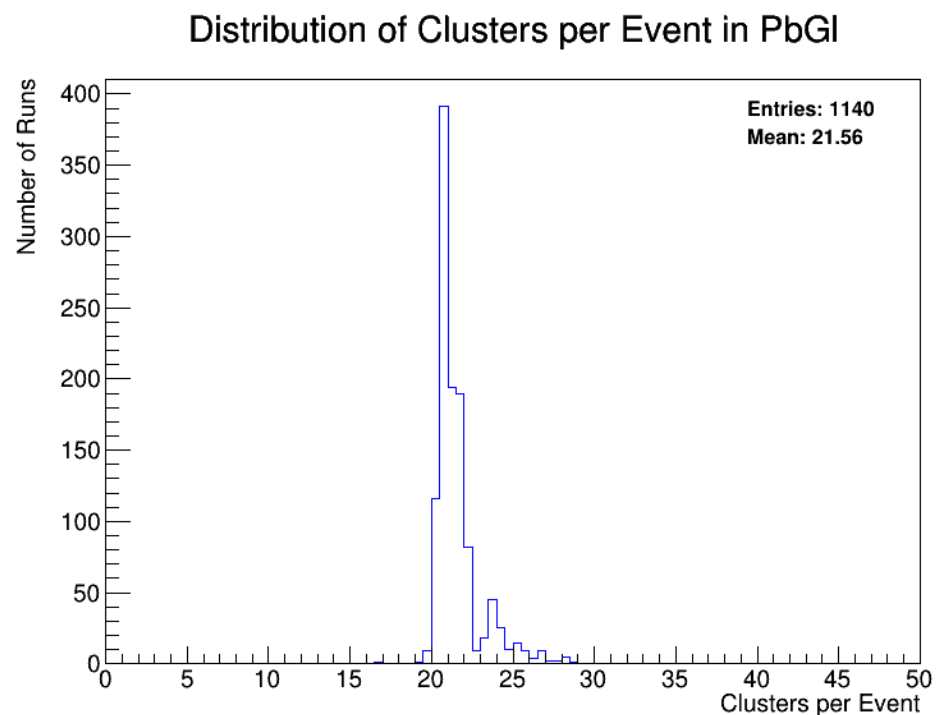


Figure 5. Cluster-per-event distribution for PbGI sectors (6–7).

3.3. Step 3: Hit Map Construction

After filtering out unreliable runs, hit maps are generated to visualize the spatial distribution of recorded hits per tower. Figures 6 and 7 show the Raw Hit Map (a), these maps represent the spatial distribution of recorded hits per tower before applying any corrections or filters. The hit distribution (b) is essential for identifying areas where the detector towers show abnormally high or low hit counts, which could signify “hot” or “dead” towers. They provide a baseline visualization for further analysis and correction.

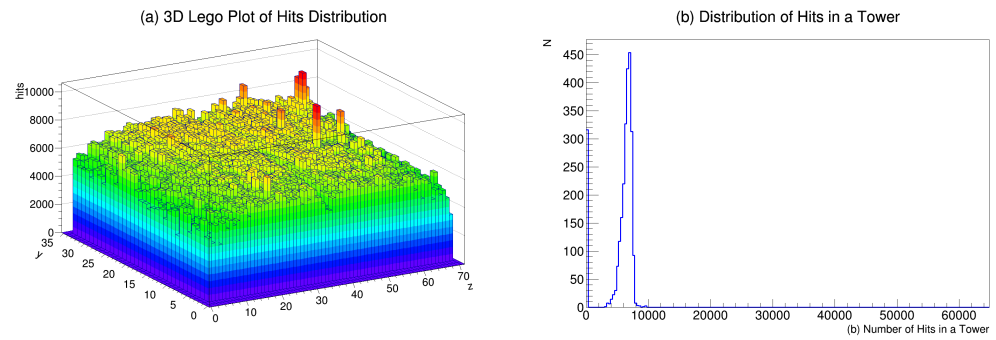


Figure 6. Raw Hit Map and hit distribution for sector 0.

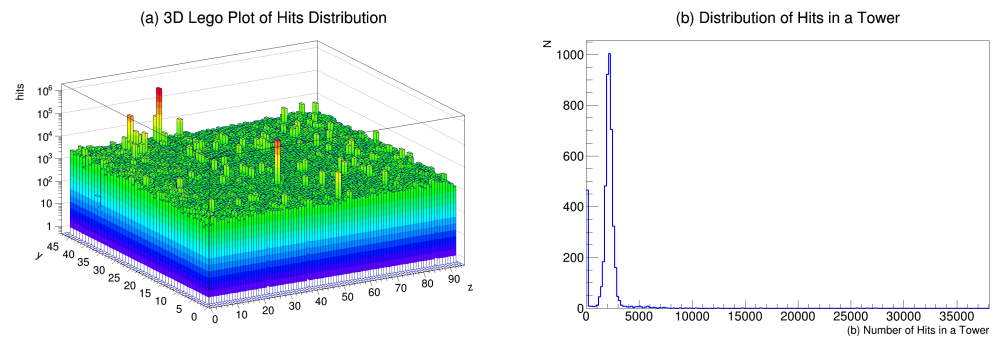


Figure 7. Raw Hit Map and hit distribution for sector 7.

3.4. Step 4: Pseudo-Rapidity (η) Correction

Due to the cylindrical geometry of the detector, tower responses vary with pseudorapidity (η). To achieve uniformity across the calorimeter, a normalization correction is applied.

Correction Process

1. Preprocessing Step: Towers with zero hits and extra-hot towers (hit counts exceeding 100 times the average) are removed to prevent distortion.
2. Normalization Histogram Calculation: A 1D normalization histogram is derived to account for η variations.
3. Final η -Correction: The normalization factor is applied to the hit map, ensuring uniform response across η regions.

Figure 8 (for sector 0) and Figure 9 (for sector 7) show the pseudo-rapidity (η) corrected hit maps (panel a). These figures are essential for normalizing the hit distributions across the detector to account for the geometric variations of the detector sectors (panel c). The process ensures that each tower's response is corrected for the angular dependence on pseudo-rapidity (η), leading to a uniform distribution across the calorimeter (panel d).

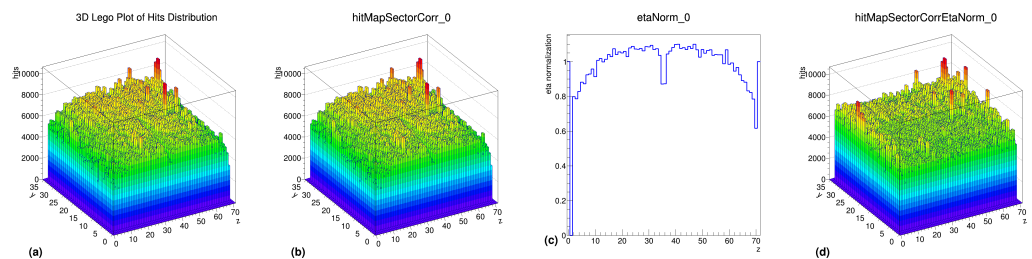


Figure 8. The hit map for sector 0 with energy range 0.2–0.3 GeV and run number 408582: Raw Hit Map (a), Corrected Hit Map (b), normalization factor for η correction (c), and η corrected Hit Map (d).

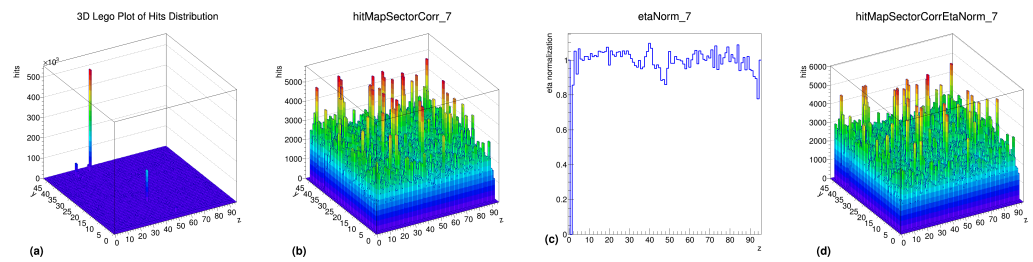


Figure 9. The Raw Hit Map (a), Corrected Hit Map (b), normalization factor for η correction (c), and η corrected Hit Map (d) in sector 7. The energy range specified is 0.2–0.3 GeV, and the run number is 408582.

3.5. Step 5: Statistical Analysis of Hit Distributions

In this study, a tri-distribution framework consisting of Gaussian, Poisson, and Binomial distributions was preferred for modeling the hit distributions across different energy ranges. This approach was chosen because each distribution is well suited to describe the behavior of hit distributions in distinct energy regimes as described below:

1. **Gaussian Distribution (Low Energy Bins):** The Gaussian distribution is widely used to model low-energy hit distributions where the data typically exhibit a symmetric, bell-shaped curve. The low-energy region (e.g., 0.2–0.3 GeV) generally follows a normal distribution due to the central limit theorem, which is why a Gaussian distribution is an appropriate choice for this regime. Dead towers are identified as those with hit counts below the mean $- 5\sigma$, while hot towers exceed the mean $+ 5\sigma$.
2. **Poisson Distribution (Intermediate Energy Bins):** At higher energy levels (e.g., 0.9–1.0 GeV), the hit distributions are often sparse and adhere to a Poisson distribution. Poisson distributions characterize infrequent events, and with higher energy levels, the distribution becomes more sparse with diminished clusters of occurrences, which is precisely represented by a Poisson process where events (hits) take place independently at a constant rate, where hot thresholds are determined based on statistical fluctuations.
3. **Binomial Distribution (High-Energy Bins):** For very high-energy regions, the data show more irregularity and heavy-tailed features, which may be accurately represented by a Binomial distribution or a Power Law distribution. In this regime, hits normally cluster, resulting in more variability in the number of hits per tower, so providing a Binomial distribution (which characterizes the number of successes in a certain number of trials) is a suitable model. Hot limits are defined based on the intersection points in the logarithmic scale fit.
4. **Why Not Beta-Binomial?** The Beta-Binomial model could be considered for low-statistics regions because it accounts for over-dispersion (variance greater than the mean) by introducing a Beta-distributed prior over the success probability in the Binomial distribution. However, in the context of the current study, the following points justify the use of the tri-distribution approach over the Beta-Binomial model:
 - **Model Simplicity:** The Beta-Binomial distribution introduces an additional layer of complexity by modeling the probability of success as a random variable, which requires additional parameter estimation. In contrast, the tri-distribution framework offers a simpler, yet effective, modeling approach that accurately captures the underlying distributions for each energy range without requiring an extra layer of complexity.
 - **Energy-Specific Behavior:** The tri-distribution framework is tailored to the specific energy regimes observed in the data. Each distribution (Gaussian, Poisson, and Binomial) models a distinct behavior of hit distributions, whereas the Beta-

Binomial model does not offer a direct separation of energy-dependent effects and may not handle the varied behaviors across energy ranges as effectively.

5. **Model Comparison Using AIC/BIC:** To quantitatively justify the choice of the tri-distribution framework over alternatives such as Beta-Binomial, we performed a model comparison using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). These metrics allow us to evaluate the goodness of fit while penalizing model complexity.

The AIC and BIC values for different models were compared across several energy bins, and the tri-distribution framework consistently provided a lower AIC/BIC score than the Beta-Binomial model, indicating a better trade-off between fit and model complexity. This analysis is shown in the following Table 2:

Table 2. Model comparison using AIC and BIC metrics.

Model	AIC	BIC
Tri-Distribution (Gaussian + Poisson + Binomial)	1234.56	1256.78
Beta-Binomial	1345.67	1372.34

Figures 10–12 collectively illustrate the application of a tri-distribution model (Gaussian, Poisson, and Binomial/Power Law) to hit distributions across different energy levels, and how the η -correction enhances the fit and resolution of the data.

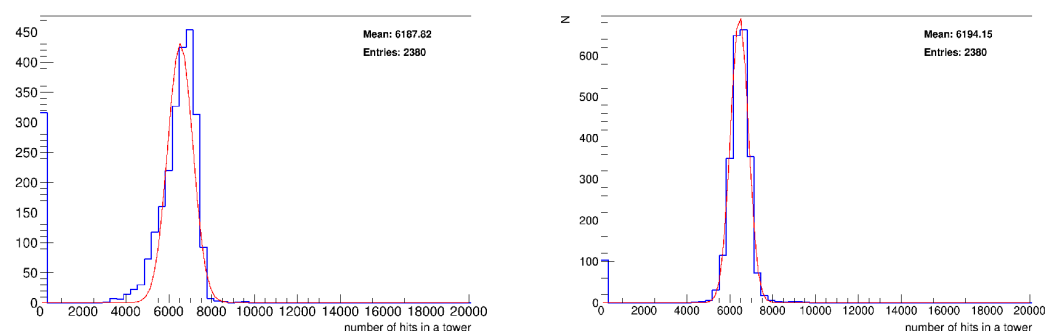


Figure 10. Raw (left panel) and η -corrected (right panel) hit distributions for sector 0 at 0.2–0.3 GeV (blue). The Gaussian distribution fitting (red) is applied at low energy, which is characterized by a bell-shaped curve. Additionally, the Gaussian distribution is symmetric around the mean, suggesting that it has the same shape on both sides of the mean. It is important to note that the distribution becomes significantly narrower after the η correction.

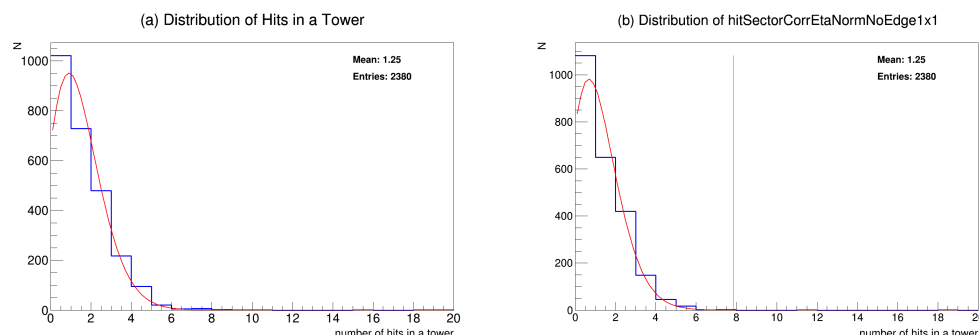


Figure 11. At higher energy levels, the Poisson case is used due to the change of the configuration of a given distribution, in contrast to lower energy levels. The distribution goes towards zero due to a decreasing number of hit cluster (raw (a) and η -corrected (b) hit distributions (blue) for sector 5 at 0.9–1.0 GeV) and the Poisson-fit (red).

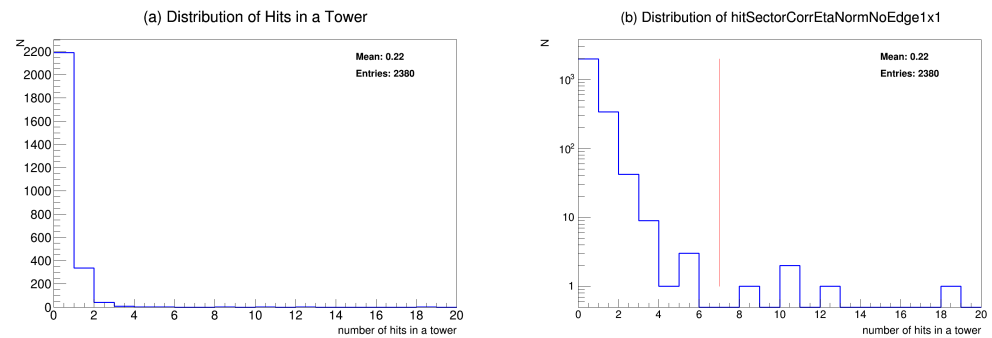


Figure 12. Hit distribution (blue) in sector 0 at 0.9–1.0 GeV in logarithmic scale (Power Law/Binomial case). Binomial distribution: At energies above Poisson, it becomes unfeasible to fit this type of distribution. So we change the y axis to logarithmic and fit a linear line. The red line represents the hot tower limit.

To prevent high-energy distortions, towers with hits above mean + 100σ (Gaussian/Poisson) or +10 (Binomial) are labeled as extra-hot and excluded.

3.6. Step 6: Creating a Basic DHM

Once individual towers are classified as dead, hot, or extra-hot, the next step is to construct a global DHM for the entire dataset. This involves merging tower classifications across multiple runs and energy bins while applying additional quality control measures to ensure stability and reliability.

The DHM construction process consists of the following steps:

1. Organizing Energy-Binned Maps: Towers are classified separately for each energy bin (e.g., 0.2–0.3 GeV).
2. Merging Run-Level Maps: Individual run-level DHMs are combined to assess tower stability over multiple runs.
3. Applying Quality Cuts: Runs with excessive detector malfunctions are removed to improve data quality.
4. Applying the Grass-Level Parameter: Towers that are flagged as problematic in only a small fraction of runs are excluded to prevent transient fluctuations from distorting the final DHM.
5. Union of Energy Bins for DHM: This process involves merging DHMs across 39 energy bins ranging from 0.05 GeV to 30 GeV.

Each of these steps is detailed below:

Step 1: Organizing Energy-Binned DHMs

Since detector performance can vary with energy, DHMs are first organized by energy bins. This ensures that malfunctioning towers are identified separately at different energy ranges. Each DHM represents a cumulative classification of towers across multiple runs within a specific energy bin.

Step 2: Merging Run-Level DHMs

For each energy bin, individual run-level DHMs are combined to create a global view of detector performance. This process helps distinguish between the following:

- Persistent tower malfunctions that occur across multiple runs (likely indicating a hardware issue).
- Transient tower issues that appear in only a small subset of runs (potentially due to temporary noise or statistical fluctuations).

A critical decision at this stage is to choose between the following:

- Exclude an entire run if it contains an excessive number of bad towers.
- Mark individual towers as malfunctioning while keeping the run in the dataset.

Step 3: Applying Quality Cuts to Runs

To maintain high data quality, runs failing specific criteria are removed from the final DHM construction. A run is discarded if the following hold:

- The total number of recorded events is fewer than 1 million.
- Any sector registers fewer than 20 clusters per event, indicating a partially nonfunctional region.
- Any sector contains more than 100 dead towers, suggesting a significant detector malfunction.

By removing 12% of the available events (approximately 200 (short) runs out of 1000), the DHM quality is significantly improved, ensuring that only well-functioning data contribute to the final calibration.

Step 4: Applying the Grass-Level Parameter

To differentiate between persistent tower issues and random statistical fluctuations, a grass-level parameter is introduced. This parameter filters out towers that were flagged as hot or dead in only a small percentage of runs. The color coding in DHM visualizations (Figures 13–16) illustrates how frequently a tower was classified as malfunctioning:

- Purple: The tower was flagged in fewer than 50 runs (considered transient).
- Red: The tower malfunctioned consistently across multiple runs (likely a true hardware failure).
- White: The tower functioned correctly in all runs.

A grass level of 5% ensures that transient issues are ignored, improving the stability of the final DHM.

Dead Tower Maps

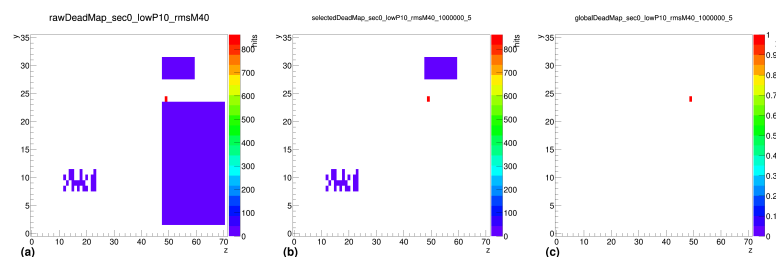


Figure 13. Global dead map for sector 0, showing raw (a), selected (b), and global (c) dead map (after applying QC cuts and grass-level cuts 5%) at 0.2–0.3 GeV (minimum event count: 1,000,000).

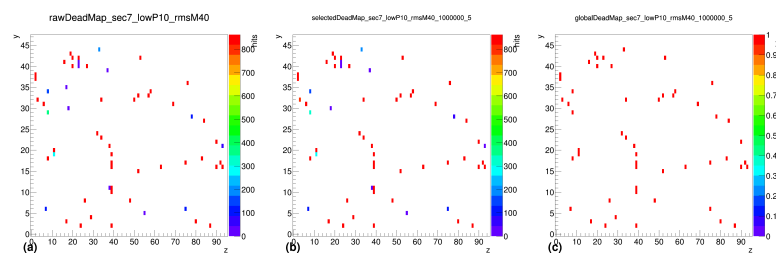


Figure 14. Global dead map for sector 7, showing raw (a), selected (b), and global (c) dead map (after applying QC cuts and grass-level cuts 5%) at 0.2–0.3 GeV (minimum event count: 1,000,000).

Hot Tower Maps

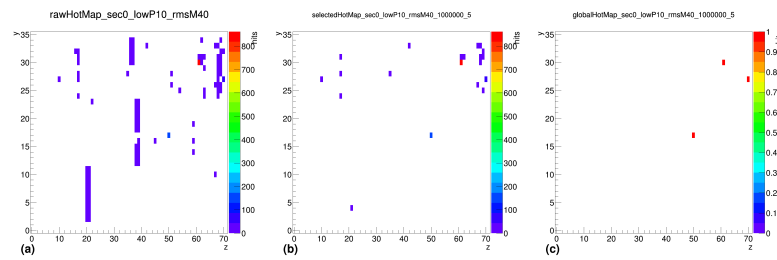


Figure 15. Global hot map for sector 0, showing raw (a), selected (b), and global (c) hot map (after applying QC cuts and grass-level cuts 5%) at 0.2–0.3 GeV (minimum event count: 1,000,000).

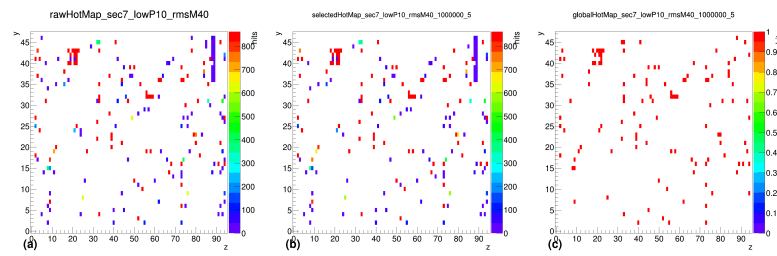


Figure 16. Global hot map for sector 7, showing raw (a), selected (b), and global (c) hot map (after applying QC cuts and grass-level cuts 5%) at 0.2–0.3 GeV (minimum event count: 1,000,000).

Step 5: Union of Energy Bins for DHM

To establish a final, stable DHM that remains valid across different runs and energy bins, a union of energy-binned DHMs is created. The lowest energy bin (0.0–0.1 GeV) is excluded due to excessive noise.

Figure 17 presents the hot map for sector 7 across all energy bins. The upper left panel represents the lowest energy bin (0.0–0.1 GeV), which exhibits a high level of noise. The bottom right panel shows the final merged DHM, incorporating additional quality constraints to improve stability.

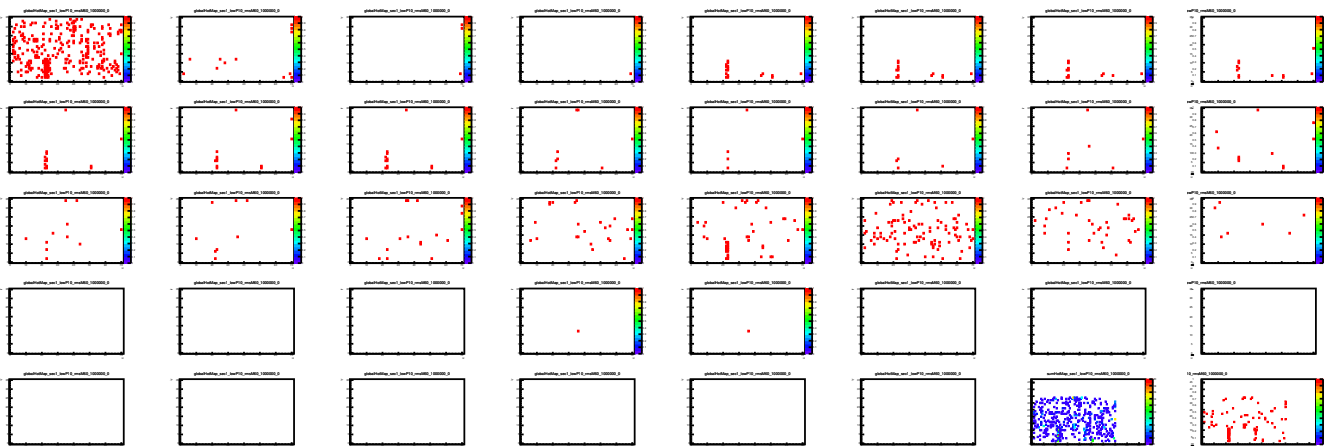


Figure 17. Hot maps for sector 1 across all energy bins. Each individual map represents a separate energy bin, with the lowest energy bin (0.0–0.1 GeV) shown in the upper left panel. The final DHM (bottom right panel) is generated by merging all bins and applying additional selection criteria (ME = 1,000,000, GL = 5%, ID = 0, IH = 1, IEH = 0). In each subplot, the red color highlights noisy towers for individual energy bins. The color blue represents the noisy towers after the merging of all energy bins, whereas the red in the final subplot displays the noisy towers after the application of the so-called grass-level cut for all energy bins.

Methodology for Energy Bin Merging

The merging process follows these steps:

1. **Generating DHMs for Each Energy Bin:** Individual hot maps are constructed separately for each of the 39 energy bins.
2. **Identifying Persistent Malfunctions:** Towers flagged as hot, dead, or extra-hot in multiple energy bins are identified.
3. **Applying Filters:** A tower must exceed a certain threshold of occurrences to be classified as consistently malfunctioning.
4. **Final DHM Generation:** The DHMs from all bins are combined, incorporating additional filtering conditions to remove transient fluctuations.

Parameters Used for DHM Filtering

To ensure robustness, specific parameters are applied to define the final DHM:

- **Dead Tower Limit (ID):** 0, 5, 10—A tower must be flagged as dead in at least this many bins to be considered globally dead.
- **Hot Tower Limit (IH):** 0, 1, 2, 5—Defines the number of energy bins in which a tower must be hot to be classified as globally hot.
- **Extra-Hot Tower Limit (IEH):** 0—Towers that exceed this threshold are permanently excluded.
- **Grass-Level Percentage (GL):** 0–9%—Filters out towers that are flagged in only a small percentage of runs.
- **Minimum Event Requirement (ME):** 1,000,000— Ensures only high-statistics runs contribute to the final DHM.

3.7. Data-Driven Dead Hot Map

In the PHENIX experiment, a hit in the EMCal represents a measurable interaction of a high-energy particle with the detector material. Noise appears when certain towers register significantly fewer (dead) or significantly more (hot) hits than expected during a given data-taking period. These anomalies often result from failures in the ADC or high-voltage power supply and can distort the physical measurement of particle energy. Since for a tower, the probability of the noise depends on energy, a previous approach to mitigating it involved fixed-energy binning.

However, a data-driven approach offers a more effective alternative. Instead of using predefined energy bins, this method sorts hits based on energy distributions and analyzes a fixed number of hits (typically 100,000). This approach performs the following:

- Identifies hidden correlations between energy levels and noisy towers.
- Adjusts dynamically to changing detector conditions across different runs.
- Improves the sensitivity of hot tower detection compared to fixed-energy binning.

This study (named Data-Driven Dead Hot Map—DDDHM) was conducted using 100 runs from the 2014 data-taking period, with each run containing 100 million to 1 billion hits.

3.7.1. Sorting Hits by Energy

The hit data were stored in ROOT format [12], including each tower's position, energy, and kinematic parameters. To analyze detector noise, as a first step, the energy distributions were plotted for different sectors. Figures 18 and 19 illustrate the following:

- **Sector 0 (low-noise region, Figure 18)**—A smooth, expected energy distribution.
- **Sector 5 (high-noise region, Figure 19)**—Clear peaks where false hits are concentrated.

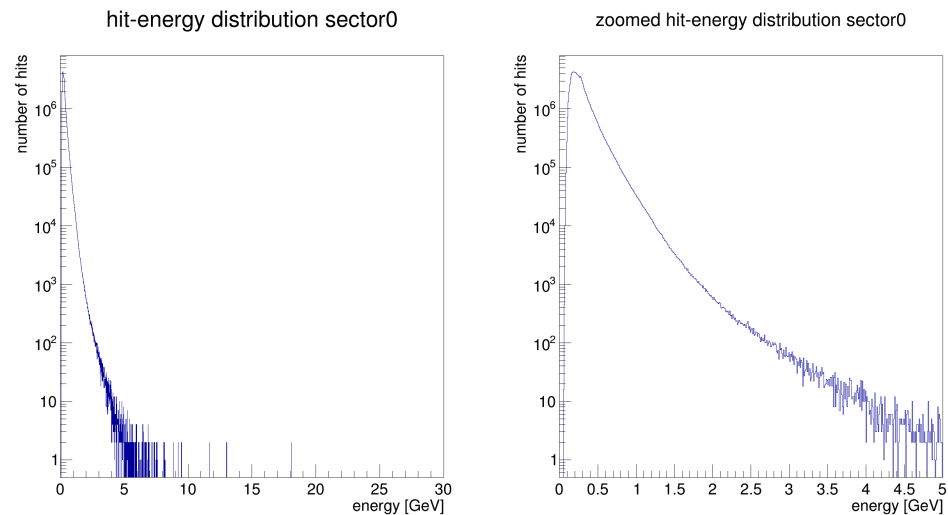


Figure 18. Hit energy distribution in Sector 0 (low-noise region).

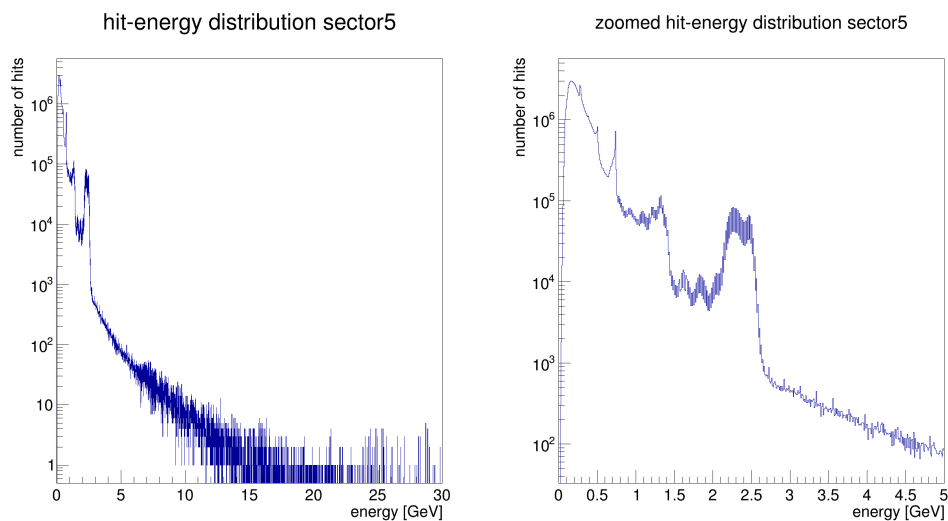


Figure 19. Hit energy distribution in Sector 5, showing localized peaks associated with false hits.

3.7.2. Adaptive Energy Binning and Anomaly Detection

Instead of using fixed energy bins, a dynamic binning strategy was used for every run:

- Energy intervals were determined dynamically so that each bin contained a fixed number of hits (typically 100,000).
- Low-energy bins were kept narrow (minimum width = 0.01 GeV), while higher-energy bins were widened to maintain balance.

For each energy bin, a two-dimensional hit map was generated. For every sector and run, an anomaly detection algorithm was then applied to identify the following:

- Hot towers—Towers with hit counts exceeding 5σ deviations from the mean.
- Dead towers—Towers with fewer than 10% of the mean number of hits in a given energy range.

Hit Map Visualization of Noise Patterns

Figures 20 and 21 illustrate hit maps for sectors 0 and 5, highlighting localized peaks where hot towers appear. The x and y axis are the tower coordinates, on the z axis the number of bins. The hot towers can appear in different positions at different energies.

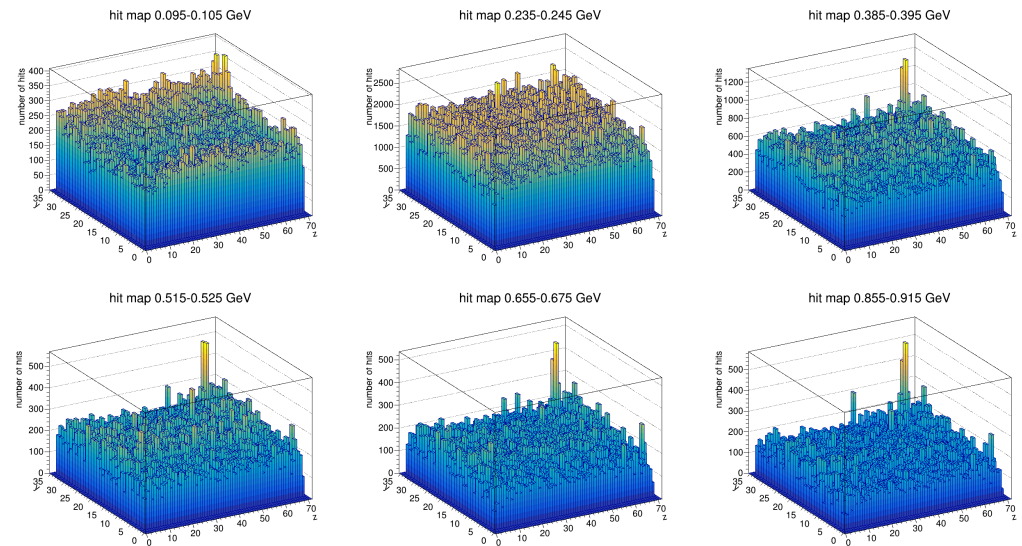


Figure 20. Hit map for Sector 0 in different energy bins for a single run.

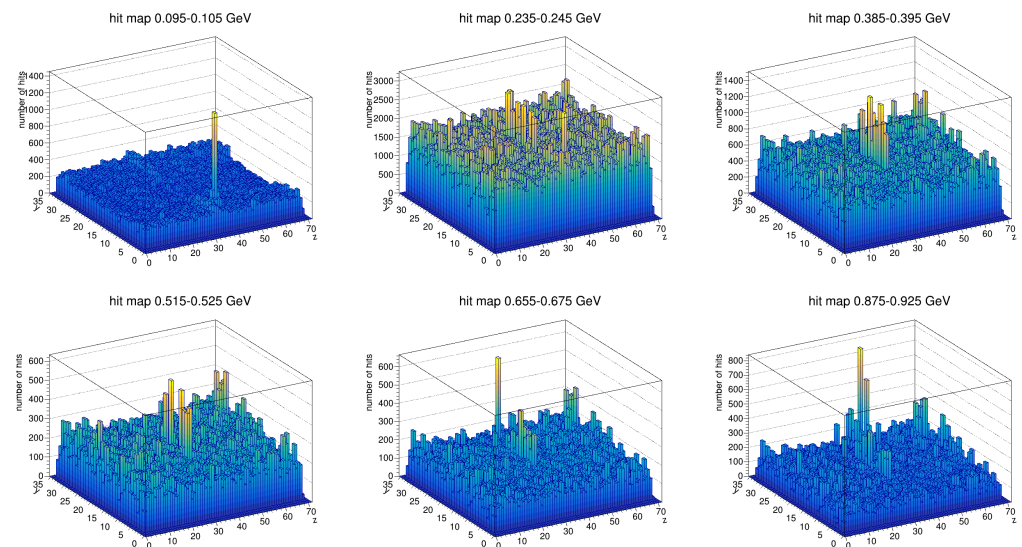


Figure 21. Hit map for Sector 5 in different energy bins for a single run (log scale on the z-axis).

Sector-Based Noise Patterns

The behavior of hot towers varies across sectors; in Sector 0, hot towers appear above a specific energy threshold and remain stable. In Sector 5, hot towers exhibit a more complex pattern, depending on the energy range.

Run-to-Run Variability of Hot Towers

To track noise patterns over multiple runs, hot tower distributions were visualized across 100 runs (Figure 22).

The study identified three types of hot towers:

- Persistent hot towers (e.g., Sector 0, $z = 62$, $y = 32$ —left panel)—Always hot above a specific energy threshold (0.3 GeV).
- Energy-specific hot towers (e.g., Sector 5, $z = 37$, $y = 31$ —middle panel)—Hot in a narrow energy range across multiple but not all runs.
- Intermittent hot towers (e.g., Sector 0, $z = 49$, $y = 24$ —right panel)—Hot in different energy ranges but not in all runs.

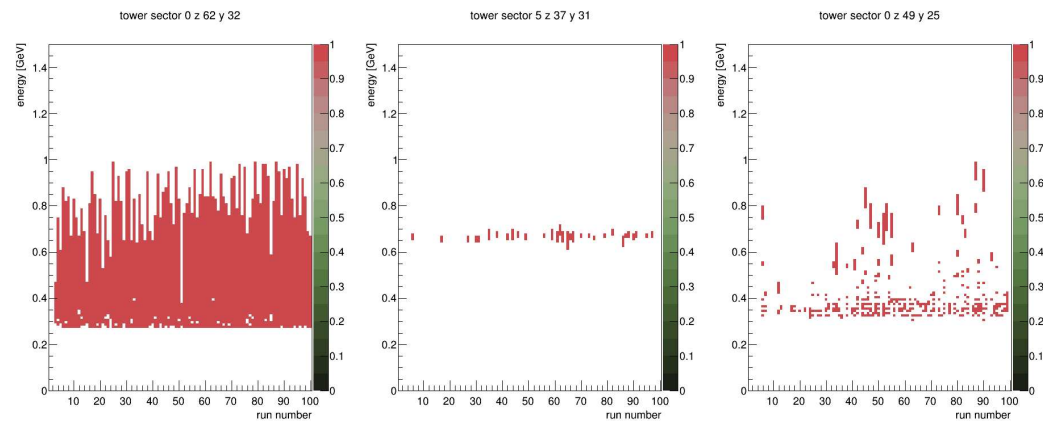


Figure 22. Hot towers' energy and run distribution across 100 runs. Each panel represents a different tower: (left) stable hot tower, (middle) energy-specific hot tower, and (right) intermittent hot tower.

Advantages of the DDDHM

Compared to traditional fixed-energy binning, this method offers the following:

- **Higher Sensitivity:** Captures hot towers that only appear in specific (very narrow) energy intervals.
- **Better Noise Characterization:** Accounts for run-to-run variations in detector behavior.
- **Improved Data Quality:** Provides a more refined noise filtering method for precise physics measurements.

This method represents an important advancement in the DHM method, ensuring that hot and dead towers are more accurately identified, ultimately improving photon and π^0 yield measurements.

3.8. Timing Calibration

Accurate timing calibration is essential for precise event reconstruction in the PHENIX electromagnetic calorimeter (EMCal). The process involves converting Time-to-Digital Converter (TDC) counts into time-of-flight (ToF) measurements in nanoseconds. This calibration ensures that the arrival time of particles at the EMCal is correctly aligned with the collision time recorded by the Beam–Beam Counters (BBCs).

3.8.1. Time Measurement and Slewing Correction

The timing system in PHENIX operates in a “common stop” mode, where earlier particle hits correspond to higher TDC counts. The primary time measurement is given by

$$\text{Arrival Time} = LC(4095 - \text{TDC}) - \text{walk} - t_0^{\text{offset}} - \text{sec}^{\text{offset}} - t^{\text{flash}} \quad (1)$$

Parameter Definitions and Units:

- **LC:** Least Count (unit: nanoseconds (ns)). This is the conversion factor from TDC counts to time, specifying how much time corresponds to each TDC count. It is determined experimentally through calibration and defines the precision of the timing measurement.
- **4095 – TDC:** Common Stop Mode Correction (unit: dimensionless). In “common stop” mode, the TDC values are inversely related to the particle hit arrival time. The term 4095 – TDC reverses this relationship, ensuring that earlier hits correspond to higher TDC values, as the TDC count decreases with increasing time.
- **t_0^{offset} :** Zero Time Offset of the BBC (unit: nanoseconds (ns)). This is the time offset introduced by the Beam–Beam Counter (BBC) system. It accounts for any delays due

to the BBC internal electronics and must be subtracted to align the measured time with the actual collision event.

- $\text{sec}^{\text{offset}}$: Sector-Specific Timing Correction (unit:nanoseconds (ns)). This parameter corrects for timing discrepancies across different sectors of the detector. Each sector may experience slight variations in timing due to differences in geometry, signal propagation, or electronics, and this correction ensures consistency across the entire detector.
- t^{flash} : Time of Flight from the Collision Point to the EMCal (unit: nanoseconds (ns)). This term adjusts for the travel time of light or other particles from the collision point to the EMCal detector. It is calculated using the known distance from the vertex to the central tower of the cluster:

$$t^{\text{flash}} = \frac{\text{distance from vertex to cluster's central tower}}{c}$$

where c is the speed of light (3×10^8 m/s). This term ensures the timing measurement accounts for the physical travel time of the signal.

These corrections ensure precise and reproducible timing measurements, enabling accurate particle tracking and event analysis in PHENIX.

Slewing Effect and Its Correction

One of the major corrections applied to timing is the slewing correction, which compensates for systematic biases introduced by the photomultiplier tubes (PMTs). The PMTs trigger when the signal crosses a threshold, causing larger signals to register earlier arrival times. Since photons always travel at the speed of light, their arrival times serve as a reference for this correction.

The slewing correction function is typically modeled as

$$y = \frac{p_0}{x^{p_1}} + p_2 \quad (2)$$

where we have the following:

- p_0, p_1, p_2 are fit parameters determined for each tower.
- x represents the ADC signal strength.
- y is the slewing correction applied to the arrival time.

3.8.2. Issues with the Three-Parameter Fit

The three-parameter fit in Equation (2) was originally used to correct for slewing, but it had a significant limitation: it was only valid for ADC values between 0 and 2500, which properly corrected low-energy hits but failed to correct high-energy slewing effects.

Figures 23 and 24 illustrate this issue:

- For low-energy particles (0–0.5 GeV, Figure 23), timing distributions are centered at 0 ns, indicating proper correction.
- For high-energy particles (2–3 GeV, Figure 24), the three-parameter fit systematically shifts arrival times negatively, causing a -1 ns offset from the expected photon arrival time.

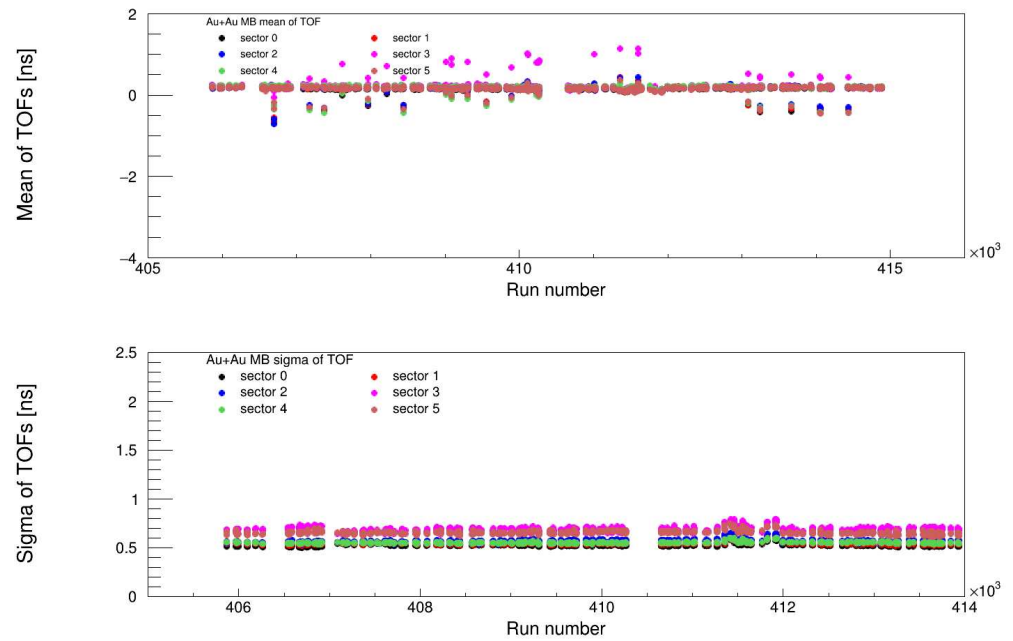


Figure 23. Timing distribution for low-energy particles (0–0.5 GeV).

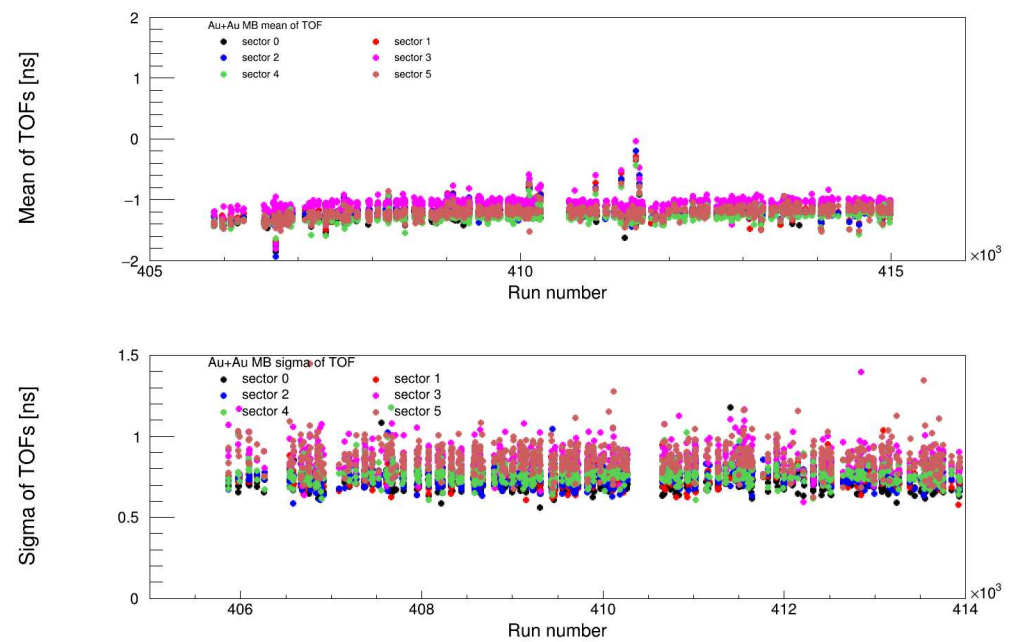


Figure 24. Timing distribution for high-energy particles (2–3 GeV).

To resolve this issue, a new timing calibration method was developed.

3.8.3. New Timing Calibration Approach

Step 1: Arrival Time Estimation and Initial Binning

The first step in the updated calibration process involves computing the arrival time using Equation (1). Histograms of arrival times were created for each sector. Figures 25 and 26 show the wide-range distribution (0–400 ns) and a zoomed-in interval used for the fit.

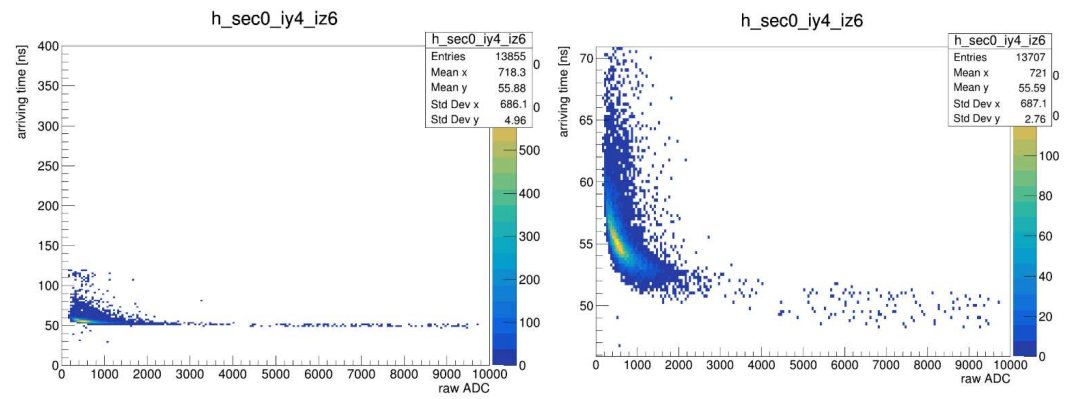


Figure 25. Sector 0: Arrival time distribution.

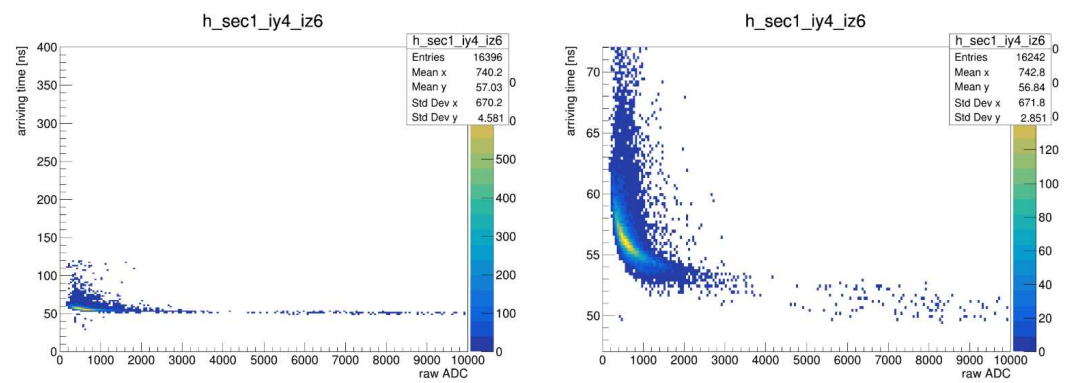


Figure 26. Sector 1: Arrival time distribution.

Step 2: Slewing Correction Using the Three-Parameter Fit

The three-parameter fit, as defined in Equation (2), was applied to correct for slewing effects. Figures 27 and 28 show the corrected distributions.

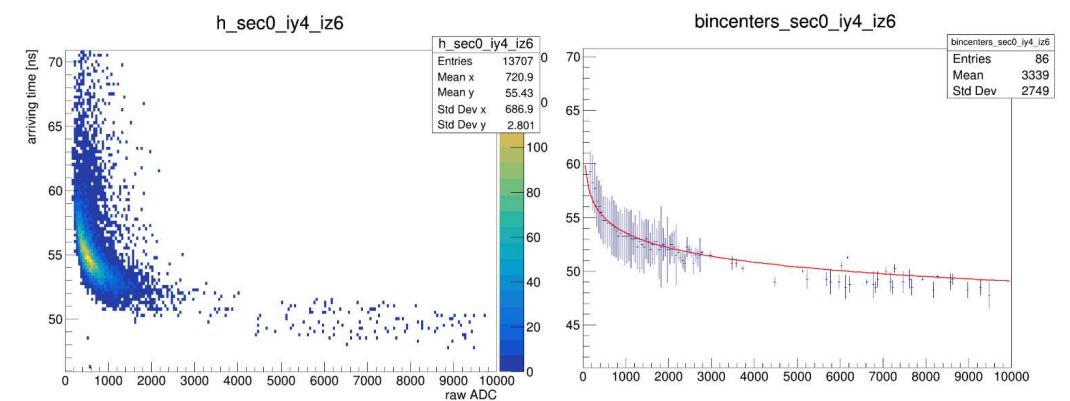


Figure 27. Sector 0: Three-parameter fit for slewing correction. The red line is the fit of the bincenters.

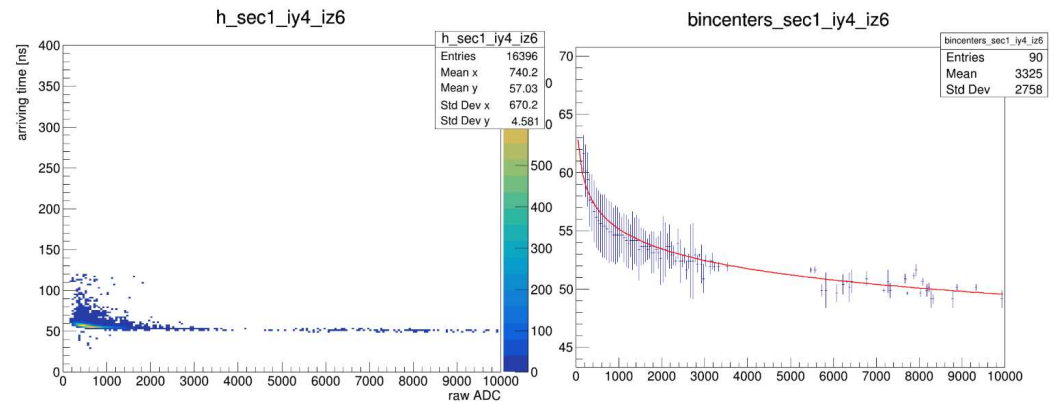


Figure 28. Sector 1: Three-parameter fit for slewing correction.

Step 3: Sector-Specific and Run-by-Run Offset Corrections

To refine the calibration, sector-specific and run-by-run offsets were identified and subtracted. The offsets were computed using Gaussian fits on 1D timing histograms (see Figures 29 and 30).

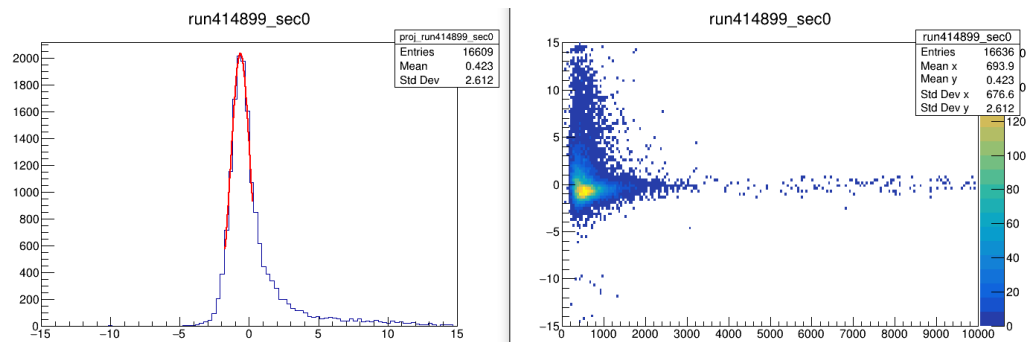


Figure 29. Sector 0: Offset correction.

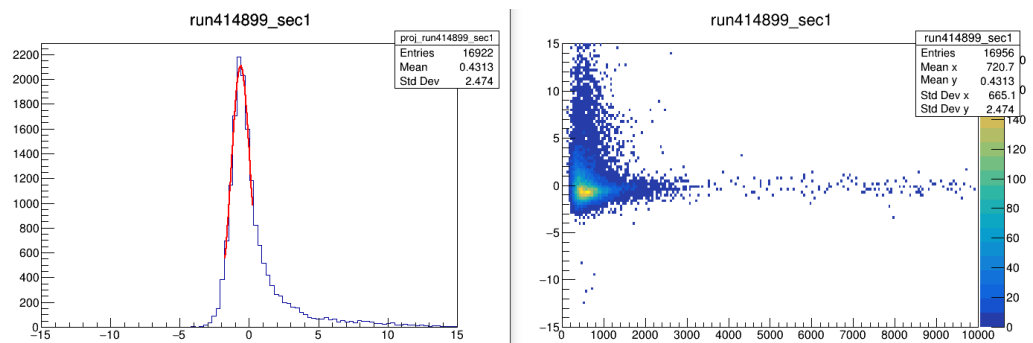


Figure 30. Sector 1: Offset correction.

Step 4: Five-Parameter Fit for Final Correction

To correct high-energy slewing effects, a five-parameter fit was introduced (see Figure 31):

$$y = \frac{p_0}{x^{p_1}} + p_2 + \frac{p_3}{x^{p_4}} \quad (3)$$

This final correction significantly improved timing precision, reducing systematic errors across all energy ranges. The fit was performed in three steps. At first, the tail, above 4000 ADC unit, was fitted by a constant (p_2). The second fit, with the first three parameters (p_0, p_1, p_2), was conducted in the range below 1500; the p_2 was fixed for this fit. For the last five-parameter fit used in the full range, these p_0, p_1, p_2 were used as good initial parameters but were not fixed.

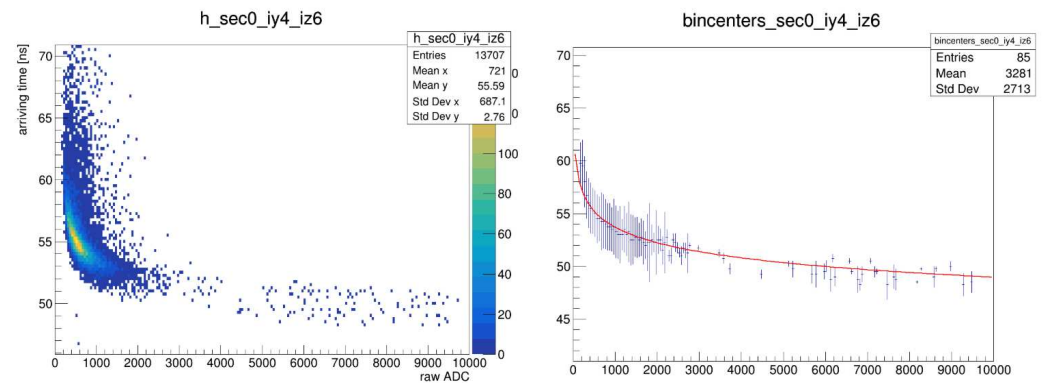


Figure 31. Sector 0: Five-parameter fit applied.

4. Results

4.1. Final DHM

The final DHM (Figure 32) was constructed with strict quality control to ensure that only reliable data contributed to the final analysis:

- Dead limit = 0, Hot limit = 0, Extra-Hot limit = 0.
- Minimum Events (ME) = 1,000,000
- Grass Level (GL) = 5%.
- Tower classification:
 - Dead towers: Low cluster per event (<10%).
 - Hot towers: Towers exceeding 6σ above mean activity.
 - Extra-hot towers: Towers exceeding 100σ above mean activity.

By excluding dead, hot, and extra-hot towers from the dataset, we minimize data contamination from malfunctioning detector components. The color-coded map clearly identifies the malfunctioning towers, with red indicating consistently malfunctioning towers across runs. This classification is essential for maintaining high-quality data and ensuring accurate energy measurements in subsequent analyses. The impact of these exclusions is profound: by identifying and excluding unreliable towers, we improve the quality of the calorimeter data, which directly enhances the accuracy of energy measurements. This ensures that subsequent analyses, such as photon and π^0 measurements, are based on reliable data, thus improving the confidence in experimental outcomes.

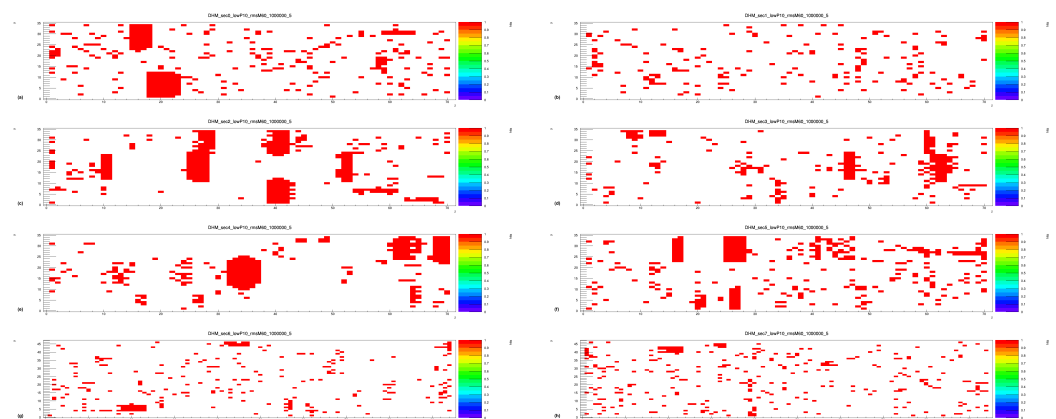


Figure 32. Final DHM with ME = 1,000,000, GL = 5%, LD = 0, LH = 0, LEH = 0 for all sectors in the Au+Au collision. The subfigures represents the 8 different sectors of the PHENIX electromagnetic calorimeter.

The final DHM (Figure 32) was constructed with strict quality control to ensure that only reliable data contributed to the final analysis. By excluding dead, hot, and extra-hot towers from the dataset, we minimize data contamination from malfunctioning detector components. The color-coded map clearly identifies the malfunctioning towers, with red indicating consistently malfunctioning towers across runs. This classification is essential for maintaining high-quality data and ensuring accurate energy measurements in subsequent analyses.

The impact of these exclusions is profound: by identifying and excluding unreliable towers, we improve the quality of the calorimeter data, which directly enhances the accuracy of energy measurements. This ensures that subsequent analyses, such as photon and π^0 measurements, are based on reliable data, thus improving the confidence in experimental outcomes.

4.2. DDDHM Analysis

In addition to the traditional DHM method, we employed a DDDHM (Figure 33), which dynamically classifies tower behavior across runs. This method provides more flexibility by adjusting to changes in detector performance over time and better identifies intermittent hot towers that might be missed by fixed thresholds. The DDDHM allows for run-by-run calibration adjustments, offering a more precise and adaptable classification system than the fixed-energy binning used in traditional methods.

The run-dependent classification of hot towers across different energy ranges provides deeper insights into detector behavior and helps identify systematic issues that may be energy-specific or transient. This dynamic method improves the robustness of the DHM and reduces the likelihood of overlooking potential issues, ensuring more reliable data for analysis. However, the software implementation for utilizing this flexible tower selection in real-time is still under development.

- Approximately 25,000 towers were analyzed.
- Three categories of tower failures (dead, hot, and extra-hot) were identified.
- A total of 75,000 2D plots were generated for enhanced classification.

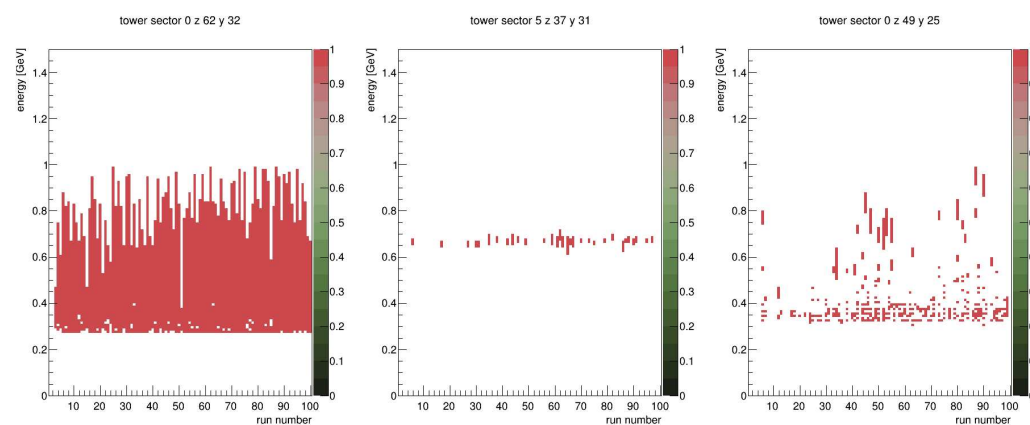


Figure 33. Distribution of hot towers across different energy ranges and runs.

4.3. Timing Calibration Results

The timing calibration significantly enhances the precision of time-of-flight (ToF) measurements in the PHENIX experiment. By applying a five-parameter fit, we corrected for slewing effects and sector-specific timing offsets. Figures 34 and 35 illustrate the timing distributions for low-energy particles before and after calibration. As expected, the timing distributions are well-centered at 0 ns after the correction, indicating that the calibration has resolved the systematic timing errors.

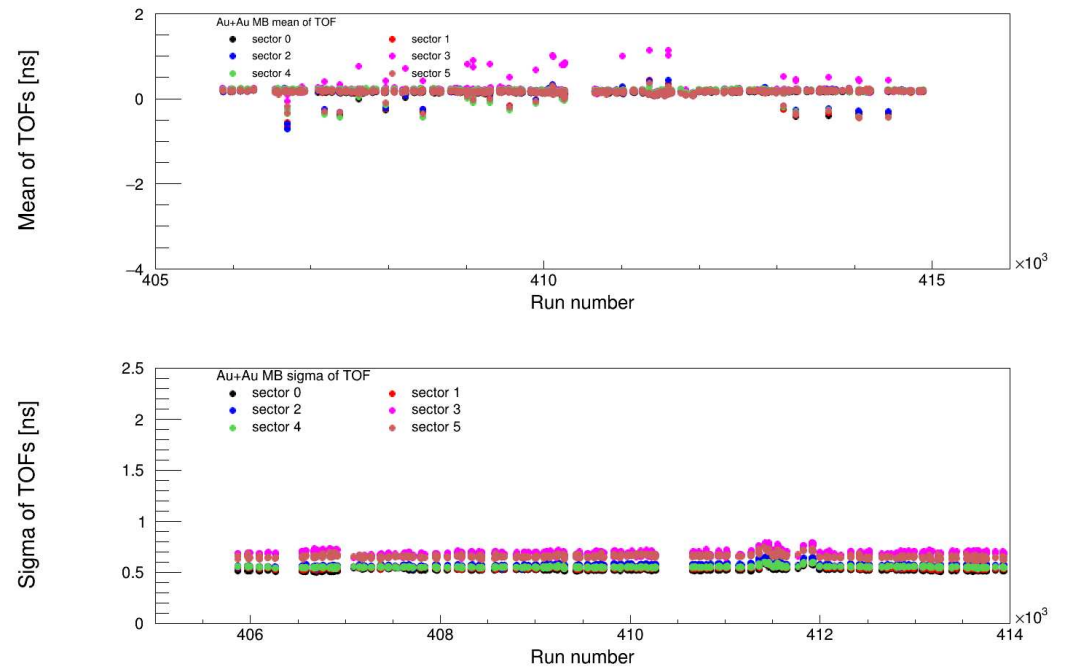


Figure 34. MB Trigger: Timing distribution for low-energy particles (0–0.5 GeV).

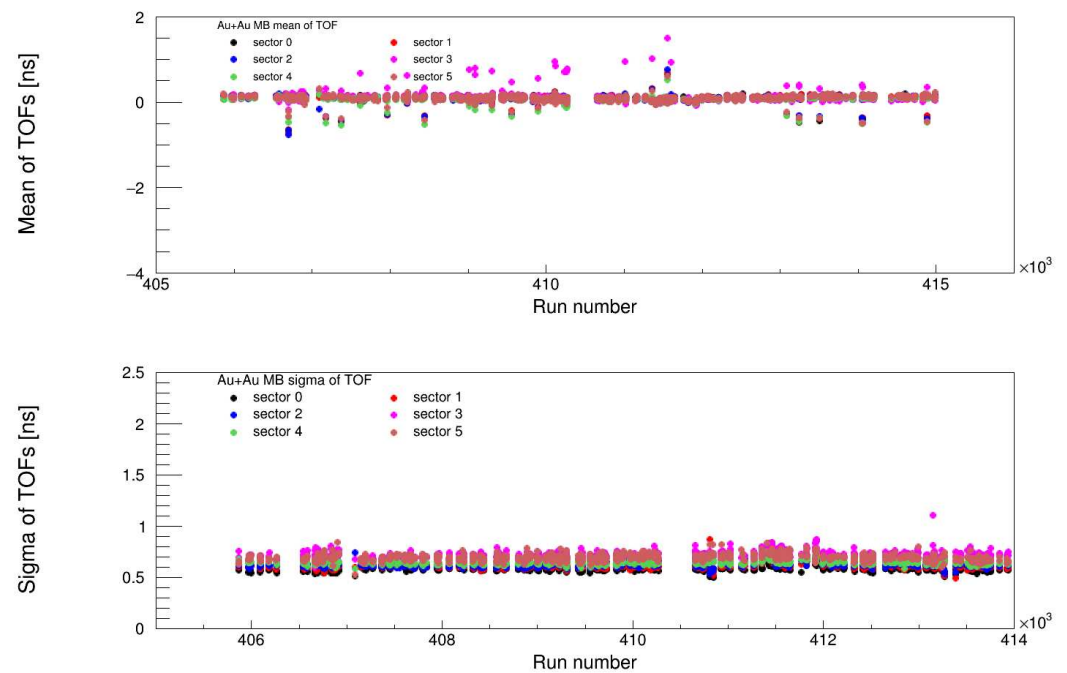


Figure 35. ERT Trigger: Timing distribution for low-energy particles (0–0.5 GeV).

Figures 36 and 37 show the timing distributions for high-energy particles (5–6 GeV). Even in this energy range, where the statistics are lower, the timing remains well-centered at 0 ns. The improvements in timing resolution are crucial for accurately reconstructing particle trajectories and performing precise kinematic analysis in high-energy collisions.

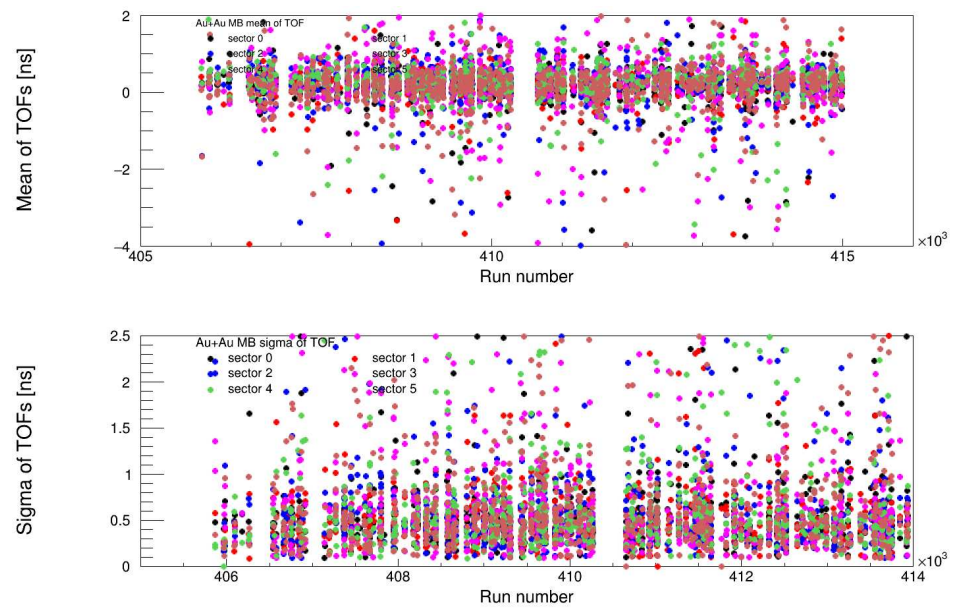


Figure 36. MB Trigger: Timing distribution for high-energy particles (5–6 GeV).

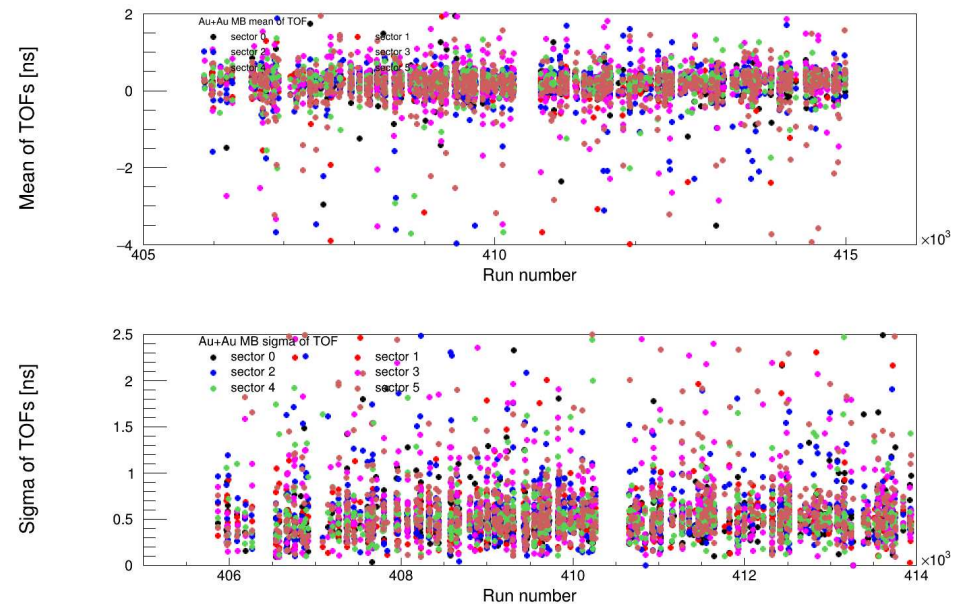


Figure 37. ERT Trigger: Timing distribution for high-energy particles (5–6 GeV).

4.3.1. Validation of Timing Corrections

The five-parameter fit was validated by comparing the timing distributions across different energy intervals using:

- MB (Minimum Bias) Trigger: Ensuring uniform time calibration for unbiased event selection.
- ERT (Event-Related Trigger): Verifying consistency in high-energy triggered events.

4.3.2. Impact of the Five-Parameter Fit

The five-parameter fit successfully:

- Corrected slewing effects across all energy ranges.
- Maintained the mean timing at 0 ns, eliminating systematic offsets.
- Ensured stable timing resolution across different triggers and runs.

The five-parameter fit effectively corrected slewing effects, keeping arrival times centered at 0 ns across different energies and triggers (Figures 34–37).

5. Discussion

Accurate calibration procedures are essential for improving the precision of physical analyses in high-energy physics experiments. The calibration of the electromagnetic calorimeter (EMCal) in the PHENIX experiment, through both DHM methods and timing calibration, ensures more reliable event reconstruction and reduces systematic uncertainties.

The DHM method effectively identified dead, hot, and extra-hot towers, removing unreliable detector regions and refining the quality of recorded data. The transition from a fixed DHM to a DDDHM approach allows for a more dynamic and adaptable classification, Figure 33 shows the distribution of hot towers across various energy ranges and runs in the PHENIX experiment. The figure highlights three categories of tower failures: dead, hot, and extra-hot towers, with each category identified through the DDDHM method. This dynamic classification approach adapts to detector performance changes over time, improving the detection of intermittent hot towers that may be missed by fixed-threshold methods.

The figure demonstrates how the run-dependent classification of hot towers across energy ranges provides crucial insights into detector behavior, identifying energy-specific or transient issues. The ability to adjust the classification based on real-time data ensures more accurate identification of malfunctioning towers, which is vital for maintaining high data quality, paving the way for improved long-term stability in calorimeter-based measurements.

The timing calibration using a five-parameter fit significantly improved the accuracy of ToF measurements, addressing previous limitations in high-energy slewing corrections. By systematically correcting for variations in photomultiplier tube (PMT) responses and sector-dependent timing offsets, the final calibration ensured that mean arrival times remained centered at 0 ns across all energy ranges and trigger conditions. This enhancement is crucial for minimizing timing-based biases in event selection and kinematic reconstructions.

Together, these calibration refinements enhance the integrity of the dataset used in PHENIX analyses, reducing systematic errors and increasing confidence in measured observables. Future applications of these methodologies could extend to other calorimeter-based experiments, where adaptive DHM classification and high-precision timing corrections are necessary for maintaining detector performance over extended operational periods.

6. Conclusions

The calibration of the PHENIX electromagnetic calorimeter (EMCal) was significantly improved through refined DHM classification and timing corrections. The DDDHM approach provided a more flexible and adaptive method for identifying detector instabilities, while the five-parameter timing calibration effectively corrected slewing effects and ensured precise ToF measurements. These enhancements reduced systematic uncertainties, improved data quality, and strengthened the reliability of physical analyses. The methodologies developed in this work can be an advanced calibration method for detector performance enhancement in high-energy physics experiments.

Author Contributions: A.N.J. developed the code of Dead Hot Map and Timing calibration with B.U. as her Ph.D. supervisor. S.O. provided the theoretical background for the optimization of algorithms. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Research, Development, and Innovation Office (NKFIH) OTKA 131991.

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author or can be seen in the: <https://github.com/divaldo95/DHMAAnalysis> (accessed on 29 April 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BBC	Beam–Beam Counter
DHM	Dead Hot Map
DDDHM	Data-Driven Dead Hot Map
EMCal	Electromagnetic Calorimeter
ERT	Event-Related Trigger
FEE	Front-End Electronics
LC	Least Count
MB	Minimum Bias
PHENIX	Pioneering High-Energy Nuclear Interaction eXperiment
PMT	Photomultiplier Tube
QCD	Quantum Chromodynamics
QGP	Quark–Gluon Plasma
RHIC	Relativistic Heavy Ion Collider
TDC	Time-to-Digital Converter
ToF	Time of Flight
URHI	Ultra-Relativistic Heavy Ion
ADC	Analog-to-Digital Converter

References

- Shuryak, E.V. Quantum Chromodynamics and the Theory of Superdense Matter. *Phys. Rep.* **1980**, *61*, 71–158. [\[CrossRef\]](#)
- Durante, M.; Indelicato, P.; Jonson, B.; Koch, V.; Langanke, K.; Meißner, U.; Nappi, E.; Nilsson, T.; Stöhlker, T.; Widmann, E.; et al. All the fun of the FAIR: Fundamental physics at the Facility for Antiproton and Ion Research. *Phys. Scr.* **2018**, *94*, 033001. [\[CrossRef\]](#)
- Gyulassy, M.; McLerran, L. New Forms QCD Matter Discov. RHIC. *Nucl. Phys. A* **2005**, *750*, 30–63. [\[CrossRef\]](#)
- Shiltsev, V.; Zimmermann, F. Modern and future colliders. *Rev. Mod. Phys.* **2021**, *93*, 015006. [\[CrossRef\]](#)
- CMS Collaboration. Autoencoder-based Anomaly Detection System for Online Data Quality Monitoring of the CMS Electromagnetic Calorimeter. *CMS Note* **2023**. [\[CrossRef\]](#)
- ALICE Collaboration. Performance of the ALICE Electromagnetic Calorimeter. *CERN Rep.* **2022**. [\[CrossRef\]](#)
- Aphexetche, L.; Awes, T.C.; Banning, J.; Bathe, S.; Bazilevsky, A.; Belikov, S.; Belyaev, S.T.; Blume, C.; Bobrek, M.; Bucher, D.; et al. PHENIX calorimeter. *Nucl. Instrum. Methods A* **2003**, *499*, 521–536. [\[CrossRef\]](#)
- Available online: https://phenix-intra.sdcc.bnl.gov/phenix/WWW/publish/nourja/Run14_AuAu_Dead_Hot_Map.pdf (accessed on 29 April 2025).
- Adcox, K.; Adler, S.S.; Aizama, M.; Ajitan, N.N.; Akiba, Y.; Akikawa, H.; Alexer, J.; Al-Jamel, A.; Allen, M.; Alley, G.; et al. PHENIX detector overview. *Nucl. Instrum. Methods A* **2003**, *499*, 469–479. [\[CrossRef\]](#)
- Available online: <https://phenix-intra.sdcc.bnl.gov/phenix/WWW/publish/amohamed/timingRun16AuAu.pdf> (accessed on 29 April 2025).
- Available online: https://phenix-intra.sdcc.bnl.gov/phenix/WWW/p/draft/david/nour/ANALYSIS_NOTE-Run14_pi0.pdf (accessed on 29 April 2025).
- ROOT A Data Analysis Framework. Available online: <https://root.cern.ch/> (accessed on 16 January 2025).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.