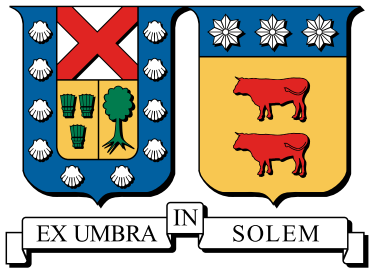
UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
VALPARAÍSO - CHILE



Deep learning for extensive air showers reconstruction: an interpretable multi-task CNN-Transformer model for the CONDOR observatory

Luis Felipe Navarro Farías

Thesis presented in partial fulfillment of the requirements for the degree of
Magíster en Ciencias de la Ingeniería Informática

Advisor
Raquel Pezoa, Ph.D.

Co-Advisor
Nicolás Viaux, Ph.D.

Internal Committee Member
Claudio Torres, Ph.D.

External Committee Member
Gonzalo Farías, Ph.D.

November 24, 2026

CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☐ Memoria o trabajo de título; ☒ Tesis de Postgrado;

Título del trabajo: Deep learning for extensive air showers reconstruction: an interpretable multi-task CNN-Transformer model for the CONDOR observatory

Nombre del candidato(a): Luis Felipe Navarro Farías

Carrera / Grado: Magíster en Ciencias de la Ingeniería Informática

Campus: Casa Central Valparaíso**;** **Departamento**: Departamento de Informática

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Raquel Pezoa, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL

☒ El trabajo **NO contiene información que amerite confidencialidad** y puede ser publicado de inmediato en repositorio con acceso abierto.

☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (embargo) por:

☐ 6 meses; ☐ 12 meses; ☐ 2 años; ☐ 3 años; ☐ 5 años; ☐ 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 05-11-26 ; **Firma:** 

Estudiante o Candidato(a):

Fecha: 05-11-26 ; **Firma:** 

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

THESIS TITLE:

Deep learning for extensive air showers reconstruction: an interpretable multi-task CNN-Transformer model for the CONDOR observatory

AUTHOR:

Luis Felipe Navarro Farías

Thesis presented in partial fulfillment of the requirements for the degree of Magíster en Ciencias de la Ingeniería Informática of the Universidad Técnica Federico Santa María.

Raquel Pezoa, Ph.D. _____
Advisor

Nicolás Viaux, Ph.D. _____
Co-Advisor

Claudio Torres, Ph.D. _____
Internal Committee Member

Gonzalo Farías, Ph.D. _____
External Committee Member

Valparaíso, Chile.
November 24, 2026.

To my beloved Liza, my parents,
and my dear little cat Thorfinn

Acknowledgments

This work was supported by ANID CCTVal (CIA250027) and by Proyectos Internos de Investigación USM 2025 (PI_LII_25_06). The author also acknowledges the support of the Becas Magíster Científico-Tecnológico of the Universidad Técnica Federico Santa María.

Resumen

La reconstrucción de duchas atmosféricas extensas (extensive air showers, EAS) en el régimen sub-TeV es un desafío central para los observatorios de astropartículas de gran altitud. En este trabajo desarrollamos y evaluamos un método de aprendizaje profundo multitarea que reconstruye simultáneamente tres propiedades de las EAS — discriminación gamma-hadrón, ángulo cenital y energía primaria — en el observatorio CONDOR (Desierto de Atacama, Chile, 5.300 m s.n.m.). El modelo es una arquitectura híbrida CNN–Transformer con dos streams de atención por tarea: uno sobre las características procesadas por la CNN y otro directamente sobre la secuencia cruda de hits. Entrenado con datos simulados por CORSIKA en el rango 300–800 GeV, el modelo alcanza un AUC de 0,991 y un F1-score de 0,970 en la discriminación gamma-hadrón, un error angular medio de 0,484° en el ángulo cenital y un error absoluto medio de 103,72 GeV en la energía. La importancia de características por permutación (PFI) y el análisis de ablación coinciden en que la información temporal domina la reconstrucción angular, mientras que las características espaciales y de energía depositada gobiernan la clasificación. La atención sobre la secuencia cruda muestra que cada tarea pondera una parte distinta de la secuencia: la clasificación, su borde inicial; la reconstrucción angular, su parte tardía. Un estudio de dropout de detectores muestra que la discriminación gamma-hadrón es la tarea más tolerante a fallos, y que un núcleo denso favorece la clasificación y el ángulo, mientras que la energía requiere cobertura lateral; la optimización del *layout* queda como trabajo futuro. Estos resultados muestran que el aprendizaje multitarea con explicabilidad integrada es un enfoque viable para la reconstrucción de EAS en observatorios de próxima generación.

Palabras clave: Rayos cósmicos, rayos gamma, duchas atmosféricas extensas, aprendizaje profundo, redes neuronales convolucionales, transformers, aprendizaje multitarea, inteligencia artificial explicable, mecanismos de atención.

Abstract

The reconstruction of extensive air showers (EAS) in the sub-TeV regime is a central challenge for high-altitude astroparticle observatories. In this work, we develop and evaluate a multi-task deep learning method that simultaneously reconstructs three EAS properties — gamma-hadron discrimination, zenith angle, and primary energy — at the CONDOR Observatory (Atacama Desert, Chile, 5,300 m a.s.l.). The model is a hybrid CNN–Transformer architecture with two attention streams per task: one on CNN-processed features and one directly on the raw hit sequence. Trained on CORSIKA-simulated events in the 300–800 GeV range, the model achieves an AUC of 0.991 and an F1-score of 0.970 for gamma-hadron discrimination, a mean angular error of 0.484° for the zenith angle, and a mean absolute error of 103.72 GeV for the energy. Permutation feature importance (PFI) and feature ablation agree that temporal information dominates angular reconstruction, while spatial and deposited-energy features govern classification. Attention on the raw sequence shows that each task weights a different part of the sequence: classification, its leading edge; angular reconstruction, its late portion. A detector dropout study shows that gamma-hadron discrimination is the most fault-tolerant task, and that a dense core favours classification and angular reconstruction, while energy needs lateral coverage; a full layout optimisation is left for future work. These results show that multi-task learning with integrated explainability is a viable approach to EAS reconstruction in next-generation astroparticle observatories.

Keywords: Cosmic ray, gamma ray, extensive air showers, deep learning, convolutional neural networks, transformers, multi-task learning, explainable artificial intelligence, attention mechanisms.

Contents

Resumen **III**

Abstract **IV**

Contents **VII**

List of Figures **VIII**

List of Tables **IX**

1 Introduction **1**
1.1 From Cosmic Rays to the CONDOR Observatory 1
1.2 Deep Learning and Explainability 2
1.3 Objectives 3
1.4 Contributions Beyond the Published Work 4
1.5 Thesis Structure 5

2 Theoretical Background **6**
2.1 Cosmic Rays and High-Energy Astrophysics 6
2.1.1 The Energy Spectrum 7
2.1.2 Gamma Rays as Cosmic Ray Tracers 7
2.1.3 The Sub-TeV Window 8
2.2 Physics of Extensive Air Showers 9
2.2.1 Electromagnetic Showers 9
2.2.2 Hadronic Showers 11
2.2.3 Shower Observables 12
2.3 Ground-Based Detection of EAS 15
2.4 The CONDOR Observatory 15
2.4.1 Design and High-Altitude Advantage 16
2.4.2 Array Layout and Detector Architecture 16
2.4.3 Data Acquisition and Sub-Nanosecond Synchronization 18

2.5 The CORSIKA Simulation Framework . . . 19
2.6 Convolutional Neural Networks . . . 20
2.7 Transformer Architectures . . . 21
2.8 Multi-Task Learning . . . 23
2.9 Explainable Artificial Intelligence . . . 24
2.9.1 Permutation Feature Importance . . . 24
2.9.2 Feature Ablation . . . 25
2.9.3 Attention Pattern Analysis . . . 25
2.10 Reconstruction of EAS: State of the Art . . . 26
2.10.1 Classical Parametric Reconstruction . . . 26
2.10.2 Deep Learning for Air Shower Reconstruction . . . 27

3 Methodology **29**
3.1 Experimental Setup . . . 29
3.2 Dataset Construction and Preprocessing . . . 30
3.3 Data Stratification and Balance Strategy . . . 32
3.4 Model Architecture . . . 36
3.5 Hyperparameter Optimization . . . 39
3.6 Training Configuration . . . 39
3.7 Code and Data Availability . . . 41

4 Results and Explainability **42**
4.1 Multi-Task Performance . . . 42
4.1.1 Gamma-Hadron Discrimination . . . 43
4.1.2 Zenith Angle Reconstruction . . . 45
4.1.3 Energy Estimation . . . 46
4.1.4 Comparison with the State of the Art . . . 49
4.2 Explainability Analysis . . . 50
4.2.1 Feature Importance . . . 51
4.2.2 Attention Analysis . . . 52
4.3 Detector Robustness and Array Optimization . . . 55
4.3.1 Single-Detector Importance . . . 55
4.3.2 Regional Dropout . . . 55
4.3.3 Detector Count and Layout . . . 58

5 Conclusions **62**
5.1 Summary of Contributions . . . 62
5.2 Limitations and Future Work . . . 63

References **66**

A Supplementary Material **70**
A.1 Atmospheric Depth 70
A.2 Model Architecture Summary 72
A.3 Global Feature Definitions 73
A.4 Quality Factor Definition 74
A.5 Angular Resolution: PSF_{68} Definition 74

List of Figures

2.1 The cosmic ray energy spectrum. 8
2.2 Branching structure of an air shower. 10
2.3 Average shower footprint: proton vs photon comparison. 12
2.4 Lateral particle density predicted by the NKG function. 13
2.5 Arrival geometry of an extensive air shower. 14
2.6 Angular evolution of the average EAS footprint. 14
2.7 Schematic of the CONDOR detector array layout. 16
2.8 Simulated CONDOR detector-module geometry and identifiers. . . . 17

3.1 Representative individual EAS footprints by particle type and energy. 32
3.2 Average EAS footprints by particle type and energy. 33
3.3 Event distributions before and after stratified balancing. 35
3.4 Flow diagram of the CNN-Transformer multi-task model. 38

4.1 Training and validation curves across the three tasks. 43
4.2 Gamma-hadron discrimination results. 44
4.3 ROC curves grouped by energy and zenith angle. 45
4.4 Zenith angle reconstruction results. 46
4.5 Energy pull distribution overall and by particle type. 48
4.6 Energy migration matrix. 49
4.7 Feature ablation study results. 52
4.8 Permutation feature importance results. 53
4.9 Attention distribution of the RAW streams. 54
4.10 Single-detector importance map. 56
4.11 Detector-group definitions for the regional dropout analysis. 56
4.12 Impact of switching off whole detector groups. 57
4.13 Reconstruction under progressive random detector failure. 59
4.14 Array geometry trade-off at matched detector counts. 60

List of Tables

3.1 Dual-input feature structure. 31
3.2 Dataset composition before and after balancing. 34
3.3 Dataset composition after stratified balancing. 35
3.4 Consolidated hyperparameter configuration. 40

4.1 Summary of multi-task reconstruction performance. 43
4.2 Gamma-hadron discrimination performance on the test set. 44
4.3 Zenith angle reconstruction metrics on the test set. 46
4.4 Energy estimation metrics on the test set. 47
4.5 Energy bias and resolution split by particle type. 47
4.6 Comparison with state-of-the-art EAS reconstruction. 50

A.1 Vertical atmospheric depth at several observation levels. 72
A.2 Layer-by-layer model summary. 73
A.3 Global feature computation and ordering. 73

Chapter 1

Introduction

The universe is permeated by a continuous flux of high-energy particles arriving at Earth from all directions. These particles, recognized as cosmic rays, were first identified in the early twentieth century (Hess, 1912) and remain an active area of astrophysical research (Bell, 2013). Despite significant experimental progress, the precise locations of their astrophysical sources, as well as the physical mechanisms governing their acceleration to ultra-high energies, have not yet been definitively established (Blasi, 2013). Understanding cosmic rays provides an observational window into some of the most energetic and extreme environments in the universe.

1.1 From Cosmic Rays to the CONDOR Observatory

A central goal of high-energy astrophysics is identifying the acceleration sites of cosmic rays. Because cosmic rays are predominantly charged particles, their trajectories are deflected by galactic magnetic fields, erasing any directional information about their sources by the time they arrive at Earth. Gamma rays, being electrically neutral, travel in straight lines from their production sites and therefore serve as direct tracers of cosmic ray activity (Morrison, 1958). However, isolating these photon signals from the overwhelming background of hadronic cosmic ray events constitutes a central challenge for ground-based observatories (Alfaro et al., 2022; Deligny & Collaboration, 2025; Schneider et al., 2026). The physics underlying this challenge is discussed in detail in Chapter 2.

When a high-energy primary particle strikes the upper atmosphere, it initiates a multiplying cascade of secondary particles known as an extensive air shower (EAS) (Rao & Sreekantan, 1998). The secondary particles that survive to the detector level form a spatiotemporal footprint across the array. Reconstructing the

primary particle's type, energy, and arrival direction from this sparse and stochastic footprint is the central problem of ground-based astroparticle physics and of this thesis.

An observational gap exists in the sub-TeV window between 100 GeV and 1 TeV: space-based instruments such as the Fermi Large Area Telescope (Fermi-LAT) are limited by collection area at higher energies (Ackermann et al., 2012), while ground-based installations at altitudes around 4,000 m impose effective energy thresholds[1] of several hundred GeV. The CONDOR observatory (Arratia et al., 2025), planned for Cerro Toco in the Atacama Astronomical Park at 5,300 m above sea level, is designed to bridge this gap by intercepting the shower front closer to its atmospheric maximum, lowering the effective energy threshold to approximately 100 GeV.

Analyzing the low-energy events accessible to CONDOR presents a significant computational challenge. At 100 GeV, the spatiotemporal footprint of a shower is sparse, noisy, and subject to large stochastic fluctuations, and extracting the kinematic properties of the primary particle from it is an inverse problem with no closed-form solution.

1.2 Deep Learning and Explainability

Historically, EAS observatories relied on classical parametric algorithms, for example, planar fitting for arrival direction reconstruction (Collaboration et al., 2014) and the Nishimura-Kamata-Greisen lateral distribution function (Kamata & Nishimura, 1958) for energy estimation (Kawata et al., 2017). While computationally lightweight, these methods show degraded performance under the noise conditions and sparse hit patterns of low-energy showers, and they cannot naturally handle the simultaneous estimation of multiple shower properties.

Deep learning (LeCun et al., 2015) shifts this paradigm by enabling end-to-end representation learning directly from raw detector hits. Convolutional neural networks (CNNs) (LeCun & Bengio, 1998) have proven effective at extracting local features from the hit patterns of dense detector arrays (Okukawa et al., 2024), while Transformer architectures (Vaswani et al., 2017), through their self-attention mechanism, are well suited for capturing the global, long-range correlations that span the full shower footprint (Lin et al., 2022; Langner et al., 2024). Despite recent advances, existing frameworks for EAS reconstruction often constrain themselves to single-task objectives or rely on detector geometries dissimilar to CONDOR (Erdmann et al., 2018; Okukawa et al., 2024), making direct translation difficult. The predecessor

[1]The *effective energy threshold* is the lowest primary energy at which an array can still detect and reconstruct showers with reasonable efficiency. It is set by how much of the shower survives down to the observation altitude, and is discussed further in Chapter 2.

to this work (Navarro et al., 2025) introduced a CNN-LSTM architecture for CONDOR that addressed two of the three tasks — particle classification and zenith angle regression — but remained limited by sequential processing constraints and the absence of energy estimation as a third reconstruction task. This thesis extends that foundation with a hybrid CNN-Transformer architecture capable of simultaneously addressing gamma-hadron discrimination, zenith angle reconstruction, and primary energy estimation within a unified multi-task framework.

An ongoing challenge in the scientific adoption of deep learning is the tension between predictive performance and physical interpretability (Ribeiro et al., 2016). A model achieving high reconstruction accuracy yields limited scientific value if its predictions rely on non-physical or spurious correlations in the training data. This concern is particularly relevant when the model is trained on data synthesized by a Monte Carlo framework like CORSIKA (Heck et al., 1998), the standard simulation code for extensive air showers, which is described in detail in Section 2.5; under these conditions, the risk of learning simulation artifacts rather than genuine shower physics cannot be dismissed without explicit verification. This is why the adoption of explainable artificial intelligence (XAI) methods (Adadi & Berrada, 2018) is indispensable for understanding the model's reasoning when generating predictions.

This thesis addresses this requirement through two complementary feature importance methods — permutation feature importance (Altmann et al., 2010) and a feature ablation study, which quantify the contribution of each input channel to each reconstruction task by intervening on the input — together with an analysis of the attention patterns, which describes the temporal focus of the Transformer streams along the shower front. Confirming that these analyses align with physical expectations constitutes a key scientific deliverable of this work, alongside the reconstruction performance itself.

1.3 Objectives

The primary objective of this thesis is to design, evaluate, and interpret a deep learning model for the simultaneous reconstruction of multiple physical properties of extensive air showers, specifically developed for the high-altitude conditions of the CONDOR observatory.

The specific objectives are:

1. **To design** and implement a hybrid CNN-Transformer multi-task architecture capable of simultaneously performing gamma-hadron discrimination, zenith angle regression, and primary energy estimation from variable-length sequences of spatiotemporal detector hits.

2. **To devise** a data preprocessing and stratified balancing strategy that mitigates the class and configuration imbalances inherent in Monte Carlo simulation datasets.

3. **To evaluate** the architecture's performance on a held-out test set and compare its reconstruction accuracy against classical baselines and prior deep learning approaches for CONDOR.

4. **To interpret** the learned representations using permutation feature importance, feature ablation, and attention pattern analysis, validating their physical consistency against established shower phenomenology.

5. **To identify** the structural and algorithmic limitations of the current framework and provide a transparent methodological foundation for future adaptation to real detector data once CONDOR begins operations.

1.4 Contributions Beyond the Published Work

The research carried out in this thesis has given rise to the following published contributions:

1. **JCC 2025.** A first two-task CNN-LSTM model for CONDOR was presented at the Jornadas Chilenas de la Computación (Navarro et al., 2025), addressing gamma-hadron classification and zenith angle regression.

2. **Astronomy and Computing.** The core multi-task reconstruction results and the three-stream CNN-Transformer architecture were published in a companion journal paper (Navarro et al., 2026a).

3. **ICHEP 2026.** The main results of the journal paper were presented at the 43rd International Conference on High Energy Physics (Navarro et al., 2026b).

This thesis extends those published works in several directions that are not covered by the papers:

- An extended six-stream architecture with independent RAW attention streams per task, enabling hit-level interpretability without convolutional indirection.

- A feature ablation study as a second, independent method for validating feature importance, complementing the permutation-based analysis reported in the paper.

- A ROC analysis grouped by primary energy and zenith angle, revealing the physically consistent energy dependence of discrimination performance.
- A detector robustness study quantifying the fault tolerance of the trained model to missing detectors across the three tasks.

1.5 Thesis Structure

The remainder of this thesis is organized as follows. Chapter 2 establishes the theoretical foundations of the work, covering the physics of cosmic rays and extensive air showers, the CONDOR observatory and its simulation framework, and the deep learning architectures and explainability methods employed. Chapter 3 describes the full methodological pipeline, from dataset construction and preprocessing through model architecture, hyperparameter optimization, and training configuration. Chapter 4 presents the results across all three reconstruction tasks and the explainability analysis. Chapter 5 concludes with a summary of the main findings, a discussion of limitations, and directions for future work.

Chapter 2

Theoretical Background

This chapter establishes the physical and computational foundations necessary for the reconstruction of EAS at CONDOR. The chapter begins with the properties and astrophysical context of cosmic rays, followed by the phenomenology of atmospheric particle cascades and the principles of ground-based detection. It then describes the CONDOR observatory and the Monte Carlo simulation framework used to generate the training data. Finally, it introduces the deep learning architectures and explainability methods that form the technical core of this thesis.

2.1 Cosmic Rays and High-Energy Astrophysics

The study of cosmic rays sits at the intersection of particle physics, astrophysics, and cosmology. First discovered in 1912 by Victor Hess (Hess, 1912) through a series of balloon ascents that measured increasing ionization rates with altitude, cosmic rays are not electromagnetic radiation but high-energy subatomic particles, primarily bare atomic nuclei, traveling through space at relativistic speeds (Longair, 2011). They carry information about some of the most extreme physical environments in the universe, yet the precise mechanisms behind their acceleration remain one of the central open problems in modern astrophysics (Blasi, 2013).

Decades of space-based and balloon-borne measurements have established a precise picture of the composition of cosmic rays at sub-TeV energies. By particle flux, approximately 89% of primary cosmic rays are protons, roughly 10% are helium nuclei, and the remaining fraction consists of heavier atomic nuclei up to iron, alongside electrons, positrons, and traces of antimatter (Navas et al., 2024).

Because these primaries are electrically charged, their propagation through the interstellar and intergalactic medium is governed by the Lorentz force, which continuously deflects their trajectories as they pass through irregular galactic magnetic fields. Over astronomical distances, this deflection randomizes their arrival directions

at Earth, permanently erasing the geometric line of sight between their production sites and the observer. As a consequence, identifying cosmic ray sources through arrival direction alone is not generally possible, and indirect tracers such as gamma rays and neutrinos become essential tools.

2.1.1 The Energy Spectrum

The energy spectrum of cosmic rays spans more than twelve orders of magnitude and follows a sequence of power laws of the form

$$\Phi(E) \propto E^{-\gamma} \tag{2.1}$$

where $\Phi(E)$ is the differential flux and γ is the spectral index. Rather than a single unbroken power law, the spectrum exhibits well-characterized structural breaks that reflect transitions between different source populations, acceleration limits, and propagation regimes. For example, as illustrated in Figure 2.1, precise measurements of this spectrum reveal distinct steepenings and flattenings at specific energy scales — commonly referred to as the *knee*, the *second knee*, and the *ankle* — each associated with changes in the dominant physical processes governing where and how cosmic rays are produced and transported through the Galaxy. Observations of the spectrum and mass composition are therefore fundamental tools for understanding the origin and acceleration mechanisms of high-energy cosmic rays, and they directly motivate the design of observatories sensitive to specific energy windows, such as the sub-TeV range targeted by CONDOR.

2.1.2 Gamma Rays as Cosmic Ray Tracers

A central motivation for gamma-ray astronomy in the context of cosmic ray physics is that photons, being electrically neutral, are not deflected by interstellar magnetic fields and therefore travel in straight lines from their production sites. This makes them valuable probes of cosmic ray acceleration environments. When a high-energy proton interacts with ambient matter near its source, it produces pions, and neutral pions decay promptly into photon pairs

$$\pi^0 \rightarrow \gamma + \gamma, \tag{2.2}$$

generating gamma-ray emission that directly reflects the parent proton spectrum. Detecting this hadronic signature from a source constitutes direct evidence of cosmic ray acceleration at that site.

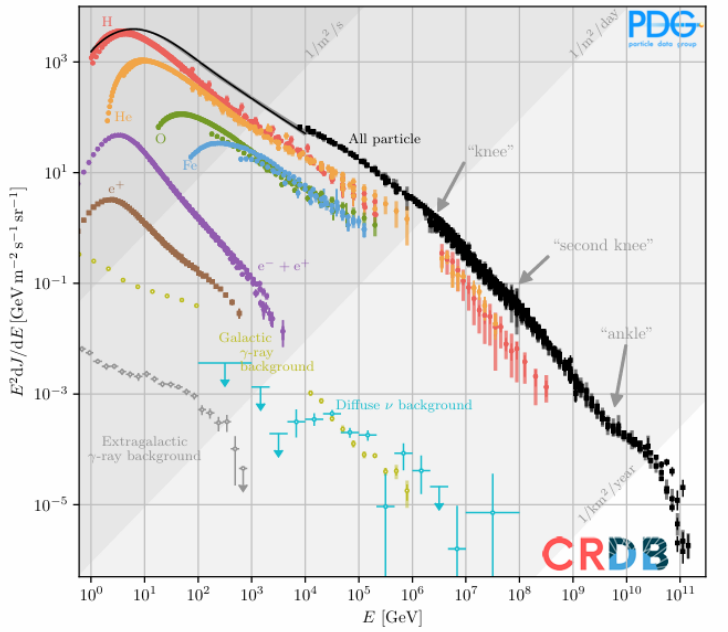


Figure 2.1: The cosmic ray energy spectrum. The intensity of charged and neutral cosmic rays, multiplied by kinetic energy squared, is shown as a function of energy. Characteristic breaks such as the knee, second knee, and ankle are clearly visible.

However, high-energy electrons can also produce gamma rays through inverse Compton scattering[1] and bremsstrahlung[2], generating leptonic emission that closely resembles the hadronic signature spectrally. Distinguishing between the two is a central observational challenge, and it is precisely this challenge that motivates the task of gamma-hadron discrimination at ground-based detectors.

2.1.3 The Sub-TeV Window

The energy range between 100 GeV and a few TeV marks the boundary between what is accessible to space-based instruments and what requires ground-based observation. The Fermi-LAT, a space-based pair-conversion gamma-ray telescope operating aboard a satellite above the Earth's atmosphere, effectively detects gamma rays up to a few hundred GeV but is limited by its collection area at higher energies (Ackermann et al., 2012). Above this threshold, ground-based detectors exploit the EAS that cosmic rays and gamma rays produce in the atmosphere, reaching

[1]Inverse Compton scattering is the process by which a relativistic electron transfers energy to a low-energy photon, up-scattering it to gamma-ray energies; it is the leptonic counterpart to the hadronic π^0 channel.

[2]Bremsstrahlung (German for "braking radiation") is the radiation emitted when a charged particle, typically an electron, is deflected by the electric field of an atomic nucleus.

the detector level as measurable patterns of secondary particles. CONDOR is designed to operate precisely in this sub-TeV window, and the reconstruction of shower properties in this regime is the central problem addressed in this thesis.

2.2 Physics of Extensive Air Showers

When a high-energy primary particle enters the Earth's atmosphere, it does not reach the ground directly. Instead, its interaction with atmospheric nuclei initiates a cascade of secondary particles that propagates downward at nearly the speed of light, growing in number until the energy per particle drops below the threshold for further multiplication. These cascades, known as extensive air showers, are the fundamental observable that ground-based gamma-ray and cosmic ray observatories exploit (Rossi & Greisen, 1941).

Figure 2.2 summarizes the sequence of processes that follow the first interaction. The primary produces a large multiplicity of secondary hadrons, mostly pions; the charged ones sustain the hadronic cascade and eventually decay into muons, while the neutral ones decay into photon pairs and transfer their energy into electromagnetic sub-cascades. The three components that reach the ground — hadronic, muonic and electromagnetic — are what a surface array actually samples, and the following subsections describe the electromagnetic and hadronic branches in turn.

2.2.1 Electromagnetic Showers

The simplest type of EAS is initiated by a high-energy photon or electron. When a photon of sufficient energy passes close to an atmospheric nucleus, it converts into an electron-positron pair in the screened Coulomb field of that nucleus. This process, known as pair production, requires a photon energy of at least twice the electron rest mass ($2m_ec^2 \approx 1.022$ MeV) and relies on the nearby nucleus to absorb the recoil momentum so that energy and momentum are simultaneously conserved. Each member of the resulting pair is then deflected in the same nuclear fields and radiates a bremsstrahlung photon, which can in turn produce a further pair, and so on. The cascade multiplies rapidly, roughly doubling the particle number at each step, until the energy per particle is no longer sufficient to sustain further pair production.

A useful framework for understanding this process is the Heitler model (Heitler, 1984; Matthews, 2005), which describes the shower as a binary splitting process. After each splitting length, every particle splits into two secondaries of half the original energy. Starting from a primary of energy E_0, the shower reaches its maximum particle number when the energy per particle equals the critical energy E_c, below

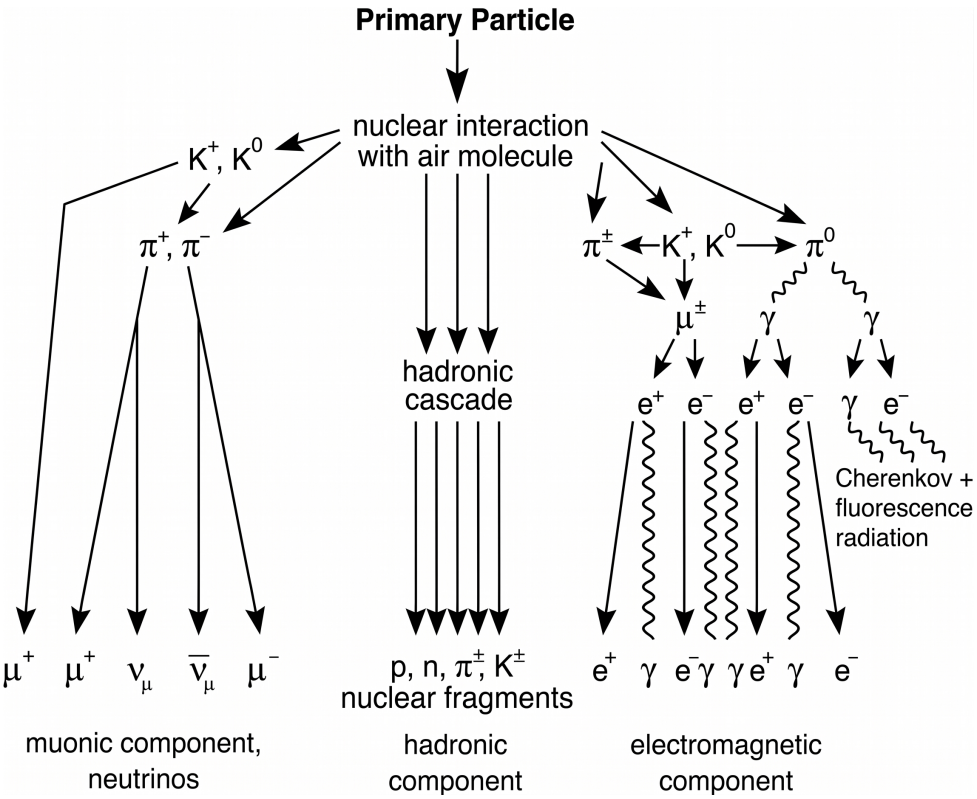


Figure 2.2: Branching structure of an air shower, grouped into the three components sampled at ground level. **Left:** the muonic and neutrino component. Charged kaons and pions ($K^{\pm}$, K^0, $\pi^{\pm}$) decay into muons and muon neutrinos, which are weakly interacting and reach the ground largely undisturbed. **Centre:** the hadronic component. Successive nuclear interactions sustain the hadronic cascade, whose surviving nucleons and mesons (p, n, $\pi^{\pm}$, $K^{\pm}$) constitute the nuclear fragments that reach the ground. **Right:** the electromagnetic component. Neutral pions decay promptly into photon pairs, which alternate pair production and bremsstrahlung to build an electron-positron-photon cascade; this component is what a Cherenkov or fluorescence detector observes directly, and what dominates the particle count of the shower.

which ionization losses dominate over radiative processes. The number of generations required is $n = \log_2(E_0/E_c)$, so the maximum occurs at an atmospheric depth

$$X_{\max} = \frac{X_0}{\ln 2} \ln \frac{E_0}{E_c} \,, \tag{2.3}$$

measured in g cm^{-2} of traversed atmosphere, where X_0 is the radiation length[3] and

[3]The radiation length is the mean atmospheric depth over which a high-energy electron loses all but a fraction $1/e$ of its energy through bremsstrahlung; it is the natural length scale of electromagnetic shower development.

the factor $X_0/\ln 2$ is the effective splitting length of the cascade. The quantity $X_{\max}$ is one of the most important observables in EAS physics: it depends logarithmically on the primary energy and is sensitive to the primary particle type, making it a key discriminant between electromagnetic and hadronic showers.

Shower development is tracked in units of *atmospheric depth* rather than geometric distance, which is why $X_{\max}$ carries units of g cm^{-2}. Atmospheric depth is the mass of air per unit area contained in the column above a given point, and it is the natural coordinate for a cascade because the number of interactions a particle undergoes depends on how much matter it has crossed and not on how far it has moved. Appendix A.1 develops this in full, together with its two consequences for this work: the dependence of the traversed depth on the zenith angle, and the advantage of a high-altitude observation site.

2.2.2 Hadronic Showers

Showers initiated by protons or heavier nuclei are considerably more complex. The primary particle interacts inelastically with an atmospheric nucleus, producing a large multiplicity of secondary hadrons, predominantly pions. Charged pions $\pi^{\pm}$ continue the hadronic cascade through further inelastic interactions, while neutral pions π^0 decay almost immediately into photon pairs, seeding electromagnetic subshowers at every generation. A hadronic shower is therefore a composite structure: a hadronic core that continuously feeds energy into a superposition of electromagnetic cascades developing at different depths.

This composite structure has two important consequences for detection. First, hadronic showers are intrinsically wider and more irregular in their lateral extent than purely electromagnetic showers, because hadronic interactions transfer significant transverse momentum and the multiple electromagnetic sub-showers develop independently along different axes. Second, a fraction of the primary energy is carried away by neutrinos produced in charged pion and muon decay, making hadronic showers less efficient at producing detectable particles at ground level than gamma-induced showers of the same primary energy. These morphological and energetic differences form the physical basis for gamma-hadron discrimination (Rao & Sreekantan, 1998).

Figure 2.3 illustrates these differences at the level of the ground detector array. The average footprint of gamma-induced showers is more compact and concentrated near the shower core, while proton-induced showers exhibit a broader, less symmetric lateral distribution and higher particle counts at peripheral detectors. The difference map (right panel) highlights the core-dominated excess for gamma events and the extended halo for hadronic events, providing a direct visual foundation for the

morphological discrimination task.

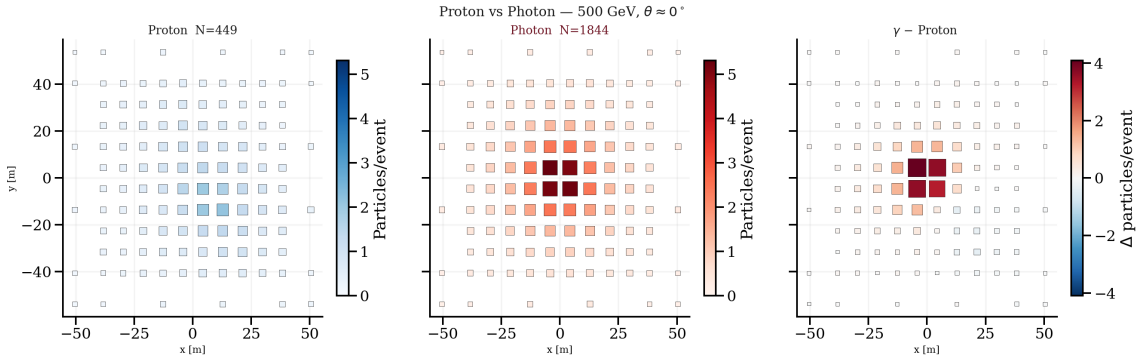


Figure 2.3: Average shower footprint at 500 GeV and $\theta \approx 0°$. **Left:** mean particle count per detector for proton-induced showers. **Center:** mean particle count per detector for gamma-induced showers. **Right:** difference map (γ − proton), revealing the more compact core signature of electromagnetic showers versus the wider lateral spread of hadronic showers. Marker size is proportional to particle count; position corresponds to physical detector location on the array.

2.2.3 Shower Observables

At ground level, an EAS manifests as a propagating disk of particles sweeping across the detector array and leaving a spatiotemporal pattern of hits. The key observables recorded by a surface array are the number of particles at each detector, the deposited energy, and the arrival time of the shower front.

The lateral distribution describes how the density of secondary particles varies with the distance from the shower core at the detector altitude. For the electromagnetic component of a shower, this distribution is well approximated by the Nishimura-Kamata-Greisen (NKG) function (Kamata & Nishimura, 1958),

$$\rho(r) = \frac{N_e}{2\pi r_M^2} \frac{\Gamma(4.5-s)}{\Gamma(s)\Gamma(4.5-2s)} \left(\frac{r}{r_M}\right)^{s-2} \left(1+\frac{r}{r_M}\right)^{s-4.5}, \tag{2.4}$$

where N_e is the total number of electrons, r is the distance from the shower core, r_M is the Molière radius,[4] and s is the shower age parameter that tracks the developmental stage of the cascade ($s = 1$ at shower maximum). The NKG function is relevant here as the classical parametric description of the lateral distribution: traditional reconstruction pipelines fit it to the measured hit pattern to estimate the shower

[4]The Molière radius is the characteristic transverse scale of an electromagnetic shower, defined as the radius of the cylinder around the shower axis that contains, on average, about 90% of the deposited energy; it sets the lateral width of the particle distribution.

size and core position. The deep learning model developed in this thesis does not fit any fixed functional form; instead, it learns the lateral and temporal structure of the shower directly from the detector hits, and therefore does not assume the shape given by the NKG function in advance. Figure 2.4 shows the distribution predicted by this function.

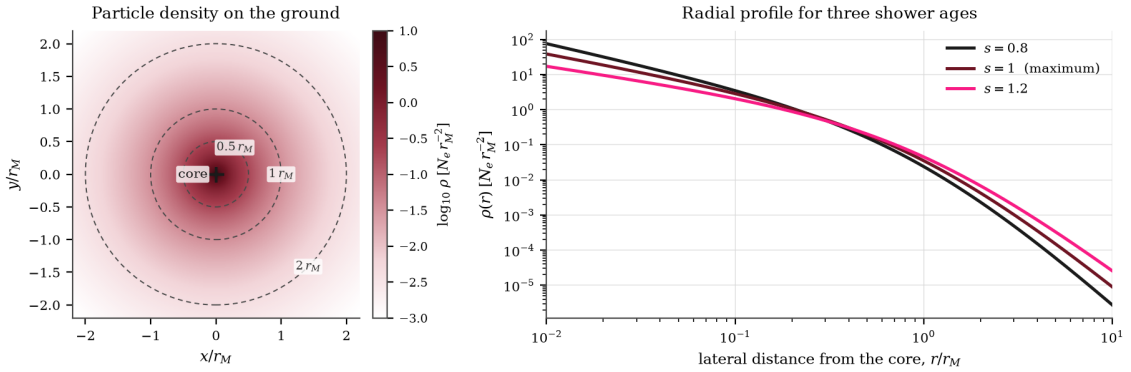


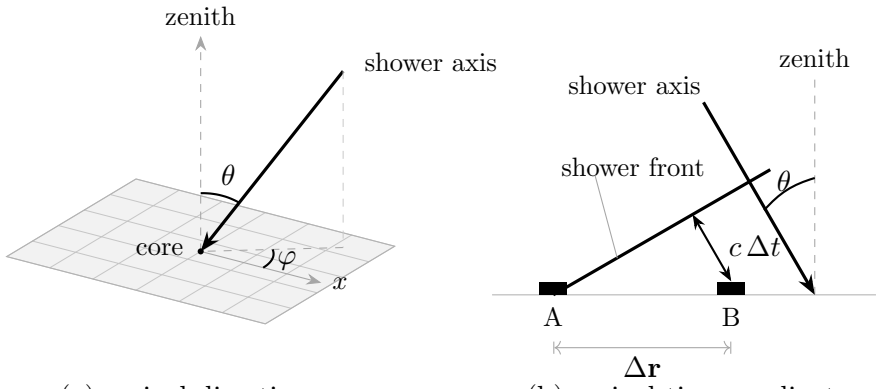
Figure 2.4: Lateral particle density predicted by the NKG function of Equation 2.4, in units of the Molière radius r_M so that it does not depend on the air density at the observation altitude. **Left:** density on the ground at shower maximum ($s = 1$), with rings at 0.5, 1 and $2\,r_M$. **Right:** radial profile for three shower ages: a young shower is more concentrated on the core, an older one flatter and wider.

The timing of the shower front encodes the arrival geometry of the primary particle. As shown in Figure 2.5, for an ideal planar front traveling at the speed of light, the arrival time difference between two detectors separated by a displacement vector $\Delta\mathbf{r}$ is

$$\Delta t = -\frac{\hat{n} \cdot \Delta\mathbf{r}}{c}, \tag{2.5}$$

where $\hat{n}$ is the unit vector in the shower propagation direction. In practice, shower front curvature and stochastic fluctuations introduce deviations from this ideal model, but the timing information remains the primary observable for reconstructing the arrival direction of the incoming primary. In this thesis the azimuthal angle is held fixed in the simulations (Section 2.5), so the arrival-time gradient is exploited specifically for the reconstruction of the zenith angle θ — one of the three target tasks of the model — which is why this observable is emphasized here.

Figure 2.6 shows how the average shower footprint evolves with the zenith angle for 500 GeV proton-induced showers simulated at CONDOR. As θ increases, the footprint elongates along the projection of the shower axis onto the detector plane, the lateral spread widens, and the arrival-time gradient across the array steepens. These systematic morphological changes underlie the angular reconstruction task addressed in this thesis.



(a) arrival direction (b) arrival-time gradient

Figure 2.5: Arrival geometry of an extensive air shower. **(a)** The direction of the primary is given by the zenith angle θ, from the vertical, and the azimuth φ, on the plane of the array; the axis meets that plane at the core. **(b)** Side view. The front is perpendicular to the axis, so it reaches A first and still has $c\,\Delta t$ to travel before reaching B; that delay is the projection of their separation onto the direction of propagation, Equation 2.5. Here φ is fixed at 0° (Section 2.5), so only θ is reconstructed.

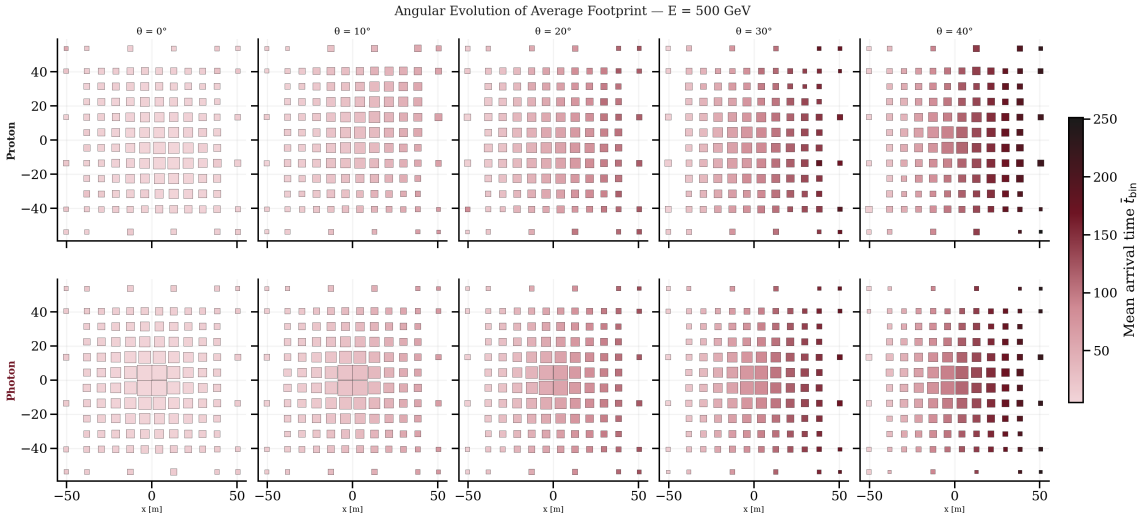


Figure 2.6: Angular evolution of the average shower footprint for proton-induced events at $E = 500\,\mathrm{GeV}$. Each panel shows the mean particle count per detector averaged over all events within the corresponding zenith angle bin ($\theta \approx 0°$, 10°, 20°, 30°, 40°). As the zenith angle increases, the shower footprint elongates along the projected shower axis and the lateral spread broadens, providing the morphological signature exploited for zenith angle reconstruction. Color encodes the mean arrival time $\bar{t}_{\mathrm{bin}}$; marker size is proportional to particle count.

2.3 Ground-Based Detection of EAS

Ground-based observatories exploit the EAS as an indirect measurement device: rather than detecting the primary particle directly, they sample the shower footprint at ground level and use the recorded hit pattern to infer the properties of the primary cosmic ray, such as its type, energy, and arrival direction. This approach enables observation over large sky areas simultaneously and is not limited by collection area in the way that satellite detectors are.

In the sub-TeV and TeV energy regimes, ground-based arrays fall into two dominant technology categories. Water Cherenkov detectors, used by observatories such as HAWC (Abeysekara et al., 2023) and LHAASO (Cao et al., 2022), consist of large tanks of purified water in which secondary particles produce Cherenkov light[5] recorded by photomultiplier tubes. They offer good muon sensitivity, which is advantageous for gamma-hadron discrimination, but require large structural investment at high altitude. The same technology has been adopted for the Southern Wide-field Gamma-ray Observatory (SWGO) (Abreu et al., 2019), a next-generation array proposed for the southern hemisphere, which would extend this coverage to a sky region that HAWC and LHAASO cannot observe. Plastic scintillator arrays, used by operating observatories such as Tibet ASγ (Amenomori et al., 2019) and adopted for the future CONDOR array (Arratia et al., 2025), consist of segmented scintillator panels that produce light pulses when charged particles pass through them, read out by silicon photomultipliers (SiPMs) through wavelength-shifting fibers. Scintillator arrays offer better timing resolution, higher modularity, and lower deployment weight than water Cherenkov detectors, making them suitable for operation at the highest altitudes.

2.4 The CONDOR Observatory

The CONDOR (**CO**mpact **N**etwork of **D**etectors with **O**rbital **R**ange) observatory is a proposed high-altitude gamma-ray and cosmic ray array designed to observe EAS in the 100 GeV to 1 TeV range (Arratia et al., 2025). It is currently in its design and validation phase, so all data used in this thesis are generated through Monte Carlo simulations that reproduce the expected response of the CONDOR array under its planned high-altitude operating conditions; the simulation framework is described in detail in Section 2.5. Figure 2.7 shows a schematic representation of the array layout.

[5] Cherenkov light is the faint electromagnetic radiation emitted when a charged particle traverses a medium faster than the local phase velocity of light in that medium; its intensity and angular profile encode the particle's velocity and direction.

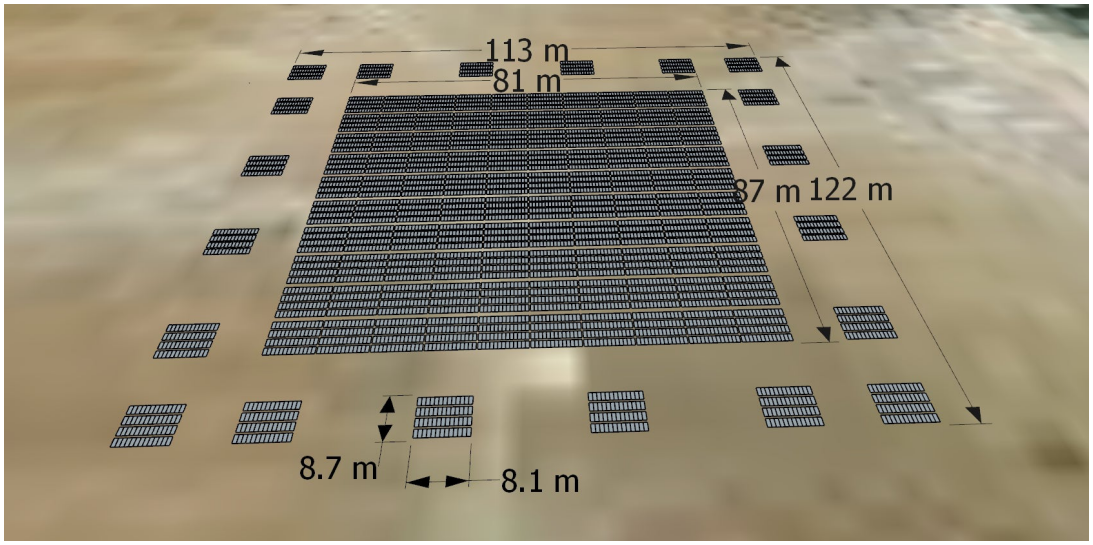


Figure 2.7: Schematic representation of the CONDOR observatory layout. The array achieves a high fill factor in its inner region using thousands of plastic scintillator units, surrounded by an outer veto ring to suppress backgrounds. Source: CONDOR Observatory (Arratia et al., 2025).

2.4.1 Design and High-Altitude Advantage

CONDOR will be located within the Atacama Astronomical Park at Cerro Toco, Chile, at an altitude of 5,300 meters above sea level, placing it above established facilities such as HAWC at 4,100 m and LHAASO at 4,410 m. This altitude is the central design advantage of CONDOR. The number of secondary electrons reaching the ground level depends strongly on atmospheric depth: by operating closer to the shower maximum $X_{\max}$, CONDOR avoids the progressive attenuation experienced by cascades traveling to lower elevations. This enables CONDOR to lower its effective energy threshold to approximately 100 GeV, accessing a range that is difficult to reach for lower-altitude arrays and that partially overlaps with the upper sensitivity limit of space-based gamma-ray instruments such as the Fermi-LAT.

2.4.2 Array Layout and Detector Architecture

The CONDOR array consists of 5,300 plastic scintillator detector units arranged in a compact central grid covering an area of approximately 6,000 m^2, supplemented by 1,040 additional peripheral units forming a veto ring that encloses a total instrumented area of 14,000 m^2. The central region achieves a fill factor of 90%, meaning that nearly the entire surface is actively sensitive to shower particles. This high fill factor is essential for low-energy operation, where the shower footprint is small and

sparse hits must be captured efficiently.

For the purpose of simulation and reconstruction, these individual detector units are grouped into **detector modules**, each covering an area of approximately 8.1 × 8.7 m^2. As shown in Figure 2.8, the full array comprises 120 modules: 100 arranged in a 10 × 10 central grid, plus 20 peripheral modules. Each module aggregates the signals from its constituent scintillator panels into a single readout channel.

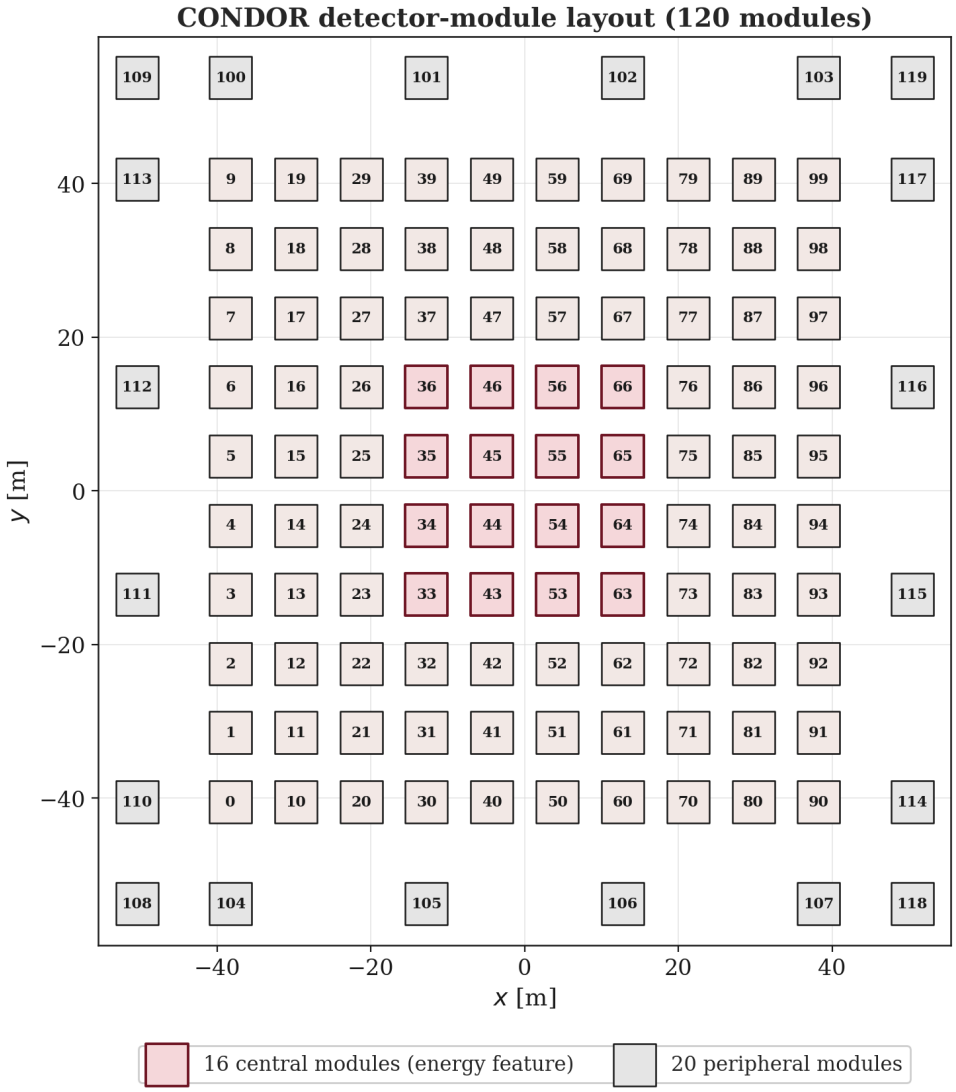


Figure 2.8: Simulated CONDOR detector-module geometry used throughout this thesis. The 100 central modules (identifiers 0–99) form a 10 × 10 grid spanning approximately 87 × 81 m^2, surrounded by 20 peripheral modules (identifiers 100–119), matching the planned observatory configuration. Each module aggregates the signals of its constituent scintillator units into a single readout channel. The module identifier is used directly as the `detector_id` input feature of the model.

It is important to distinguish the two levels at which the shower is represented. The CORSIKA simulation itself does not know about detector modules: it tracks the air shower as individual particles and records their type, energy, and arrival position and time at the ground. The mapping onto the 120 detector modules is performed afterwards, during the data-processing stage, where the raw particle hits falling within each module's footprint are aggregated into the per-module quantities used by the model (Chapter 3). Throughout this thesis, the term *detector* therefore refers to these modules unless otherwise specified, and the analysis operates at the module level rather than at the level of individual scintillator units or raw particle hits.

Each detector unit consists of extruded plastic scintillator bars, each coated with a 0.25 mm layer of titanium dioxide to enhance internal light reflection. A wavelength-shifting optical fiber runs through a central hole in each bar, collecting the fluorescent light produced by ionizing particles and guiding it to a silicon photomultiplier. The chosen SiPM model, the Hamamatsu S14160-6015PS, features 159,565 pixels and a photon detection efficiency of 32%, providing single-photon sensitivity and precise timing. The units are enclosed in 5 mm aluminum housings for protection against the harsh Atacama environment, including the large daily temperature variations that require remote bias voltage adjustments to maintain stable detector performance.

2.4.3 Data Acquisition and Sub-Nanosecond Synchronization

Reconstructing the arrival direction of a shower front traveling at nearly the speed of light requires timing accuracy across thousands of detector nodes separated by tens to hundreds of meters. The intrinsic time resolution of each CONDOR unit is designed to be below one nanosecond, and sub-nanosecond synchronization across the full array is therefore a system-level requirement.

CONDOR is designed to achieve this through White Rabbit technology (Moreira et al., 2009), an Ethernet-based timing protocol based on the IEEE 1588 standard[6] that distributes a pulse-per-second signal derived from the Global Positioning System (GPS) through fiber optic cables to White Rabbit nodes distributed across the array. Each node serves approximately ten detector panels, providing sub-nanosecond timing to each readout board. This synchronization allows offline discrimination between genuine shower hits and uncorrelated dark noise pulses, and it is the foundation on which the timing-based reconstruction of arrival direction depends.

[6]IEEE 1588, also known as the Precision Time Protocol (PTP), is an international standard for synchronizing clocks across nodes of a packet-switched network to sub-microsecond accuracy.

2.5 The CORSIKA Simulation Framework

Because the CONDOR observatory has not yet begun scientific operations, all data used in this thesis are generated through Monte Carlo simulation. The simulation tool is CORSIKA (COsmic Ray SImulations for KAscade), developed at the Karlsruhe Institute of Technology and the standard framework for EAS simulation across the ground-based cosmic ray and gamma-ray community (Heck et al., 1998).

CORSIKA tracks the propagation, interaction, and decay of every secondary particle produced in the cascade, from the first interaction of the primary in the upper atmosphere down to a configurable observation level. The cascade is treated as a set of coupled components — hadronic, electromagnetic, and muonic — each governed by the relevant physics. Hadronic interactions are split into two energy regimes handled by different interaction models matched at a transition energy. The high-energy regime — the first, most energetic collisions of the primary and its leading secondaries — is simulated with the EPOS model (Pierog & Werner, 2009), which provides a comprehensive treatment of minijet production, string fragmentation, and nuclear effects. The low-energy regime — the more numerous, softer interactions that dominate the later stages of shower development — is handled by the GHEISHA model (Fesefeldt, 1985). The neutral pions produced in these hadronic interactions decay almost immediately into photon pairs and feed the electromagnetic component, whose pair-production and bremsstrahlung sub-cascades are followed explicitly down to user-defined energy thresholds, below which particles are no longer tracked. The atmosphere is modeled as a series of layers with exponentially varying density, and the observation level is set to 5,300 m above sea level to match the planned altitude of CONDOR.

Conceptually, CORSIKA provides a full Monte Carlo realization — with all the stochastic fluctuations of real showers — of the average cascade development captured analytically by the Heitler model introduced earlier in this chapter. Whereas the Heitler model yields the mean depth of shower maximum and the mean particle number as smooth functions of primary energy, CORSIKA samples individual showers event by event, reproducing the shower-to-shower variability and the correlations between observables that the reconstruction model must learn to exploit.

The simulations cover proton and gamma-ray primaries at three discrete energies, 300, 500, and 800 GeV, with zenith angles sampled at uniform intervals from 0° to 40° in steps of 2°, producing 21 discrete inclinations, and the azimuthal angle fixed at $\phi = 0°$. Events producing fewer than 30 secondary particles at the observation level are discarded. The full dataset construction and balancing procedure is described in Chapter 3.

2.6 Convolutional Neural Networks

Convolutional neural networks (CNNs) are a class of deep learning architectures designed to process data with local spatial or temporal structure (LeCun & Bengio, 1998). Their defining property is the use of shared learned filters applied across the input, which allows the network to detect the same local pattern regardless of where it appears in the sequence. In this thesis, a CNN backbone is used to extract local temporal features from the ordered sequence of detector hits that constitutes the primary input to the reconstruction model.

The convolution operation. For a one-dimensional input sequence $\mathbf{x} \in \mathbb{R}^{T \times C_{\text{in}}}$, where T is the sequence length and C_{in} is the number of input channels, and a learned filter $\mathbf{w} \in \mathbb{R}^{k \times C_{\text{in}} \times C_{\text{out}}}$ of kernel size k, the output at position t for output channel j is

$$y_j(t) = \sum_{c=1}^{C_{\text{in}}} \sum_{i=0}^{k-1} w_{i,c,j} \cdot x_{t+i,c} + b_j, \tag{2.6}$$

where b_j is learnable bias. The filter $\mathbf{w}$ is shared across all positions t, which gives CNNs two important properties: translational equivariance, meaning that a pattern detected at one position will also be detected if it appears elsewhere in the sequence, and a substantially reduced parameter count compared to a fully connected network operating on the same input.

Pooling and normalization. Following each convolution, max pooling reduces the temporal resolution by selecting the maximum activation within non-overlapping windows of size p,

$$z_j(t) = \max_{0 \le i < p} y_j(tp + i) \ , \tag{2.7}$$

retaining the presence and magnitude of detected features while discarding fine positional detail. Batch normalization (Ioffe & Szegedy, 2015) then standardizes each feature map across the batch dimension before applying learnable scale and shift parameters,

$$\hat{y}_j = \frac{y_j - \mu_j}{\sqrt{\sigma_j^2 + \epsilon}}, \quad \tilde{y}_j = \gamma_j \hat{y}_j + \beta_j \tag{2.8}$$

where μ_j and σ_j^2 are the batch mean and variance for channel j, and γ_j, β_j are trainable. Batch normalization stabilizes training and permits higher learning rates without destabilizing the optimization. It was originally motivated as a way to reduce internal covariate shift, although subsequent work attributes its effect primarily to a smoothing of the optimization landscape rather than to covariate-shift reduction.

Activation. The nonlinearity applied after normalization is the Exponential Linear Unit (ELU) (Clevert et al., 2015),

$$\mathrm{ELU}(x) = \begin{cases} x & \text{if } x > 0, \\ \alpha(e^x - 1) & \text{if } x \leq 0, \end{cases} \tag{2.9}$$

where the scale parameter α controls the saturation value for negative inputs; the standard choice $\alpha = 1$ is used, which makes the activation continuously differentiable at $x = 0$ (the left derivative $\alpha e^x|_{x=0} = \alpha$ matches the right derivative of 1) and sets the negative saturation level to -1. Unlike the rectified linear unit (ReLU), ELU produces negative outputs for negative inputs, preserving a non-zero mean activation and accelerating convergence in deep networks by reducing the bias shift effect.

The specific CNN architecture used in this work, including the number of layers, filter configurations, and projection heads, is detailed in Chapter 3.

2.7 Transformer Architectures

The Transformer is a deep learning architecture introduced by Vaswani et al. (2017) that processes sequences through scaled dot-product attention. Unlike recurrent architectures such as long short-term memory networks (Hochreiter & Schmidhuber, 1997), which propagate information sequentially through a hidden state and struggle with long-range dependencies, the Transformer computes all pairwise interactions between sequence positions in parallel, making it both more expressive for long-range structure and more efficient on modern hardware. In this thesis, Transformer streams process both the CNN-extracted feature sequence and the raw masked hit sequence to capture global temporal dependencies across the full shower hit pattern; the full multi-stream configuration is detailed in Chapter 3.

Scaled dot-product attention. The input to the attention mechanism is a sequence of *token embeddings*: vector representations of the discrete elements of the sequence (in this work, each token corresponds to one detector hit, encoded either as its raw feature vector or as a CNN-derived feature embedding). Three learned linear projections map this sequence to queries, keys, and values: the *query* of a position asks "what information am I looking for?", the *key* of every other position answers "what information do I expose?", and the *value* carries the actual content that will be propagated. Concretely, the three projections produce $\mathbf{Q} \in \mathbb{R}^{T \times d_k}$, $\mathbf{K} \in \mathbb{R}^{T \times d_k}$, and $\mathbf{V} \in \mathbb{R}^{T \times d_v}$, and the attention output is

$$\mathrm{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V}. \tag{2.10}$$

The matrix $\mathbf{Q}\mathbf{K}^\top/\sqrt{d_k}$ contains the scaled similarity score between every pair of positions. After the softmax, each row becomes a probability distribution over the

sequence, and the output at each position is a weighted sum of the value vectors. The scaling by $1/\sqrt{d_k}$ prevents the dot products from growing large in magnitude as d_k increases, which would push the softmax into regions of near-zero gradient. The resulting attention weights α_{ij} admit a direct reading in terms of the computation itself: a large α_{ij} indicates that position i draws strongly on position j when forming its own representation. This is a statement about how information flows inside the model, and it carries no claim about the physical process that produced the shower. These weights do not constitute an explanation in themselves, but they expose an internal quantity of the model that can be inspected and analyzed; this potential to be probed for interpretability — rather than interpretability as a given — is one of the primary motivations for using attention in a physics context. The analysis of attention patterns, together with the caveats on its faithfulness as an explanation, forms part of the explainability study in Chapter 4.

The Transformer encoder block. A full encoder block combines attention with a position-wise feed-forward network applied independently to each position,

$$\mathrm{FFN}(\mathbf{x}) = \mathrm{ReLU}(\mathbf{x}\mathbf{W}_1 + \mathbf{b}_1)\mathbf{W}_2 + \mathbf{b}_2 \tag{2.11}$$

and wraps both sub-layers with residual connections (He et al., 2016), which add the sub-layer input back to its output so that gradients can flow directly through the network and training does not degrade as depth increases, and layer normalization (Ba et al., 2016), which rescales each token's activations to zero mean and unit variance before the sub-layer is applied, stabilizing the magnitude of the representations across positions and layers. This work uses the pre-norm variant, in which normalization is applied before each sub-layer rather than after,

$$\begin{aligned} \mathbf{x}' &= \mathbf{x} + \mathrm{Attention}(\mathrm{LN}(\mathbf{x})) \ , \\ \mathbf{x}'' &= \mathbf{x}' + \mathrm{FFN}(\mathrm{LN}(\mathbf{x}')) \ , \end{aligned}$$

where LN denotes layer normalization. The pre-norm formulation has been shown to improve training stability in deep Transformer models (Xiong et al., 2020) by ensuring that the residual path carries a well-conditioned gradient signal regardless of network depth.

The specific Transformer configuration used in this work — including the number of streams, their inputs, and the design choices motivated by interpretability — is detailed in Chapter 3.

On positional encoding. The standard Transformer architecture uses additive positional encodings to inject sequence order information, since the attention mechanism is inherently permutation-invariant. In this work, explicit positional encodings are not used because the input sequence already contains the temporal position

of each hit as a feature (`t_bin`) and the spatial coordinates (`x_center`, `y_center`, `detector_id`). These features provide the model with direct access to both temporal and spatial ordering within each event, rendering separate positional encodings redundant. Additionally, the CNN backbone processes the time-ordered sequence before the Transformer streams, embedding local temporal structure into the feature representations.

2.8 Multi-Task Learning

Multi-task learning (MTL) is a training paradigm in which a single model is optimized to perform multiple related tasks simultaneously, sharing internal representations across them (Caruana, 1997). Rather than training independent models for each objective, MTL exploits the joint learning signal from all tasks, which can improve generalization and reduce overfitting when the tasks are physically correlated. In EAS reconstruction, gamma-hadron discrimination, zenith angle estimation, and primary energy estimation are correlated in exactly this sense: the particle type, arrival direction, and energy all influence the same shower observables, and a representation that captures one aspect of the shower is likely to be informative for the others.

Loss formulation. The total training loss is a weighted sum of the individual task losses,

$$\mathcal{L}_{\text{total}} = \sum_{k=1}^{K} \lambda_k \mathcal{L}_k \ , \tag{2.12}$$

where $\mathcal{L}_k$ is the loss for task k and λ_k controls its relative contribution to the gradient updates. For gamma-hadron discrimination, the binary cross-entropy loss is used,

$$\mathcal{L}_{\text{cls}} = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)] \ , \tag{2.13}$$

where $y_i \in \{0, 1\}$ is the true label and $\hat{y}_i \in (0, 1)$ is the predicted probability from a sigmoid activation. For zenith angle reconstruction and energy estimation, the Huber loss is used (Huber, 1992),

$$\mathcal{L}_{\text{Huber}}(r) = \begin{cases} \frac{1}{2} r^2 & \text{if } |r| \leq \delta \ , \\ \delta\big(|r| - \frac{\delta}{2}\big) & \text{if } |r| > \delta \ , \end{cases} \tag{2.14}$$

where $r = \hat{y}_i - y_i$ is the residual and δ is threshold parameter. The Huber loss behaves as squared error for small residuals, providing a smooth gradient signal near the optimum, but transitions to linear error for large residuals, limiting the influence of

outlier events that arise from shower-to-shower fluctuations. This robustness makes it more appropriate than mean squared error for the regression tasks in this setting.

Task weights and regularization. The choice of λ_k is non-trivial: if one task dominates the gradient signal, the shared encoder will be biased toward that objective at the expense of the others. The weights used in this work, $\lambda_{\text{cls}} = 0.6$, $\lambda_\theta = 1.0$, and $\lambda_E = 1.5$, were determined through Bayesian hyperparameter search as described in Section 3.5. The higher weight on energy estimation reflects the difficulty of that task and prevents the classification objective, which produces smaller raw loss values early in training, from monopolizing the shared encoder updates. Task-specific dropout is applied with rate 0.35 for the classification head, no dropout for the zenith angle head, and rate 0.5 for the energy head, reflecting the different regularization requirements of each task.

2.9 Explainable Artificial Intelligence

The adoption of deep learning in experimental physics introduces a fundamental tension between predictive performance and scientific transparency (Ribeiro et al., 2016). A model that achieves high reconstruction accuracy without revealing which physical features drive its predictions cannot be fully trusted in a scientific context, because there is no way to verify that the learned representations correspond to genuine shower physics rather than spurious correlations in the simulation. Explainable artificial intelligence (XAI) refers to the set of methods developed to address this problem by providing interpretable descriptions of model behavior. In this thesis, two complementary techniques are applied across all three reconstruction tasks simultaneously.

2.9.1 Permutation Feature Importance

Permutation feature importance (PFI) is a model-agnostic method — that is, one that treats the model as a black box and does not require access to its internal architecture or gradients — for quantifying the contribution of each input feature to a model's predictive performance (Altmann et al., 2010). The core idea is to measure how much performance degrades when the values of a given feature are randomly permuted across the dataset, breaking any statistical association between that feature and the target. Formally, let S denote a performance metric evaluated on the original dataset and $S_j^{(\pi)}$ the same metric after randomly permuting feature j. The importance of feature j is defined as

$$I_j = S - \mathbb{E}_\pi\left[S_j^\pi\right] , \tag{2.15}$$

where the expectation is taken over multiple random permutations. A large positive I_j indicates that the model relies heavily on feature j; a value near zero indicates that the feature contributes little to the predictions. In this work, PFI is computed separately for each of the three reconstruction tasks, which allows the analysis to reveal whether different tasks prioritize different input features. Alignment between the importance rankings and physical expectations about which channels should be informative for each objective constitutes a form of validation that the learned representations are physically consistent.

2.9.2 Feature Ablation

Feature ablation provides a complementary perspective to PFI. Whereas PFI breaks the statistical association between a feature and the target by random permutation across events, ablation neutralizes the feature in a controlled way — here, by replacing its values with the dataset median — and measures the resulting drop in task performance. Both techniques estimate input relevance without modifying the model, but they probe different counterfactuals: PFI asks how the model behaves when the feature is replaced by an arbitrary sample from its marginal distribution, while ablation asks how it behaves when the feature is reduced to a single, representative constant. Agreement between the two rankings provides cross-validation of feature importance and reduces sensitivity to the artifacts of any single method. The ablation procedure used in this work is applied jointly across the three reconstruction tasks and is described in detail in Chapter 4.

2.9.3 Attention Pattern Analysis

The attention weights α_{ij} produced by the Transformer streams, introduced in Section 2.7, have a natural interpretability property: they explicitly quantify the degree to which temporal position i in the input sequence attends to position j when forming its output representation. Because the input sequence consists of detector hits ordered by arrival time, each position in the sequence corresponds to a specific phase of the shower front's development. Analyzing the attention patterns therefore amounts to asking which temporal phases of the shower the model considers most informative for each reconstruction task.

Unlike PFI, which operates at the level of input feature channels averaged over the entire dataset, attention analysis reveals position-specific and event-specific patterns: for a given event, the attention matrix shows precisely where the model focuses when performing each task. Aggregating these matrices over the test set provides a statistical picture of the model's temporal focus for each objective, which can be

compared against physical expectations. For example, the arrival-time structure of the shower front is expected to carry the directional signal used for zenith angle reconstruction, while the total particle count and lateral spread are expected to be more relevant for energy estimation. It should be noted that the use of attention weights as model explanations has been the subject of an ongoing methodological debate. On one side, Jain and Wallace (2019) show that, in several architectures, attention distributions can be substantially altered without affecting predictions, casting doubt on their faithfulness as explanations; subsequent work has both refined this critique and identified conditions under which attention can still serve as a useful diagnostic of model behavior. The position adopted in this thesis is therefore not that attention weights constitute a formal explanation by themselves, but that they expose an internal model quantity worth analyzing alongside complementary methods. In the present context the use of single-head attention per task produces unambiguous weight distributions, and the agreement between attention patterns and the independent feature importance methods (PFI and ablation) provides mutual validation. The combination of these three complementary views of model behavior forms the explainability framework analyzed in detail in Chapter 4.

2.10 Reconstruction of EAS: State of the Art

The reconstruction of the properties of a primary particle from the ground-level footprint of its shower has been approached in two ways. The first fits the recorded pattern with functional forms derived from shower physics; the second learns the mapping from simulated events. This section describes both, and states the gap that this thesis addresses. The quantitative comparison of the results obtained here against published figures is given in Section 4.1.4.

2.10.1 Classical Parametric Reconstruction

Traditional pipelines estimate each shower property with a dedicated fit. The arrival direction follows from the arrival times recorded across the array: assuming a planar front travelling at the speed of light, the direction is obtained by fitting Equation 2.5 to the measured times, with curvature corrections for showers whose front departs appreciably from planarity (Collaboration et al., 2014). The core position and the shower size are obtained by fitting the NKG function of Equation 2.4 to the lateral density profile, and the primary energy is then inferred from the fitted particle density at a fixed distance from the core, a quantity chosen because it is comparatively insensitive to shower-to-shower fluctuations (Kawata et al., 2017). In arrays with sensitivity to the muonic component, gamma-hadron discrimination is performed by

cutting on the estimated muon content, which is larger in hadronic showers, or on shape parameters of the footprint.

These methods are computationally light, their assumptions are explicit, and their failure modes are understood. Their limitations become relevant precisely in the regime that CONDOR targets. Each property is estimated separately, so the correlations between them are not exploited. The functional forms describe averages over shower-to-shower fluctuations, and those fluctuations are largest at low primary energy. The fits also degrade when few modules record a signal, which is the normal situation for a sub-TeV shower. The classical reconstruction reported for CONDOR attains an angular resolution of approximately 1–2° (Arratia et al., 2025).

2.10.2 Deep Learning for Air Shower Reconstruction

Deep learning replaces the fitted functional form with a mapping learned directly from simulated events. Erdmann et al. (2018) applied a convolutional network to the signal and timing traces of the Pierre Auger Observatory, at 1,400 m and over 3–100 EeV, obtaining an angular resolution of 1.2° for the arrival direction alone. Okukawa et al. (2024) trained neural networks for gamma-hadron separation on the Tibet ASγ scintillator array at 4,300 m over 10–100 TeV, reporting AUC values between 0.781 and 0.901 for a single-task classifier. Meng et al. (2025) addressed energy reconstruction for the same experiment over 1 TeV–10 PeV, again as a separate task. Closer to the energy window of CONDOR, Conceição et al. (2025) discriminated gamma- from hadron-induced showers with a pretrained vision transformer applied to the footprint of a water-Cherenkov array at 5,200 m over 200 GeV–1 TeV, reporting a proton rejection near 97% at a gamma efficiency of 80%. Transformer architectures have also been applied to composition-sensitive observables in hybrid arrays (Langner et al., 2024); their general properties as sequence models are reviewed by Lin et al. (2022).

The closest precedent to this thesis is its own predecessor. Navarro et al. (2025) introduced a CNN-LSTM model for CONDOR that reconstructed particle type and zenith angle jointly over 150–800 GeV, reaching an F1-score of 0.89 and an angular MAE of 1.14°. That model established the multi-task formulation for this detector, but it processed the hit sequence recurrently, did not estimate the primary energy, and included no explainability analysis. The architecture developed here, whose three-stream version has been published in Navarro et al. (2026a), addresses those three points.

Three characteristics are common to this body of work. First, the reconstruction tasks are almost always addressed one at a time, so the physical correlations between particle type, arrival direction and energy are left unexploited; the exception

is Navarro et al. (2025), which covers two of the three. Second, the detector configurations differ from CONDOR in altitude, in technology, or in both, which limits how directly a published figure transfers: a network trained on a water-Cherenkov footprint does not describe a scintillator array, and a resolution quoted at EeV energies says little about the sub-TeV regime. Third, the representation learned by the network is rarely examined; performance is reported, but which physical observables the model has come to rely on is left open.

This thesis addresses the three points together, for one detector and with one model: the three reconstruction tasks are solved simultaneously in the 300–800 GeV range at 5,300 m, and the representation the model arrives at is examined with feature attribution and attention analysis.

Chapter 3

Methodology

This chapter describes the complete methodological pipeline developed for the reconstruction of EAS at CONDOR using deep learning. It covers the experimental conditions under which the data were generated, the construction and preprocessing of the dataset, the architecture of the proposed model, the strategy used to identify its optimal configuration, and the training procedure that resulted from that search. Together, these components constitute the full technical contribution of this thesis.

3.1 Experimental Setup

The experimental conditions of this work are determined by the design and operational constraints of the CONDOR observatory, described in detail in Section 2.4, and the simulation framework introduced in Section 2.5. Because the observatory has not yet begun scientific operations, all data are generated through CORSIKA simulations configured to match CONDOR's geometry and high-altitude observation level of 5,300 m above sea level.

Two primary particle types are simulated, protons and gamma rays, at three discrete energies: 300, 500, and 800 GeV. This discrete sampling was a deliberate design choice to maintain adequate statistical representation per energy bin while limiting the overall computational cost of the simulation campaign, and it challenges the model to generalize to continuous energy values between training points. Zenith angles are sampled from 0° to 40° in steps of 2°, producing 21 discrete inclinations, while the azimuthal angle is fixed at $\phi = 0$ for all simulations. This is a deliberate simplification that reduces the simulation cost and isolates the zenith-angle dependence.

It should be noted that the array geometry offers little symmetry to lean on: in the proposed CONDOR layout, both the individual modules and the central grid they form are rectangular rather than square in their two horizontal dimensions, so the

array maps onto itself only under a 180° rotation and possesses two-fold rather than four-fold or continuous rotational symmetry. The fixed azimuth therefore introduces a directional bias that is examined explicitly in Chapter 4. Events producing fewer than 30 secondary particles at the observation level are discarded, removing low-multiplicity cases that carry insufficient information for reliable reconstruction.

3.2 Dataset Construction and Preprocessing

The raw output of the CORSIKA simulations is particle-level: for each event it records the secondary particles reaching the observation level, with their ground positions, arrival times, and energies, rather than any notion of detector modules. The first preprocessing step aggregates these particle hits onto the 120 detector modules of the array, so that each event is subsequently described by the set of modules that registered at least one hit during the shower passage, together with the physical quantities recorded at each active module. The simulations employ the EPOS model (Pierog & Werner, 2009) for high-energy hadronic interactions and the GHEISHA model (Fesefeldt, 1985) for low-energy hadronic and electromagnetic processes, following the same configuration used by Arratia et al. (2025). Transforming this raw output into a structured representation suitable for deep learning requires two steps: the construction of a dual-input encoding that captures complementary aspects of each shower event, and the normalization of all input features to a common scale.

Sequential Features. The first input branch encodes the spatiotemporal development of the shower front across the detector array. For each event, all active detector hits are collected and sorted in ascending order of their recorded time bin. Each hit is represented as a six-dimensional vector containing the detector identifier, the particle count registered at that module, the time bin index t_{bin}, the total deposited energy ε, and the horizontal coordinates x_{center} and y_{center} of the module's position within the array. The resulting input for a single event is a matrix of shape $n \times 6$, where n is the number of active detector hits ordered chronologically. This ordering is physically meaningful: the earliest hits correspond to the leading edge of the shower front, which carries the strongest directional signal, while later hits trace the lateral spread and tail of the shower development. Because n varies across events, sequences are zero-padded to a common maximum length, and a boolean mask is passed alongside each sequence to prevent padded positions from contributing to the attention computation in the Transformer component.

Global Features. The second input branch encodes the aggregate properties of the event as a fixed-length vector of five statistics, in the exact order implemented in the preprocessing pipeline: (1) the total particle count summed across all modules,

(2) the total deposited energy summed across all modules, (3) the number of active detectors, (4) the temporal duration of the event, defined as the difference between the last and first recorded time bins, and (5) the deposited energy recorded by the central detectors of the array. These quantities capture the overall scale and spatial extent of the shower without encoding its internal temporal structure, providing a complementary signal that is particularly informative for energy estimation, where the total shower size is more directly relevant than the sequence of individual hits. The exact computation and normalization of each global feature are detailed in Appendix A.3, and Table 3.1 illustrates the dual-input data structure and the physical meaning of each feature branch.

Table 3.1: Dual-input feature structure of the CNN-Transformer model.

Input	**Shape**	**Feature**	**Description**
Sequential	$(n \times 6)$	ID_i	Detector identifier
		count_i	Particle count
		t_i	Binned arrival time
		ε_i	Deposited energy
		$x_i,\ y_i$	Detector coordinates
Global	(5)	`total_particles`	Total particle count
		`total_energy`	Total deposited energy
		`active_detectors`	Active detector count
		`duration`	Shower duration
		`central_energy`	Central array energy

n: number of hits per event, padded to `maxlen`; sequences masked at padded positions.

Figure 3.1 shows representative individual events for each combination of particle type and primary energy, illustrating the variability in shower morphology at the single-event level. Figure 3.2 shows the corresponding average footprints, computed over roughly 770 events per cell, which suppress stochastic fluctuations and reveal the systematic morphological differences exploited by the model.

Averaging over many events per bin removes this variability and exposes the underlying morphology. Figure 3.2 shows the mean footprints computed over roughly 770 events per cell: gamma-induced showers concentrate the deposited energy in a compact central core, whereas proton-induced showers spread it more widely and irregularly across the array. This morphological asymmetry is the primary discriminant the classification head learns to exploit.

Both input branches are normalized before training. The sequential features are

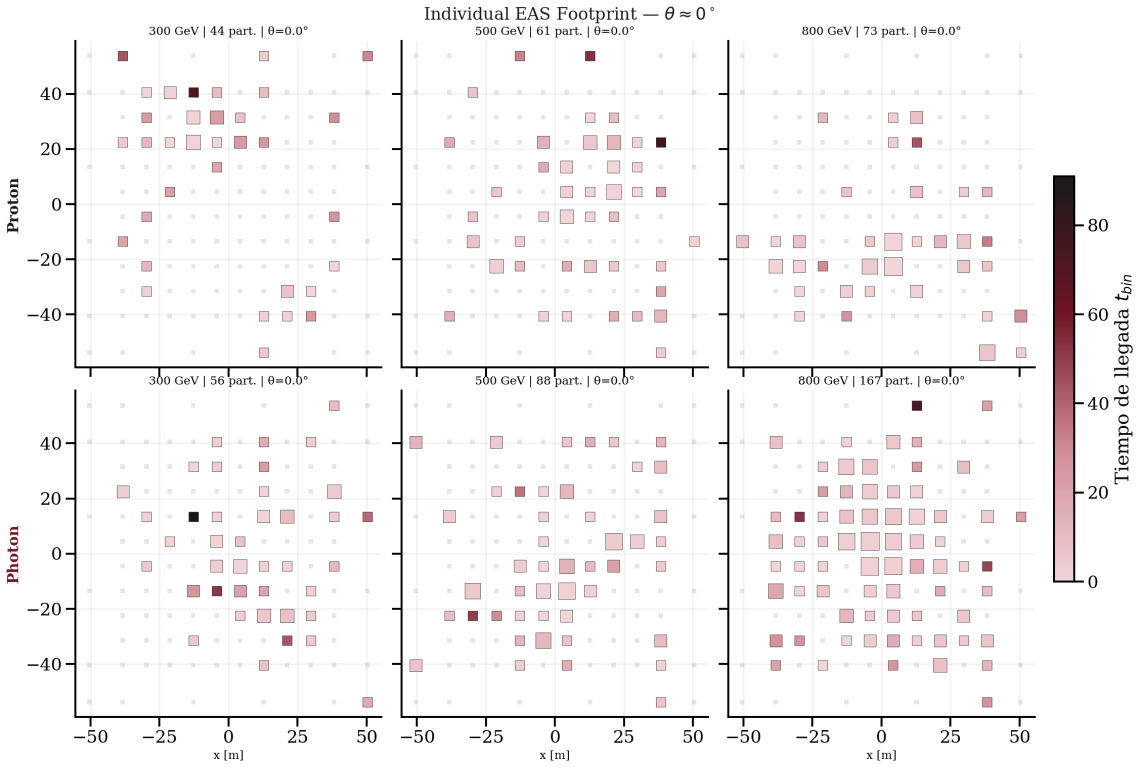


Figure 3.1: Representative individual EAS footprints at $\theta \approx 0°$, for proton-induced (top row) and gamma-induced (bottom row) showers at 300, 500, and 800 GeV primary energy. Each event is selected as the one closest to the median total particle count for its group. Color encodes the arrival time t_{bin} of particles at each detector; marker size is proportional to particle count. Individual events exhibit significant stochastic fluctuations in lateral spread and hit multiplicity.

standardized channel-wise using the mean and standard deviation computed over the training set, and the global features are scaled to the unit interval using training set minimum and maximum values. Normalization statistics are computed exclusively on the training partition and applied without modification to the validation and test sets, ensuring that no information from the evaluation data influences the preprocessing.

3.3 Data Stratification and Balance Strategy

The raw simulated dataset is not uniformly distributed across the space of primary particle types, energies, and zenith angles. This imbalance arises from the interaction between the simulation configuration and the multiplicity filter: showers at large zenith angles or low energies produce fewer secondary particles at detector level on

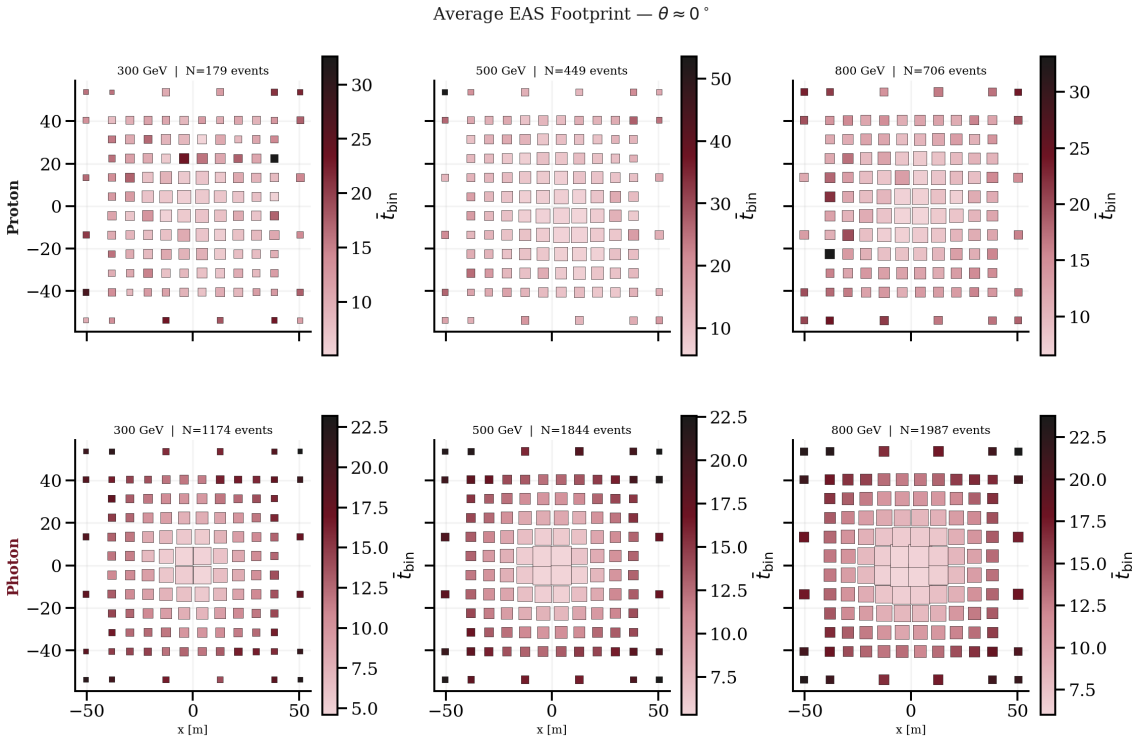


Figure 3.2: Average EAS footprints at $\theta \approx 0°$, computed over all events in each particle-type and energy bin (∼770 events per cell). Averaging suppresses single-event fluctuations and reveals the systematic morphological signatures of each class: gamma-induced showers produce a more compact, centrally concentrated deposit, while proton-induced showers exhibit a broader lateral spread and higher peripheral particle counts. Color encodes the mean arrival time $\bar{t}_{\text{bin}}$; marker size is proportional to mean particle count.

average and are therefore more frequently discarded. Left uncorrected, this imbalance would bias the shared encoder toward the conditions most heavily represented in the training data, degrading performance on underrepresented configurations and compromising the fairness of the multi-task optimization. To address this, a stratified resampling procedure is applied before the dataset split. Let

$$S = \{s_1, s_2, \ldots, s_K\} \tag{3.1}$$

denote the set of all K strata, where each stratum s_i is defined by a unique combination of primary energy $E \in \{300, 500, 800\}$ GeV, particle type $p \in \{\text{proton}, \gamma\}$, and discretized zenith angle bin θ. For each stratum s_i, let n_i denote its event count after multiplicity filtering. The target population size is defined as

$$n_{\text{target}} = \text{median}(\{n_1, n_2, \ldots, n_K\}) \tag{3.2}$$

which evaluates to $n_{\text{target}} = 593$ for this dataset. The median is preferred over the mean as a balancing objective because it is insensitive to outliers and extreme population disparities: under the median criterion, half of the strata are downsampled and half are upsampled, without being influenced by extremely large or small strata. The resampling procedure operates independently on each stratum:

$$n_i^{\text{balanced}} = \begin{cases} \text{undersample}(s_i, n_{\text{target}}) & \text{if } n_i > n_{\text{target}} \\ \text{oversample}(s_i, n_{\text{target}}) & \text{if } n_i < n_{\text{target}} \\ n_i & \text{if } n_i = n_{\text{target}} \end{cases} \tag{3.3}$$

where undersampling randomly selects n_{target} events from the n_i available events without replacement, effectively removing $n_i - n_{\text{target}}$ events. Oversampling duplicates events with replacement until the stratum size reaches n_{target}, introducing $n_{\text{target}} - n_i$ duplicate samples. This ensures that after resampling, all strata satisfy $n_i^{\text{balanced}} = n_{\text{target}}$.

The resulting balanced dataset contains approximately 74,700 events in total. These are split into training, validation, and test partitions in a ratio of approximately 70:15:15, yielding roughly 52,300 training events, 11,200 validation events, and 11,200 test events. The split is performed after stratified resampling and maintains the balanced distribution across all three partitions. No event appears in more than one partition, and the test set is held out entirely until final evaluation. Figure 3.3 illustrates the original event distributions per energy and the resampling target, and Table 3.2 summarizes the dataset composition before and after balancing.

Table 3.2: Dataset composition before and after balancing. γ: # gamma events, **p**: # proton events.

E (GeV)	γ	**p**	**Total**
	Original		
300	14474	1958	**16432**
500	27611	5306	**32917**
800	36143	11496	**47639**
	Balanced		
300	12453	12453	**24906**
500	12453	12453	**24906**
800	12453	12453	**24906**

Figure 3.3 illustrates the effect of the resampling strategy across the three marginal distributions. After balancing, the label distribution is equalised between proton and

gamma classes, the zenith angle distribution becomes uniform across the full 0°–40° range, and the energy distribution is balanced across the three primary energy levels. The resulting dataset satisfies $n_i^{\text{balanced}} = n_{\text{target}}$ for all $i \in \{1, \ldots, K\}$, ensuring equitable representation across all physics configurations.

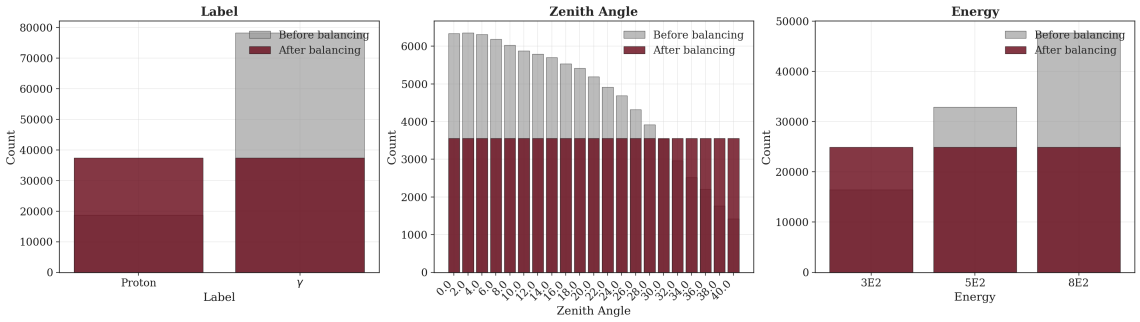


Figure 3.3: Distribution of events by label, zenith angle, and primary energy before (grey) and after (dark red) stratified balancing. Grey bars extending above the dark red bars indicate categories that were undersampled to reach the median target of 593 events per stratum; categories where both bars align were already close to the target or were oversampled.

This combined over- and undersampling approach ensures that each physics configuration is represented equitably during training. The final balanced dataset, whose composition by energy and particle type is summarized in Table 3.3, helps the model to develop unbiased decision boundaries across the entire parameter space.

Table 3.3: Dataset composition after stratified balancing, showing the number of events by primary energy, particle type, and split. Energy labels: 3E2 = 300 GeV, 5E2 = 500 GeV, 8E2 = 800 GeV.

Split	**3E2-γ**	**3E2-p**	**5E2-γ**	**5E2-p**	**8E2-γ**	**8E2-p**	**Total**
Train	8678	8684	8728	8718	8745	8749	**52302**
Val	1892	1894	1838	1875	1872	1836	**11207**
Test	1883	1875	1887	1860	1836	1868	**11209**

A methodological note on the interaction between oversampling and the data split is warranted. Because the resampling procedure is applied before the split into training, validation, and test partitions, oversampled strata contain duplicated events. The splitting procedure assigns each event instance to exactly one partition, but for strata where $n_i < n_{\text{target}}$, different copies of the same physical event may appear in different partitions, creating a potential information leak between training and test data.

The magnitude of this effect is not negligible and is strongly class-dependent. As shown in Table 3.2, the gamma class is abundant at all three energies (78,228 events before balancing) and is therefore downsampled in every stratum, so all gamma events in the balanced dataset are unique. The proton class, by contrast, is scarce (18,760 events before balancing) and falls below the median target in every stratum, so all proton strata are oversampled. Consequently the balanced dataset contains $37{,}359 - 18{,}760 = 18{,}599$ duplicated proton instances, which amounts to roughly 25% of the total dataset and approximately half of the proton class. Duplicate copies of a given proton event are distributed across the partitions, so a proton event in the test set frequently has an identical copy in the training set.

This asymmetry has a concrete implication for interpretation: it can selectively inflate the apparent performance on proton events. In particular, the markedly better energy resolution obtained for protons (16–18%) compared to gammas (28–41%, Section 4.1.3) is attributed below to shower physics, but a fraction of the gap could also reflect partial memorization of duplicated proton events. The gamma-class metrics are not subject to this effect, since gamma events are never duplicated. This caveat should be kept in mind when reading the per-class results. Alternative strategies that avoid the leak entirely — performing the split before resampling, undersampling instead of oversampling, or replacing duplication with class-weighted losses or post-split augmentation — would remove this ambiguity and represent a natural direction for future refinement.

3.4 Model Architecture

The reconstruction model is a hybrid CNN-Transformer architecture designed for multi-task learning over dual-input shower data. Its design reflects three requirements: the ability to extract local temporal features from the sequential input, the ability to model global dependencies across the full hit sequence, and the flexibility to specialize independently for each of the three reconstruction tasks without forcing a single shared representation to serve all objectives equally. A schematic diagram of the full architecture is shown in Figure 3.4.

CNN backbone. The sequential input is first processed by a convolutional backbone consisting of three successive Conv1D layers, each with 128 filters, kernel size 7, same-padding, and ELU activation. Batch normalization is applied after each convolutional layer. After the third convolutional layer, a single max-pooling operation with stride 2 halves the temporal resolution before the representation is passed to the projection layers. The three-layer structure progressively expands the effective receptive field: the first layer captures local patterns within a handful of consecutive hits, while the third integrates information over a wider temporal

window. The max-pooled output is then projected through three fully connected layers of dimensions 128, 256, and 128 with ELU activations, and dropout of rate 0.4 is applied after the 256-dimensional layer.

Task-specific Transformer streams. The architecture employs six independent Transformer streams, two per reconstruction task. For each task, one stream operates on the 128-dimensional CNN output sequence (the *CNN stream*), and a second stream operates directly on the raw masked six-dimensional hit sequence (the *RAW stream*). Each stream applies a single pre-norm Transformer encoder block with one attention head and key dimension $d_k = 32$, as described in Section 2.7. The CNN streams capture high-level temporal patterns abstracted by the convolutional backbone. The RAW streams attend directly to individual detector hits, enabling attention maps that can be interpreted in terms of specific physical observables, such as which detectors and time bins the model prioritizes for each task, without the indirection introduced by convolutional transformations. Using two dedicated streams per task ensures that each reconstruction objective can develop an independent attentional strategy without competing with the others for a shared mechanism.

Fusion and output heads. The output of each stream is summarized by global average pooling. For each task, the pooled CNN-stream vector (128 dimensions) and the pooled RAW-stream vector (6 dimensions, matching the raw feature depth) are concatenated to form a 134-dimensional fused attention representation, which is then concatenated with the five-dimensional global feature vector to yield a 139-dimensional fused representation. This representation is passed through a task-specific dense layer of 128 units with ELU activation and then mapped to the final prediction: a sigmoid-activated scalar for gamma-hadron classification and a linear-activated scalar for each regression task. Task-specific dropout is applied before the dense head, with rates of 0.35 for classification, no dropout for zenith angle regression, and 0.5 for energy estimation. The total model contains 531,375 parameters, of which 530,607 are trainable and 768 are non-trainable batch normalization statistics.

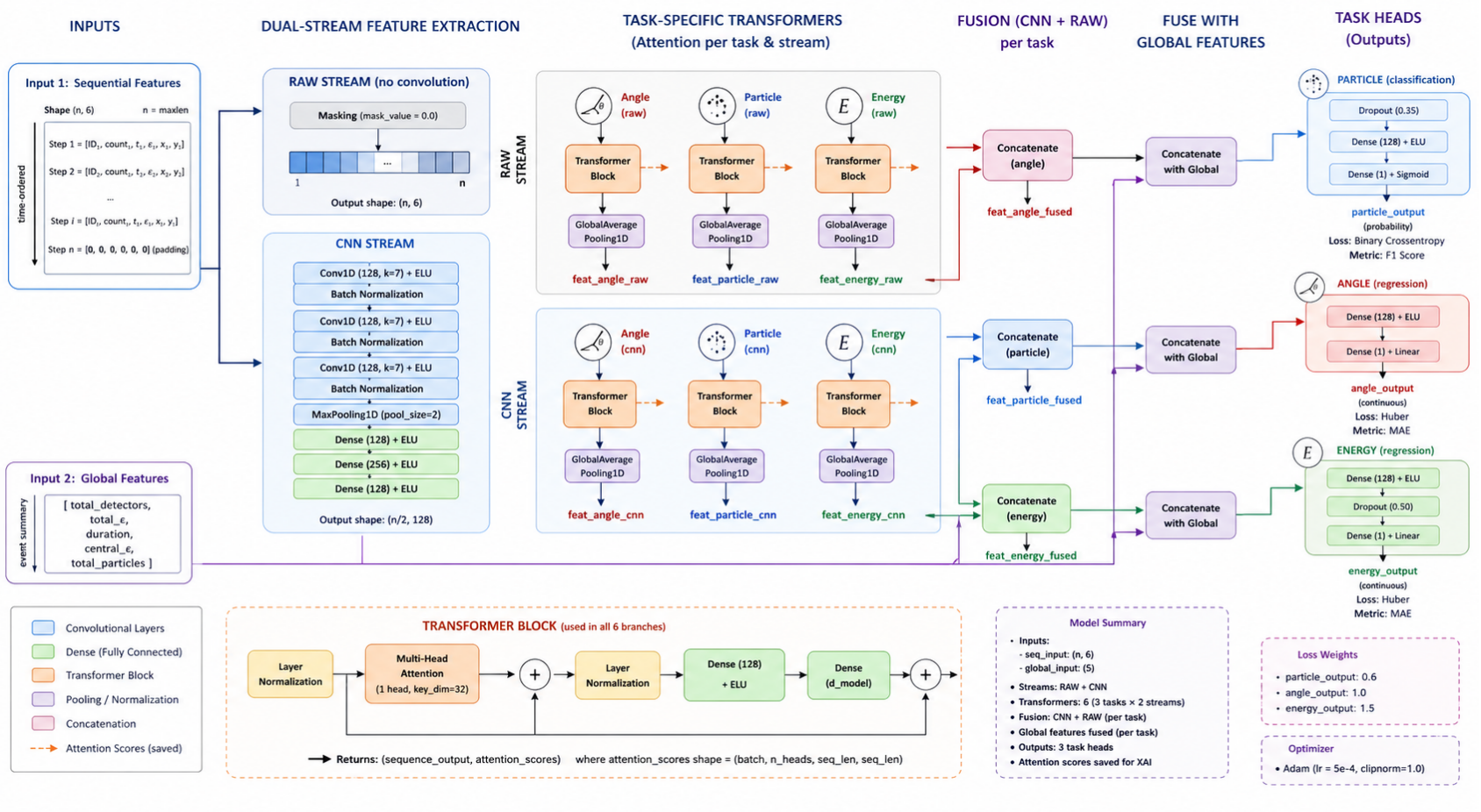


Figure 3.4: Flow diagram of the CNN-Transformer multi-task model.

3.5 Hyperparameter Optimization

The model contains a number of hyperparameters that influence training dynamics and final performance, including the learning rate, task loss weights, and the dropout rates for the classification and energy heads. An initial search was conducted using Optuna (Akiba et al., 2019), a framework for automated hyperparameter optimization based on Bayesian search with tree-structured Parzen estimators (TPE) (Bergstra et al., 2011). TPE models the distributions of hyperparameter configurations that lead to good and poor outcomes separately, and uses the ratio of these distributions to propose new trials that are more likely to improve the objective. This approach is substantially more sample-efficient than grid or random search, particularly when the hyperparameter space is high-dimensional.

The search space covered the learning rate over the range $[10^{-4}, 10^{-2}]$ on a logarithmic scale, the task loss weights λ_{cls}, λ_θ, and λ_E each over $[0.1, 2.0]$, and the dropout rates for the classification and energy heads independently over $[0.1, 0.6]$. Each trial trained the model for a fixed number of epochs and evaluated the combined validation loss as the optimization objective; divergent trials were pruned early using Optuna's built-in median pruner. The Optuna search provided initial estimates for the hyperparameter ranges, which were then refined through targeted manual tuning of the loss weights and dropout rates. The final configuration, reported in Table 3.4, was selected based on validation performance across all three tasks simultaneously.

The final task weights are

$$\begin{aligned} \lambda_{\text{cls}} &= 0.6 \\ \lambda_\theta &= 1.0 \\ \lambda_E &= 1.5, \end{aligned}$$

reflecting the relative difficulty of each task and the need to prevent the binary classification objective from dominating the regression signals early in training.

3.6 Training Configuration

The model is trained by minimizing the weighted multi-task loss:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{cls}}\mathcal{L}_{\text{cls}} + \lambda_\theta\mathcal{L}_\theta + \lambda_E\mathcal{L}_E \tag{3.4}$$

where $\mathcal{L}_{\text{cls}}$ is the binary cross-entropy loss for gamma-hadron classification and $\mathcal{L}_\theta$, $\mathcal{L}_E$ are Huber losses for zenith angle and energy regression respectively, as defined in Section 2.8. Optimization is performed using the Adam optimizer (Kingma & Ba, 2015) with learning rate 5×10^{-4} and gradient norm clipping at 1.0. The learning rate

Table 3.4: Consolidated hyperparameter configuration for the final model.

Category	**Parameter**	**Value**
CNN backbone	Conv1D filters	128
	Kernel size	7
	CNN dropout	0.4
Transformer	Attention heads	1
	Key dimension d_k	32
Task weights	λ_{cls}	0.6
	λ_θ	1.0
	λ_E	1.5
Task dropout	Classification	0.35
	Angle	0.0
	Energy	0.5
Training	Optimizer	Adam
	Learning rate	5×10^{-4}
	Gradient clip norm	1.0
	LR reduction patience	7 epochs
	Early stopping patience	15 epochs
Data	Batch size	64
	Max sequence length	472

is reduced by a factor of 0.5 whenever the combined validation loss fails to improve for seven consecutive epochs, with a minimum learning rate floor of 10^{-6}. This plateau reduction schedule allows the optimizer to take larger steps during the initial training phases and shift to finer adjustments as the model approaches convergence, which proved effective for balancing the three task objectives simultaneously.

Training is stopped early when the combined validation loss fails to improve for 15 consecutive epochs, and the model checkpoint achieving the lowest validation loss is restored for final evaluation. No information from the test set is used at any point during training or hyperparameter search. Table 3.4 consolidates all hyperparameters of the final configuration.

All experiments were conducted on a single NVIDIA GeForce RTX 3050 Ti Laptop GPU (4 GB VRAM) using TensorFlow 2.10 with CUDA 11.2. A full training run to convergence (61 epochs, with the best checkpoint at epoch 46) required approximately 45 minutes. The total model contains 531,375 parameters, of which 530,607 are trainable and 768 are non-trainable batch normalization statistics.

3.7 Code and Data Availability

The source code of this thesis is publicly available in the repository given below. It contains the data pipeline, the model definition and training notebook, the scripts for every analysis in Chapter 4, the trained model, the result files from which all reported metrics and figures are computed, and the LaTeX source of this document. The processed dataset is archived on Zenodo (Navarro Farías, 2025).

`https://github.com/luisnavarrof/Thesis_CONDOR-EAS-DL-Reconstruction`

Chapter 4

Results and Explainability

This chapter presents the evaluation of the CNN-Transformer model on the held-out test set, together with the full suite of explainability analyses and the detector robustness study. All results correspond to the six-stream architecture described in Chapter 3, which extends the three-stream CNN-Transformer model introduced in Navarro et al. (2026a) with independent RAW attention streams per task, enabling attention maps at the level of individual detector hits. The test partition comprises 11,209 events and was withheld entirely during training and hyperparameter search. Section 4.1 reports quantitative performance across the three reconstruction tasks and compares them with the state of the art. Section 4.2 presents two complementary feature importance analyses and the attention pattern study. Section 4.3 quantifies model robustness to detector failures and derives first data-driven insights on array design.

4.1 Multi-Task Performance

Figure 4.1 shows the training and validation loss and metric curves across epochs. The model trained for 61 epochs before early stopping was triggered, with the best checkpoint at epoch 46 restored for evaluation. All three tasks converge jointly without any task dominating the optimization, and the gap between training and validation curves is narrow throughout, indicating that the regularization strategy is effective.

A note on notation is in order before the results are presented. Throughout this chapter, a circumflex denotes a quantity predicted by the model and the subscript "true" denotes the corresponding Monte Carlo value: $\hat{\theta}$ and θ_{true} are the reconstructed and true zenith angles of an event, and $\hat{E}$ and E_{true} its reconstructed and true primary energies. The *angular residual* of an event is the difference $\hat{\theta} - \theta_{\text{true}}$, and its absolute value $|\hat{\theta} - \theta_{\text{true}}|$ is the absolute angular error.

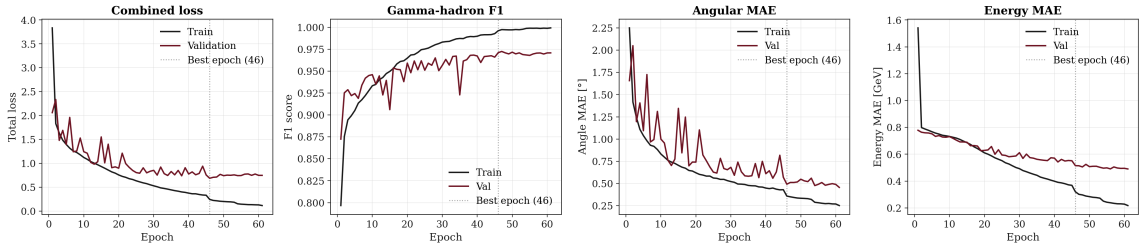


Figure 4.1: Training and validation curves as a function of epoch. From left to right: combined multi-task loss, gamma-hadron F1 score, angular MAE, and energy MAE. The vertical dotted line in each panel marks the best checkpoint epoch (minimum combined validation loss), restored for the final evaluation. All tasks converge jointly without divergence.

Table 4.1 provides a consolidated view of the primary metrics across all three tasks, including domain-specific figures of merit.

Table 4.1: Summary of multi-task reconstruction performance on the test set, including domain-specific figures of merit. $Q = \varepsilon_\gamma / \sqrt{\varepsilon_p}$ is the quality factor for gamma-hadron separation, defined in full in Section A.4. PSF_{68} is the 68% containment radius of the angular residual distribution, defined in full in Section A.5.

Task	**Primary Metric**	**Additional**
Gamma-Hadron	AUC = 0.991	Q-factor = 5.48, $\varepsilon_\gamma = 0.971$
Zenith Angle	MAE = 0.484°	$\mathrm{PSF}_{68} = 0.50°$, RMS = 0.74°
Energy	MAE = 103.72 GeV	$\sigma_E/E = 33.1\%$, $R^2 = 0.453$

4.1.1 Gamma-Hadron Discrimination

Table 4.2 summarizes the classification performance on the test set. The model achieves an area under the receiver operating characteristic (ROC) curve (AUC) of 0.991, an F1-score of 0.970, and an overall accuracy of 96.99%. The quality factor $Q = \varepsilon_\gamma / \sqrt{\varepsilon_p}$, a standard figure of merit in gamma-ray astronomy that quantifies signal-to-noise ratio, evaluates to $Q = 5.48$.

Figure 4.2 shows the row-normalized confusion matrix, with the raw event counts, and the ROC curve. Both classes are reconstructed with high fidelity: 96.9% of proton events and 97.1% of gamma events are classified correctly, and the ROC curve rises steeply to an AUC of 0.991. The few percent of misclassified events

Table 4.2: Classification performance on the test set. Precision, recall, and F1-score are reported for each class. The test partition comprises 11,209 events, 5,606 proton-induced and 5,603 gamma-induced, so the two classes are balanced to within 0.03%.

Class	**Precision**	**Recall**	**F1**
Proton	0.971	0.969	0.970
γ	0.969	0.971	0.970
Overall	Accuracy: **96.99%**		AUC: **0.991**

correspond to the small overlap between the two classes near the decision threshold at $P(\gamma) = 0.5$.

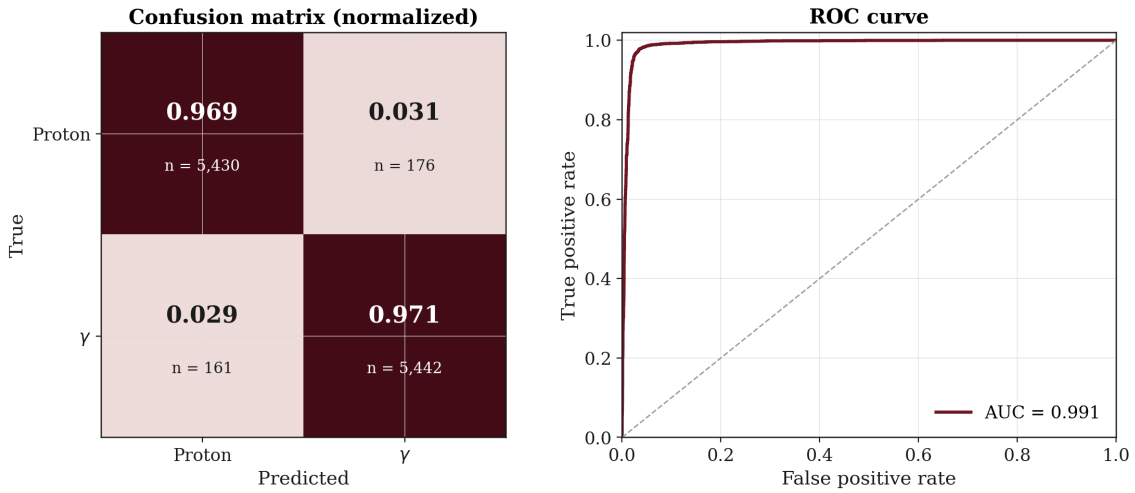


Figure 4.2: Gamma-hadron discrimination on the test set. **Left:** row-normalized confusion matrix, with the raw event count shown below each fraction. **Right:** ROC curve (AUC = 0.991).

Figure 4.3 shows the ROC curves grouped by primary energy and by zenith angle bin: each curve is computed separately within the events sharing a given value of that variable, rather than over the test set as a whole, which isolates how discrimination depends on it. Discrimination improves at lower energies: the AUC is 0.998 at 300 GeV, 0.991 at 500 GeV, and 0.981 at 800 GeV.

This trend is physically consistent. At 300 GeV, electromagnetic showers are compact with a well-defined core signature, while hadronic showers at the same energy are sparse and noisy, making the morphological contrast between the two classes large. As the primary energy increases toward 800 GeV, both shower types develop larger and more complex footprints with greater stochastic fluctuations. In

addition, a growing fraction of the energy of a hadronic shower is channelled into its electromagnetic component through neutral pion decay, the mechanism described in Section 2.2.2. Hadronic events therefore acquire progressively more of the compact, centrally concentrated signature that distinguishes gamma-induced showers in the first place. Both effects reduce the morphological contrast between the two classes.

Across zenith angle bins the AUC ranges from 0.984 to 0.994, so discrimination is essentially independent of shower inclination over the range covered by the training data.

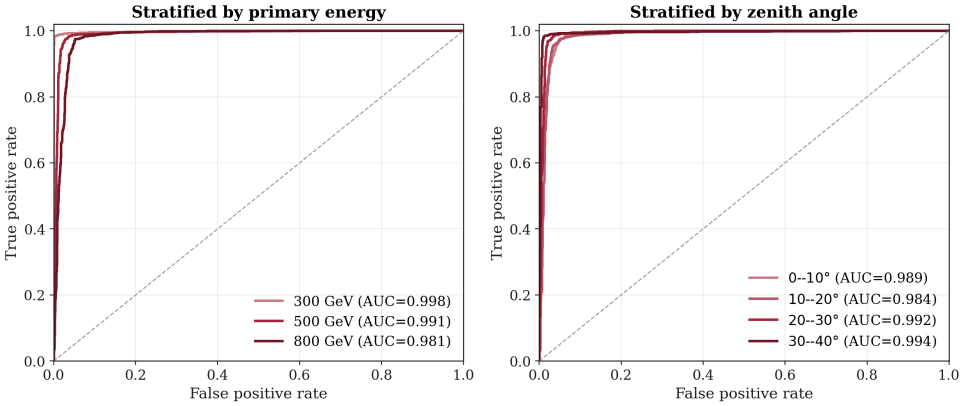


Figure 4.3: ROC curves for gamma-hadron discrimination, grouped by primary energy (left) and zenith angle bin (right). AUC values are shown in the legend for each group.

4.1.2 Zenith Angle Reconstruction

The model achieves a mean absolute error of $\mathrm{MAE}_\theta = 0.484°$ on the test set, with a coefficient of determination $R^2 = 0.996$. These results represent a significant improvement over both the traditional reconstruction baseline for CONDOR reported in Arratia et al. (2025) (approximately 1–2°) and the predecessor CNN-LSTM model of Navarro et al. (2025) ($\mathrm{MAE}_\theta = 1.14°$). Table 4.3 summarizes the angular resolution metrics.

Figure 4.4 shows the absolute and relative error distributions by particle type, together with the linearity profile $\mathbb{E}[\hat{\theta} \,|\, \theta_{\mathrm{true}}]$. The absolute error distribution is heavily concentrated below 1° for both particle types, with a mean bias of +0.036°, confirming that the model is nearly unbiased across the range. Events with $\theta_{\mathrm{true}} = 0°$ are excluded from the relative error panel, as the fractional residual $(\hat{\theta} - \theta_{\mathrm{true}})/\theta_{\mathrm{true}}$ is undefined at the origin. The linearity profile follows the ideal diagonal closely, with residuals that are symmetric and stable from 0° to 40°.

Table 4.3: Zenith angle reconstruction performance on the test set.

Metric	Value
MAE	0.484°
RMSE	0.736°
R^2	0.996

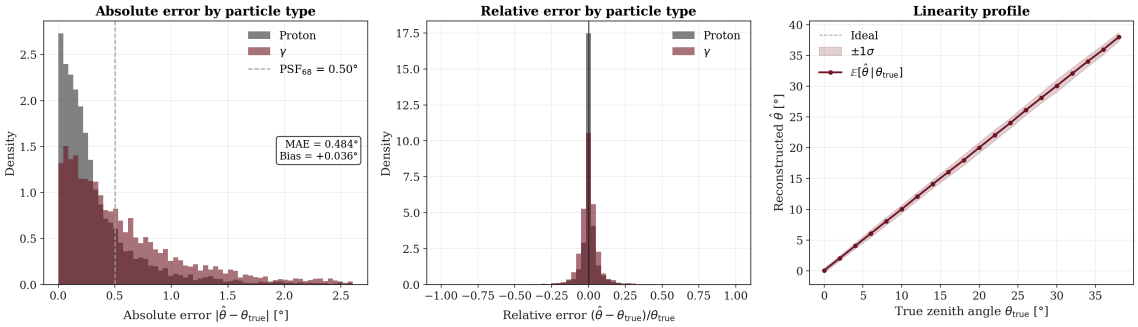


Figure 4.4: Zenith angle reconstruction on the test set. **Left:** distribution of the absolute error $|\hat{\theta} - \theta_{\text{true}}|$ for proton and gamma events, with key metrics inset. **Center:** relative error distribution $(\hat{\theta} - \theta_{\text{true}})/\theta_{\text{true}}$ by particle type. **Right:** mean reconstructed angle $\mathbb{E}[\hat{\theta} \,|\, \theta_{\text{true}}]$ with $\pm 1\sigma$ bands, confirming linearity across the full 0°–40° range.

The MAE of 0.484° corresponds to sub-degree pointing accuracy. The 68% containment radius of the angular residual distribution, $\text{PSF}_{68} = 0.50°$, is the standard figure of merit for source identification in gamma-ray astronomy. The RMSE-to-MAE ratio of 1.52 is moderately above the value expected for a pure Gaussian residual distribution ($\sqrt{\pi/2} \approx 1.25$), indicating mildly heavier-than-Gaussian tails; large outlier reconstructions are present but infrequent. Performance is stable across the full 0°–40° zenith range covered by the training data.

4.1.3 Energy Estimation

Energy estimation is the most challenging of the three tasks due to the discrete and narrow primary energy range (300–800 GeV). The model achieves $\text{MAE}_E = 103.72$ GeV and $\text{RMSE}_E = 151.80$ GeV on the test set, with an overall fractional energy resolution $\sigma_E/E = 33.1\%$ and $R^2 = 0.453$. The stratified analysis in Table 4.5 shows that the resolution varies substantially by particle type and energy, ranging from 16–20% for protons across all energy bins to 45% for gamma events at 300 GeV. The low R^2 reflects the discrete nature of the three training energies rather than a modeling

failure: the coefficient of determination penalizes the large inter-class variance that dominates when only three energy values are present. Table 4.4 summarizes the scalar metrics.

Table 4.4: Energy estimation performance on the test set.

Metric	**Value**
MAE	103.72 GeV
RMSE	151.80 GeV
Pull σ	0.331

The bias analysis reveals a systematic energy-dependent trend, summarized in Table 4.5: the model overestimates at 300 GeV and underestimates at 800 GeV, with the effect more pronounced for gamma events than for protons. This is a consequence of the discrete training energies combined with shower fluctuations. At 300 GeV, electromagnetic showers can occasionally develop late, producing a footprint more typical of a higher-energy event; at 800 GeV, shower fluctuations in the opposite direction cause the model to pull predictions toward the center of the energy range. This regression-toward-the-mean bias is expected for discrete-target regression and should be accounted for when applying the model to analysis pipelines.

Table 4.5: Energy bias and resolution per true energy bin, split by particle type. Bias = mean fractional deviation $(\hat{E} - E)/E$; Resolution = standard deviation of the fractional residual. Both expressed as percentages.

	Proton		**Gamma**	
Energy	Bias (%)	Res. (%)	Bias (%)	Res. (%)
300 GeV	+6.0	20.2	+43.8	45.3
500 GeV	+6.0	19.9	+12.0	29.2
800 GeV	−11.2	15.5	−17.7	17.5

Figure 4.5 shows the overall pull distribution and its decomposition by particle type. The overall pull is approximately Gaussian with $\mu = 0.065$ and $\sigma = 0.331$; the small positive mean indicates a systematic overestimation of approximately 6.5%, consistent with the regression-toward-the-mean effect identified in the stratified analysis: the model pulls extreme predictions toward the center of the training energy range. The per-class pull distributions confirm the same class asymmetry visible

in Table 4.5: proton events have a substantially narrower pull ($\sigma_p \approx 0.20$) than gamma events ($\sigma_\gamma \approx 0.41$). Proton-induced showers spread over many detectors with a stable lateral profile at fixed energy, whereas gamma-induced showers depend more sensitively on the electromagnetic-core position within the array, which is a larger source of variance for energy inference. Proton events are essentially unbiased ($\mu_p \approx 0.004$), while the overall positive mean of the combined distribution is driven primarily by a $\sim 13\%$ overestimate in the gamma class ($\mu_\gamma \approx 0.13$).

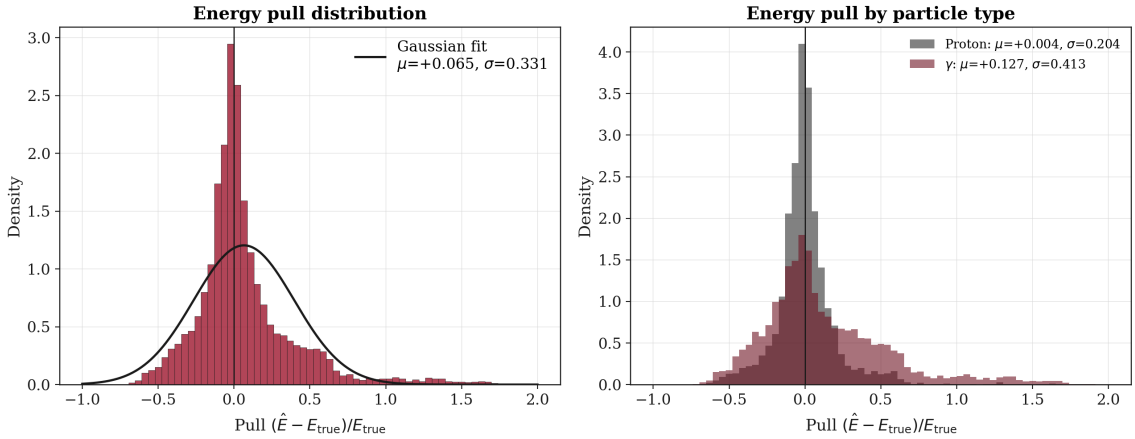


Figure 4.5: Energy estimation analysis. **Left:** overall pull distribution $(\hat{E} - E_{\text{true}})/E_{\text{true}}$ with Gaussian reference curve ($\mu = 0.065$, $\sigma = 0.331$). **Right:** pull distribution split by particle type; both classes are essentially unbiased but gamma events show roughly twice the width of proton events, consistent with the per-class resolutions in Table 4.5.

The energy head produces a continuous estimate, so a rule is required to place each prediction into one of the three energy bins of the dataset. Each event is assigned to the bin whose nominal energy is closest to its prediction, that is, to the value in $\{300, 500, 800\}$ GeV minimizing $|\hat{E} - E_{\text{bin}}|$. This is equivalent to cutting the continuous axis at the midpoints between consecutive levels, at 400 and 650 GeV. The assignment is used only to build the matrix and plays no part in training or in any of the regression metrics reported above, all of which are computed on the continuous predictions.

Figure 4.6 shows the resulting energy migration matrix, which reports for each true energy bin the fraction of events reconstructed in each of the three energy bins. The diagonal elements are 75.0%, 68.2%, and 67.5% for 300, 500, and 800 GeV respectively. The largest off-diagonal entry is the 27.8% migration from 800 GeV to the 500 GeV bin: more than a quarter of high-energy events are systematically reconstructed at a lower energy. This is consistent with the negative bias at 800 GeV identified above and reflects the inherent difficulty of separating 500 and 800

GeV showers in the sub-TeV regime, where the difference in shower size and footprint between adjacent energy bands can be within the natural fluctuation range of individual events.

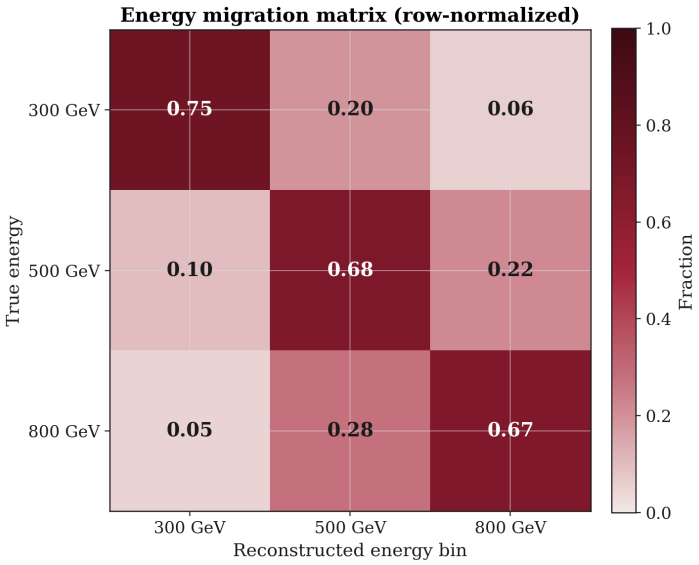


Figure 4.6: Energy migration matrix. Each row gives the fraction of events at a given true energy that are reconstructed in each of the three energy bins (row-normalized). A dominant diagonal indicates low migration; off-diagonal entries indicate systematic misreconstruction between adjacent energy bins.

4.1.4 Comparison with the State of the Art

Table 4.6 places the results in the context of prior work on deep learning for EAS reconstruction. Direct comparisons across experiments are complicated by differences in detector technology, altitude, energy range, and the number of simultaneously reconstructed quantities; the table therefore includes the relevant context for each comparison. The model presented in this thesis is the only published work to address three reconstruction tasks simultaneously, at this altitude and energy range, with an integrated explainability framework.

The methods and published results against which this work is compared are described in Section 2.10. The angular resolution of 0.484° represents a factor of 2.4 improvement over the predecessor model Navarro et al. (2025) and places the reconstruction well below the 1° threshold typically required for point-source identification in gamma-ray astronomy. The gamma-hadron AUC of 0.991 is achieved at a substantially lower energy range and with a more complex multi-task constraint than the comparable scintillator-array results of Okukawa et al. (2024), who report

Table 4.6: Comparison with selected published results on deep learning for EAS reconstruction. *Tasks*: number of simultaneously reconstructed quantities (g/h = gamma-hadron, θ = zenith angle, E = energy). Dashes indicate metrics not reported or not applicable.

Reference	Observatory	Alt.	Energy range	Tasks	g/h AUC	Ang. MAE
This work	CONDOR	5,300 m	300–800 GeV	3 (g/h, θ, E)	**0.991**	**0.484°**
Navarro et al. (2025)	CONDOR	5,300 m	150–800 GeV	2 (g/h, θ)	F1=0.89	1.14°
Arratia et al. (2025)	CONDOR	5,300 m	100 GeV–1 TeV	θ	—	1–2°
Okukawa et al. (2024)	Tibet ASγ	4,300 m	10–100 TeV	1 (g/h)	0.781–0.901	—
Erdmann et al. (2018)	Pierre Auger	1,400 m	3–100 EeV	1 (θ)	—	1.2°

AUC values of 0.781–0.901 at 10–100 TeV on a single-task network. The most directly comparable gamma-hadron study in our energy window is that of Conceição et al. (2025), who apply a pretrained vision transformer to the shower footprint of a water-Cherenkov array at 5,200 m over the 200 GeV–1 TeV range, reporting a proton rejection of approximately 97% at a gamma efficiency of 80%.

A direct AUC comparison with their work is not possible because they quote performance at fixed efficiency working points rather than a threshold-integrated metric, and their detector technology (water Cherenkov) differs from the CONDOR scintillator array; nonetheless, their results corroborate that sub-TeV footprint morphology carries strong discriminating power. Energy estimation, by contrast, is not directly comparable across experiments, as the energy ranges and figures of merit differ substantially: Meng et al. (2025), for instance, reconstruct primary energy for Tibet ASγ but over the 1 TeV–10 PeV range, where the resolution behaves intrinsically differently. The migration matrix in Figure 4.6 therefore provides the most informative characterization of energy reconstruction quality for this work.

4.2 Explainability Analysis

Two complementary feature importance methods — a feature ablation study and permutation feature importance — are applied to validate that the learned representations are physically consistent. Because both quantify the same notion (the degradation in task performance when a single input feature is disabled) and yield mutually consistent rankings, they are presented together. They are followed by an analysis of the attention patterns extracted from the RAW Transformer streams, which provide hit-level interpretability directly in the space of detector observables.

4.2.1 Feature Importance

The term *ablation* is borrowed from experimental neuroscience, where lesion studies selectively remove or disable a region of the brain and infer the function of that region from the behavioural deficit that follows; the same methodology was later adopted to probe the internal representations of artificial neural networks Meyes et al. (2019). The **feature ablation study** applies this logic at the level of the input: it measures the importance of each input feature by replacing it with its median value across the training set — computed over non-padded positions only — and evaluating the resulting degradation in test-set performance, attributing the loss to the information that the disabled feature carried.

Permutation feature importance (PFI) Altmann et al. (2010) instead randomly permutes a single feature across all test-set events, breaking its correlation with the target while preserving its marginal distribution. The two are complementary: ablation collapses a feature to a single typical value and is therefore sensitive to the loss of its typical variation, whereas PFI preserves the full distribution and is more sensitive to the information carried by a feature's extreme values. Both methods are applied independently to each of the six sequential features — `detector_id`, `particle_count`, `t_bin`, `total_energy`, `x_center`, and `y_center` — for each of the three tasks. Figures 4.7 and 4.8 report the relative degradation, expressed as a percentage of the baseline metric, for ablation and PFI respectively.

Zenith angle. Both methods agree that `t_bin` is by far the most important feature for angular reconstruction. Ablation increases the MAE from its baseline of 0.484° to 12.9° (a relative degradation of approximately 2,568%), and PFI yields a comparable ΔMAE = 12.76°. This is physically expected: the arrival time of each detector hit encodes the position of the shower wavefront, and the gradient of arrival times across the array is the primary observable from which the zenith angle is geometrically determined. The next most important features are the spatial coordinates `x_center` (ablation +9.1°, 1,889%; PFI +5.1°) and `detector_id` (ablation +7.0°, 1,456%; PFI +3.3°), consistent with the geometric nature of the task.

Gamma-hadron discrimination. For classification both methods place energy deposition and spatial position at the top, with a minor reordering between them. Ablation ranks `total_energy` first (accuracy drop of 0.224, a relative degradation of 23.1%), followed by `x_center` (0.136, 14.0%) and `t_bin` (0.118, 12.2%); PFI ranks `x_center` first (0.338, 34.8%), then `total_energy` (0.257, 26.5%) and `t_bin` (0.220, 22.7%). The prominence of `total_energy` reflects the role of the per-hit deposited energy as a tracer of local particle density: electromagnetic showers concentrate this density in a compact core while hadronic showers spread it more widely and irregularly, so the spatial pattern of per-hit energy deposits carries the morphological

signature that distinguishes the two classes. The reordering between the two methods reflects the redundancy among the positional features (`detector_id`, `x_center`, and `y_center` all encode hit location): permuting one while leaving the others intact allows partial compensation, distributing the measured importance across the correlated features. Both methods point to a discrimination strategy based on the energy-weighted lateral distribution rather than on single-detector particle identification.

Energy estimation. Spatial position dominates energy reconstruction. Ablation ranks `x_center` first (+106.4 GeV, 102.6%), followed by `detector_id` (+83.4 GeV) and `t_bin` (+72.7 GeV); PFI ranks `t_bin` first (ΔMAE = 173.3 GeV), followed by `x_center` (143.9 GeV) and `detector_id` (103.6 GeV). The spatial position of active hits encodes the shower's lateral extent, which is the strongest predictor of primary energy in ground-based EAS experiments; the prominence of `t_bin` under PFI reflects the additional information carried by the full temporal spread of the shower development. Across all three tasks and both methods, `particle_count` is the least informative feature (accuracy drop < 0.01, MAE increase < 1 GeV), indicating that the number of particles at a given detector adds little beyond what the deposited energy already captures.

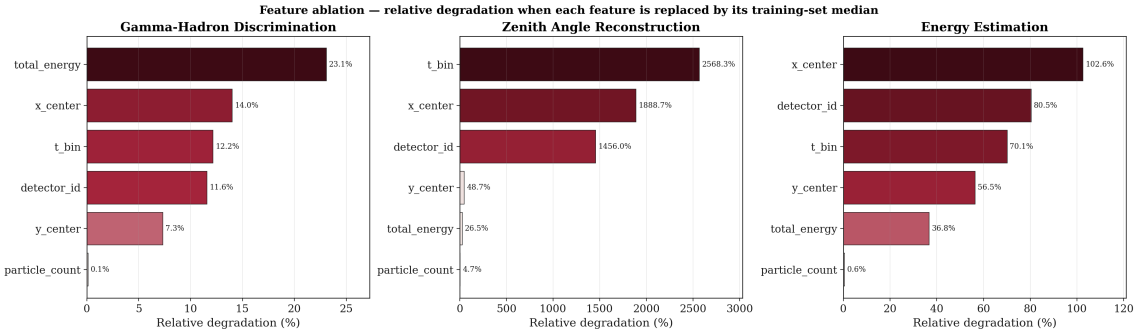


Figure 4.7: Feature ablation study: relative degradation, expressed as a percentage of the baseline metric, when each feature is replaced by its training-set median. Each panel corresponds to one reconstruction task; bars are ordered by magnitude and shaded by it.

4.2.2 Attention Analysis

The RAW Transformer streams operate directly on the masked hit sequence, so their attention weights can be examined at the level of individual detector hits. For an event with L valid hits, α_{ij} denotes the weight that query position i assigns to key position j. The weights are produced by a softmax over the valid keys, so that

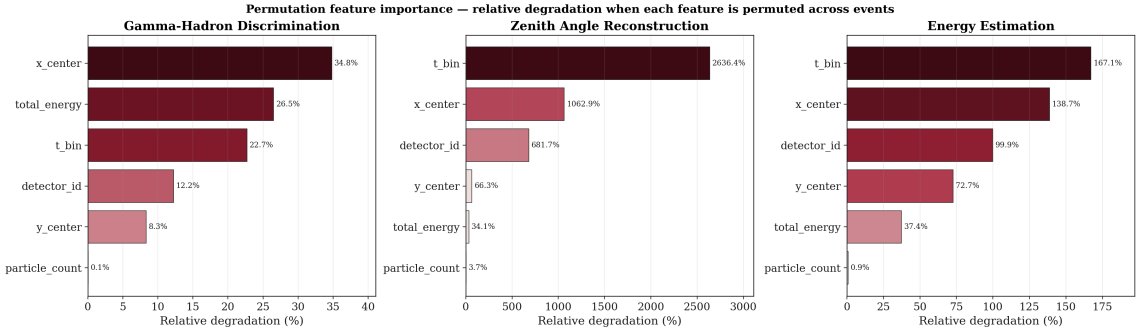


Figure 4.8: Permutation feature importance (PFI): relative degradation, expressed as a percentage of the baseline metric, when each feature is permuted across all test-set events. Each panel corresponds to one reconstruction task.

$\sum_{j=1}^{L} \alpha_{ij} = 1$. Two views are reported. The first is the mean attention matrix, averaged over events. The second is the mean attention received by each position,

$$\bar{\alpha}_{\cdot j} = \frac{1}{L} \sum_{i=1}^{L} \alpha_{ij}, \tag{4.1}$$

which identifies the positions that act as information sources for a stream. Because each row sums to unity, a position of average importance receives a weight of $1/L$; for the mean sequence length of 76.9 valid hits, this reference level is 1.3×10^{-2}. Hits are ordered chronologically, so low position indices correspond to early-arriving hits, near the leading edge of the shower front. The analysis uses a stratified subsample of 200 test events and the first 120 sequence positions.

Figure 4.9 shows both views for the three streams, and the distributions differ markedly between tasks. The classification stream concentrates its weight on the leading edge of the sequence: the first position receives a mean weight of 0.34, twenty-six times the uniform level, only the first thirteen positions exceed that level, and 82% of the profile lies within the first ten positions. Equivalently, 93% of the matrix weight is assigned to keys that precede the query.

The angle stream shows the opposite ordering. Its profile grows with position, exceeds the uniform level only beyond position 44, and peaks at 2.5×10^{-2} near position 98, with 0.7% of the profile in the first ten positions. Correspondingly, 80% of the matrix weight is assigned to keys that follow the query, producing the triangular structure visible in the mean matrix. The energy stream is intermediate: 75% of its weight also falls on later keys, but instead of the smooth ramp of the angle stream the matrix shows isolated bands at specific positions, and a weak diagonal component in which positions attend to their own neighbourhood.

These distributions are consistent with the feature importance results of Section 4.2.1. Zenith angle reconstruction is dominated by `t_bin` under both ablation

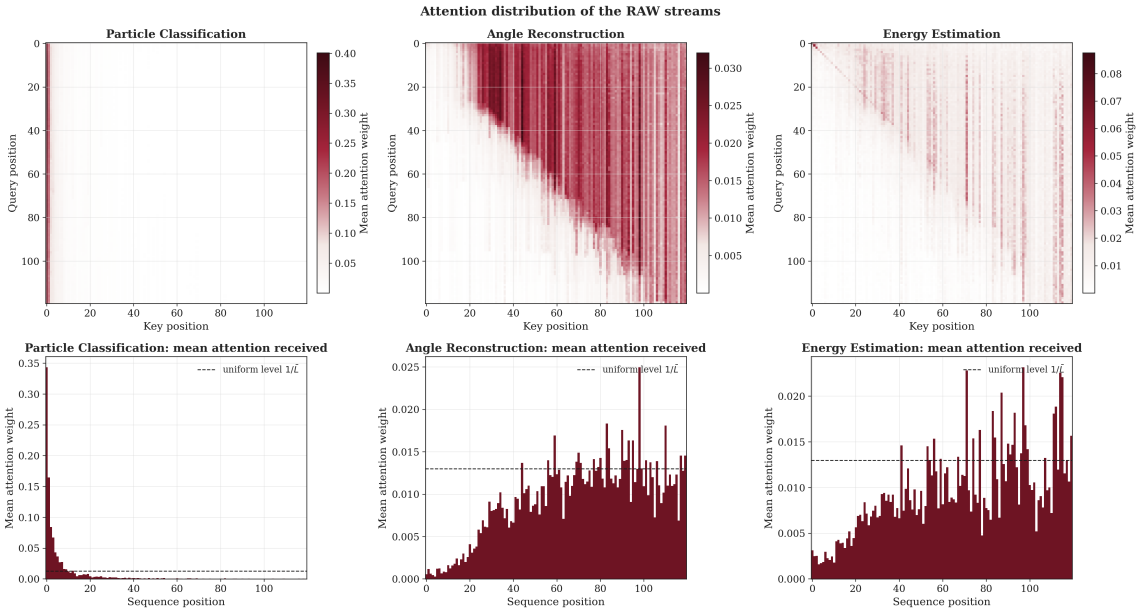


Figure 4.9: Attention distribution of the three RAW streams. **Top:** mean attention matrix, averaged over 200 test events; rows are query positions and columns are key positions. **Bottom:** mean attention received per position, Equation (4.1); the dashed line marks the uniform reference level $1/\bar{L} = 1.3 \times 10^{-2}$. Cells averaged over fewer than ten events are left blank.

and permutation, and the angle stream accordingly draws its weight from the late part of the sequence, where the arrival-time gradient across the array is defined. Gamma-hadron discrimination depends most strongly on the deposited energy and on the spatial coordinates, quantities that the leading hits already characterize, and the classification stream concentrates on precisely that region. Energy estimation shows the least specialized pattern of the three, which matches its position as the task with the lowest reconstruction quality.

Interpretive scope of the attention maps. Attention weights describe how a stream distributes its reads over the input sequence. They are not an attribution of the prediction to those inputs, and a large weight is not by itself evidence that the corresponding hit was used causally (Jain & Wallace, 2019). The distributions above are therefore reported as a description of the model's internal weighting, and the feature-level conclusions of this chapter rest on the ablation and permutation analyses of Section 4.2.1, which intervene on the input, measure the resulting change in reconstruction performance, and agree with each other. The limitations of attention-based explainability are discussed further in Section 5.2.

4.3 Detector Robustness and Array Optimization

Because CONDOR is still a proposed observatory, the trained model can be used to answer two practically important questions: how much reconstruction performance degrades as detectors fail, and whether the current array layout uses its detectors efficiently. This section addresses these questions through a *detector switch-off* procedure applied at inference time, without any retraining.

Concretely, for a chosen set of detectors to be switched off, every hit recorded by those detectors is deleted from each test event, the event's sequential representation is rebuilt from the surviving hits, and the five event-level global features (total particle count, total and central deposited energy, active-detector count, and duration) are recomputed from that reduced hit set. The already-trained model is then evaluated on these modified events, and the resulting change in each task metric relative to the intact-array baseline quantifies how much the model relied on the removed detectors. Events that lose all of their hits are excluded from the corresponding evaluation.

This procedure measures the *dependence* of the trained model on specific detectors, which serves as a proxy for their information content. It is important to note that a model retrained from scratch on a reduced detector configuration could in principle learn compensatory representations that partially recover the lost information; the degradation reported here therefore represents an upper bound on the actual performance loss under detector failure. All experiments in this section are performed on the same stratified subsample of 3,000 test events, on which the intact-array baseline is a classification accuracy of 96.7% with an F1-score of 0.967 and an AUC of 0.990, an angular MAE of 0.49°, and an energy MAE of 107 GeV.

4.3.1 Single-Detector Importance

Figure 4.10 shows the degradation in each task metric when each of the 120 detectors is individually switched off. The spatial maps reveal a clear core-periphery gradient: detectors near the geometric center of the array are individually more critical than peripheral detectors for all three tasks. This is physically expected, since the shower core passes through or near the center of the array for most events in the dataset, and central detectors sample the highest-density region of the particle distribution, which is the most discriminative for both particle type and energy.

4.3.2 Regional Dropout

Two families of geometrically coherent detector groups are switched off simultaneously, defined in Figure 4.11. The first follows how the array is built: the 100 modules of the central grid are split into the 36 closest to the array centre (*central core*) and

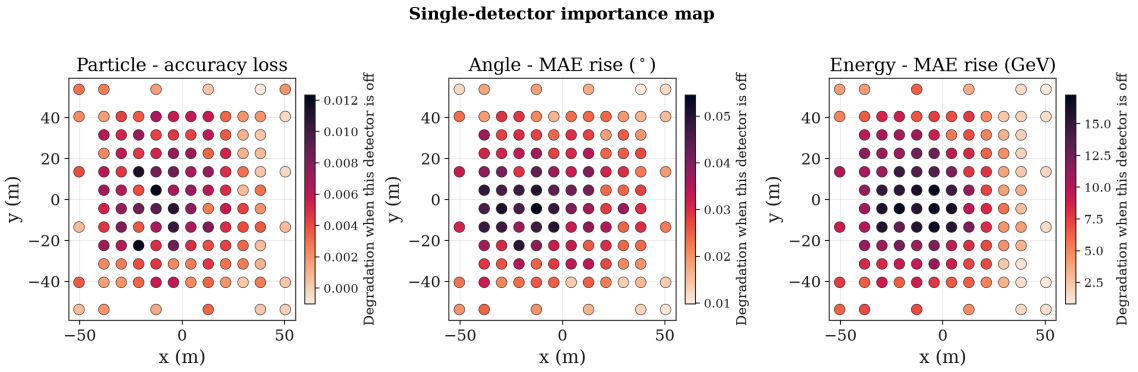


Figure 4.10: Degradation in each task metric when each detector is individually switched off. Brighter colors indicate detectors whose loss causes the largest performance drop. The core-periphery gradient is consistent across all three tasks.

the 64 that surround them (*outer grid*), and the 20 peripheral modules, which form a structurally separate sub-array, are kept as their own group. The second family consists of the four halves of the array, $x < 0$, $x > 0$, $y < 0$ and $y > 0$, each containing exactly 60 modules and together covering it completely.

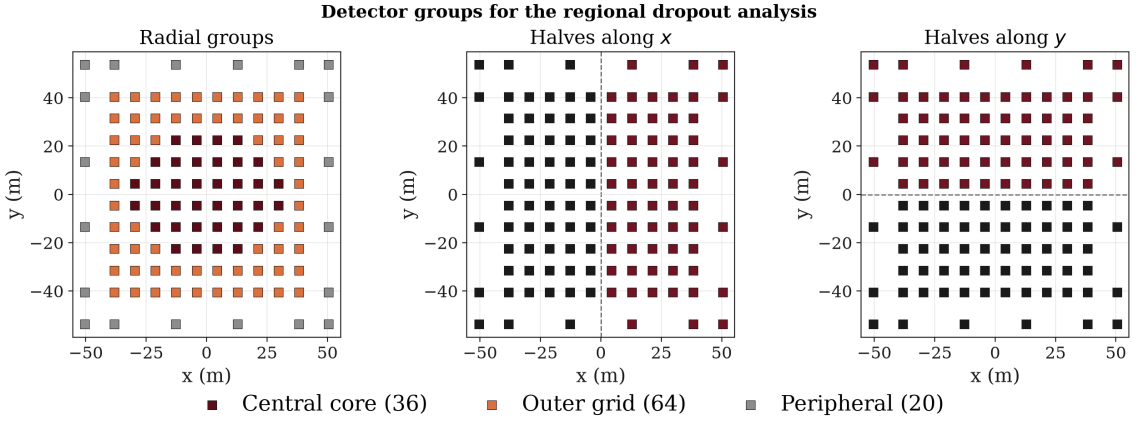


Figure 4.11: Spatial definition of the detector groups switched off in the regional dropout analysis. **Left:** the three radial groups — the 36 modules closest to the array centre, the remaining 64 modules of the central grid, and the 20 peripheral modules. **Centre and right:** the four halves, split along x and along y, of 60 modules each. The y pair acts as a control for the x pair: the simulated showers all arrive at azimuth $\phi = 0°$, which singles out the x direction and should leave the y halves equivalent. Marker positions correspond to the physical module locations on the array.

Figure 4.12 quantifies the impact of removing each of these groups. The groups

differ in size, and a larger group degrades performance more simply because it removes more detectors. Every quantity is therefore reported both as the degradation caused by removing the whole block, which corresponds to an operational failure of that block, and normalized by the number of detectors removed, which measures the contribution of an individual module. The two rankings do not coincide.

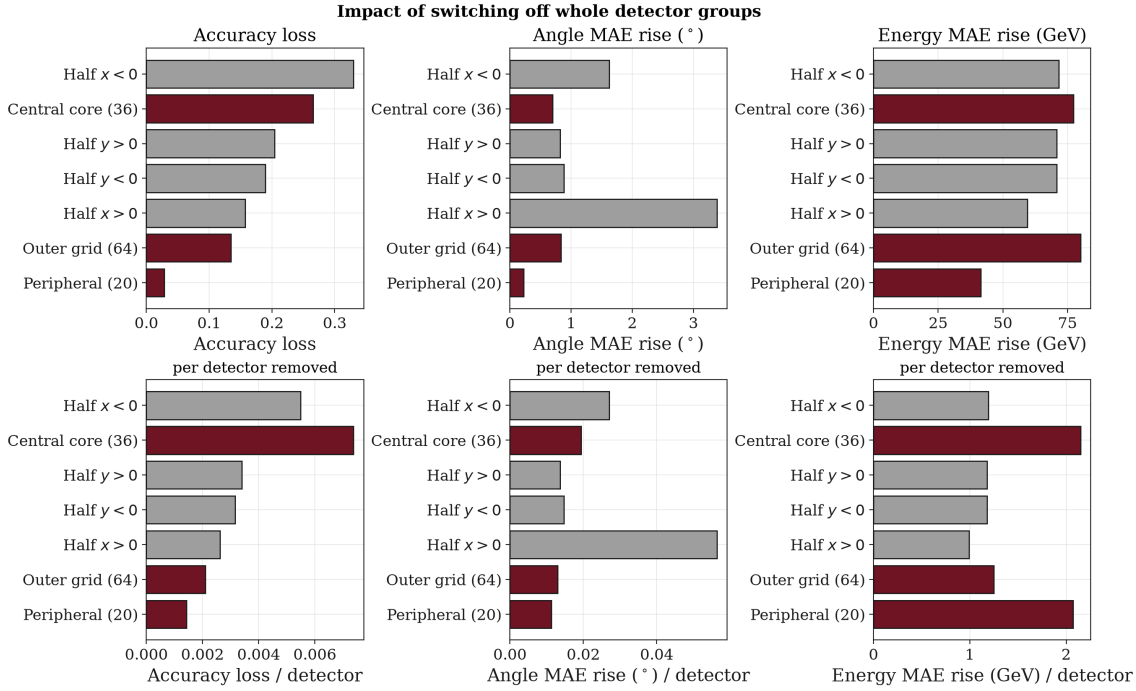


Figure 4.12: Performance degradation when entire detector groups are removed, sorted by increasing impact on particle accuracy. **Top:** degradation caused by removing the whole block. **Bottom:** the same degradation per detector removed, which controls for the different group sizes. Dark red marks the three radial groups, grey the four halves.

The three tasks do not depend on the same region of the array. For gamma-hadron classification the ordering follows the core-to-periphery one: removing the central core costs 0.265 in accuracy, a relative degradation of 27.4%, against 13.9% for the outer grid and 3.0% for the 20 peripheral modules. The core achieves this while removing half as many detectors as the outer grid, so per detector it is three and a half times more damaging (7.4×10^{-3} against 2.1×10^{-3} accuracy per module). Classification depends on the high particle density near the shower core, which only the central modules sample.

For angular reconstruction the ordering differs. Removing the outer grid raises the angular MAE by 0.83° (171%), more than the 0.70° (144%) caused by removing the central core; this is the only task for which the core is not the most damaging

block. Per detector the core remains the more critical of the two (1.9×10^{-2} against 1.3×10^{-2} degrees per module), so the outer grid leads as a block because it holds 64 modules against 36, but the per-detector gap is smaller than for classification. Zenith-angle recovery depends on the arrival-time gradient sampled across the lateral extent of the footprint rather than on the dense core alone, so the spatial spread of the modules contributes almost as much as their density.

For energy reconstruction the two normalizations disagree. In absolute terms the peripheral modules are the least important group (+41 GeV, 38.9%, against +77 and +80 GeV for the core and the outer grid). Per detector removed they are almost as costly as the central core (2.07 against 2.15 GeV per module) and above the outer grid (1.25 GeV). The 20 outermost modules are therefore not redundant for energy estimation: each contributes approximately as much as a core module, and their low block-level ranking reflects their small number. This is consistent with energy estimation requiring the lateral extent of the shower to be sampled, and with the sparsification results of Section 4.3.3.

Azimuthal dependence. The y pair acts as a control and shows no asymmetry: removing $y < 0$ or $y > 0$ gives essentially the same degradation in all three tasks (accuracy 19.6% vs 21.1%, angular MAE +0.88° vs +0.82°, energy +70.8 GeV in both cases). The x pair, by contrast, is strongly asymmetric, and in opposite directions depending on the task: for classification the $x < 0$ half is more than twice as damaging as $x > 0$ (34.1% vs 16.3%), whereas for angular reconstruction the ordering reverses and $x > 0$ is by far the more critical, raising the MAE by 3.38° against 1.63°. The two directions are geometrically equivalent in the array, so the asymmetry originates in the dataset: all showers arrive at a fixed azimuth $\phi = 0°$, an inclined shower is therefore always displaced along the same direction in x, and the two x halves sample systematically different parts of the footprint, with the trailing side carrying most of the arrival-time information on which the zenith angle depends. The agreement of the y pair indicates that the procedure introduces no directional bias of its own, so the x asymmetry is a property of the simulation configuration rather than of the array; a dataset with uniformly sampled azimuth would be expected to make both pairs symmetric.

4.3.3 Detector Count and Layout

The preceding subsections quantified which detectors the model depends on. This subsection addresses the number of detectors required and the way they are distributed over the ground, which together define an array configuration. Both are evaluated with the same switch-off procedure: the number by removing detectors at random and increasing the removed fraction, the distribution by comparing different

ways of removing the same number of detectors.

Progressive random failure. Figure 4.13 shows the degradation of each task metric as an increasing fraction of randomly chosen detectors is switched off, averaged over four independent random subsets per fraction level. This measures the failure rate the array can absorb before reconstruction degrades appreciably.

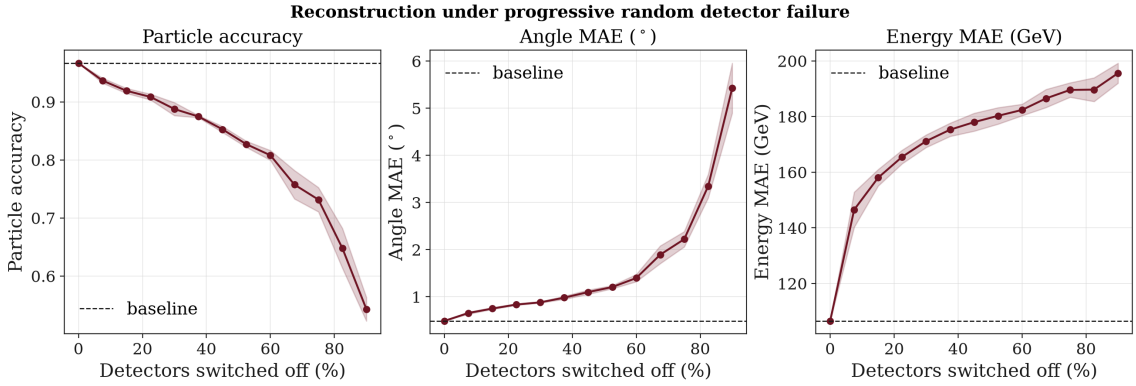


Figure 4.13: Task metric as a function of the fraction of detectors switched off at random. Shaded bands indicate one standard deviation over four random subsets per fraction. Dashed horizontal lines indicate baseline performance with all detectors active.

Gamma-hadron classification is the most fault-tolerant task: accuracy stays above 90% with up to about 22% of the detectors offline and above 85% up to roughly 45%, collapsing more steeply only beyond ∼60% loss.

The angular and energy reconstruction tasks begin to degrade more immediately. Angular MAE grows steadily, roughly doubling from its 0.49° baseline to ∼1.0° at about 45% detector loss and rising sharply thereafter. Energy MAE increases quickly at first — from 107 GeV to ∼171 GeV by 30% loss — and then saturates near 185–195 GeV, a plateau that reflects the model falling back toward predicting the mean energy of the range once too few detectors survive to constrain the shower size.

This difference in fault tolerance reflects the nature of each task: binary classification can exploit redundant evidence across many detectors and remains viable as long as some discriminative detectors survive, while metric regression requires information from a spatially distributed set of detectors to triangulate the shower geometry and measure its total energy. The knee in the particle-accuracy curve near 60–70% corresponds to the point at which too few spatially distinct measurements survive to resolve the shower footprint.

Layout at matched detector count. Random thinning reduces the detector count while leaving the covered area unchanged, so it does not separate the effect of the number of detectors from that of their arrangement. Figure 4.14 compares three

ways of reaching the same count: random thinning (detectors removed uniformly at random, sparser over the same area), central crop (only detectors within a shrinking radius are retained, denser over a smaller area), and a checkerboard pattern (every other detector removed, approximately halving the count while maintaining full coverage). At equal detector count, the differences between the three curves are attributable to the arrangement alone.

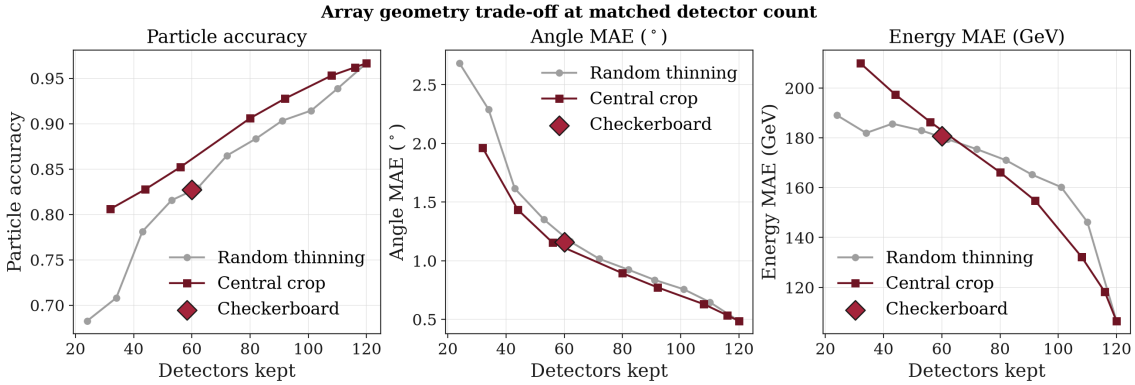


Figure 4.14: Reconstruction performance as a function of the number of active detectors, for three sparsification strategies. Central crop (denser, smaller array) outperforms random thinning (sparser, same area) at equal detector count for classification and angular reconstruction; for energy the central crop wins only above ∼65 detectors and random thinning is better below that. The checkerboard point (diamond) corresponds to approximately 60 active detectors.

The central crop strategy outperforms random thinning at matched detector counts for gamma-hadron classification and angular reconstruction at every count tested: for these two tasks, retaining the detectors closest to the array centre yields better reconstruction than retaining a random subset distributed over the full array area. Energy reconstruction, however, behaves differently. At high detector counts the central crop still performs better, but below roughly 65 active detectors the ordering reverses and random thinning gives the lower energy MAE. This crossover is consistent with the per-detector result of Section 4.3.2, where the peripheral modules were found to be nearly as valuable per unit as the core modules for energy: energy estimation relies on sampling the lateral extent of the shower, so under aggressive thinning a wide but sparse layout preserves outer footprint information that a dense central cluster discards. The checkerboard pattern (about 60 active detectors) falls close to the random-thinning curve for all three tasks, and in particular sits near this energy crossover, performing comparably to random thinning while remaining inferior to the central crop for classification and angular reconstruction.

Taken across the three studies, the results point in a consistent direction. Classi-

fication and angular reconstruction degrade least under the central crop, favouring a denser core; energy reconstruction follows the lateral spread of the modules instead and, below roughly 65 active detectors, favours a wide sparse layout over a dense central cluster. A dedicated study of the array layout, including geometries outside the present family such as hexagonal grids, variable pitch, or elliptical footprints, would re-run the CORSIKA pipeline and retrain the model for each configuration, and is left for future work.

Chapter 5

Conclusions

This thesis developed and evaluated a multi-task CNN-Transformer model for the simultaneous reconstruction of three EAS properties at CONDOR — gamma-hadron classification, zenith angle, and primary energy — on CORSIKA-simulated events in the 300–800 GeV range. The architecture extends the three-stream design of Navarro et al. (2026a) to six independent Transformer streams, two per task, one over the CNN-processed sequence and one over the raw hit sequence, which makes attention available at the level of individual detector hits.

5.1 Summary of Contributions

A multi-task CNN-Transformer model for EAS reconstruction. The central contribution is the reconstruction model itself: a single network that recovers particle type, zenith angle, and primary energy in one pass over the detector hits, with a shared convolutional backbone feeding task-specific Transformer streams. It extends the published three-stream design of Navarro et al. (2026a) to six streams, adding one attention stream per task that operates directly on the raw hit sequence, so that the weighting each task learns can be examined at the level of individual hits.

Multi-task reconstruction in the sub-TeV regime. The model reaches AUC = 0.991 (Q-factor 5.48) for gamma-hadron discrimination, angular MAE = 0.484° (PSF_{68} = 0.50°), and energy MAE = 103.72 GeV. The angular resolution improves upon Navarro et al. (2025) by a factor of 2.4 and lies well below the 1° threshold relevant for point-source identification. To the best of our knowledge, this is the only published work addressing the three EAS reconstruction tasks simultaneously at this altitude and energy range with an integrated explainability framework.

Performance by energy and zenith angle. Gamma-hadron discrimination degrades with energy in a physically consistent way (AUC from 0.998 at 300 GeV to 0.981 at 800 GeV) and is stable across the full zenith range (AUC 0.984–0.994).

Physically consistent explainability. Feature ablation and PFI agree on what each task uses: zenith reconstruction is dominated by temporal information by a wide margin (ablation: $+12.4^\circ$ of added MAE), consistent with the arrival-time gradient as the primary geometric handle for direction; gamma-hadron discrimination relies on energy deposition and spatial features; and energy estimation is governed by spatial position, consistent with the correlation between lateral extent and primary energy.

RAW attention streams and hit-level interpretability. The RAW Transformer streams produce attention matrices directly over the individual hit sequence, so the weighting learned by each task can be examined at the level of detector hits rather than of CNN features. The three streams learn distinct distributions: classification concentrates 82% of its weight on the first ten hits, angle reconstruction assigns 80% to hits later than the query and peaks near the end of the sequence, and energy estimation lies between the two, in agreement with the feature importance results.

Robustness to missing detectors. The detector dropout analysis characterises how the trained reconstruction model responds to missing detector information, measured by switching detectors off at inference without retraining. Gamma-hadron classification is the most fault-tolerant task, holding accuracy above 90% with up to approximately 22% of modules offline, while angular and energy reconstruction degrade sooner. Removing the central core is the most damaging block-level failure for classification, whereas angular reconstruction depends at least as much on the surrounding modules, which sample the lateral extent over which the arrival-time gradient is measured; and per module removed, the 20 peripheral units cost almost as much for energy as the central ones, so they are not redundant despite ranking last as a block. Because the model was trained only on the full array, these figures bound the degradation from above and indicate where the reconstruction relies on the array, rather than prescribing how many detectors to build or how to place them.

5.2 Limitations and Future Work

The limitations of this work and the lines of research that would address them are presented together, since most of the former define the latter.

Simulation-only evaluation. All results are obtained on CORSIKA-simulated events; CONDOR is a proposed observatory with no measured data yet, so generalization to real detector behaviour, with hardware noise, gain variations, detector aging, and atmospheric conditions not captured by CORSIKA, cannot be assessed here. Once the observatory begins scientific operations the model can be applied to measured events, a transition that will likely require domain adaptation, data-driven

calibration of the normalization statistics, and fine-tuning on hybrid simulation-data samples.

Energy coverage. The dataset comprises three discrete energy points (300, 500, and 800 GeV), sufficient to demonstrate multi-task learning in a controlled setting but not representative of the continuous cosmic ray spectrum; the regression-toward-the-mean bias is a direct consequence of that discretization. Performance below 300 GeV, approaching the CONDOR threshold, and above 1 TeV is likewise not characterized. Sampling the energy from a power law would remove the bias and model the flux realistically, at the cost of new CORSIKA simulations and a re-run of the preprocessing pipeline.

Fixed azimuthal angle. All simulated showers use a fixed azimuthal angle of $\phi = 0°$. The array geometry gives this simplification little support: in the proposed layout, both the modules and the central grid they form are rectangular rather than square in their two horizontal dimensions, so the array is invariant only under a 180° rotation, and the choice introduces directional biases in the learned attention patterns and in the regional dropout results (Sections 4.2.2 and 4.3.2). Because the core also lands at a different position in each event, ϕ cannot be obtained by a geometric transformation of the $\phi = 0°$ model and must be learned from a dataset with uniformly sampled azimuth, possibly with positional embeddings that encode detector coordinates explicitly.

Attention as an explainability method. The attention maps describe how each stream weights its input sequence, but they do not attribute the prediction to those inputs, and two properties of the mechanism limit what can be read from them. The weights of each row are normalized over the valid hits of the event, so their scale depends on the number of active modules, which itself differs systematically between gamma and proton showers; comparisons between classes are therefore not directly interpretable as differences in model behaviour. And a map estimated from a finite sample carries structure of its own, which a map shown without an associated uncertainty does not reveal. The analysis is accordingly restricted to the aggregate distribution of the weights per task, and the feature-level conclusions rest on the intervention-based methods of Section 4.2.

Single training run. All results come from a single training run with a fixed random seed, so the variability of the metrics across initializations is not characterized. The narrow train-validation gap and the consistency across groups suggest the results are not an artifact of a particular seed, but confidence intervals from independent runs would be stronger evidence.

Multi-task advantage not quantified. The three tasks are optimized jointly under the premise that shared representations benefit physically correlated objectives, but the model was not benchmarked against three independently trained

single-task networks, so that advantage is not established here. Training single-task variants, removing the RAW or CNN streams, and varying the number of attention heads would isolate the contribution of each component and settle the question.

Single observatory. The model is tailored to the CONDOR array geometry; the input representation, the normalization statistics, and the learned representations are specific to it, and transfer elsewhere would require retraining. Adapting the framework to instruments such as HAWC, LHAASO, or the future SWGO, which cover complementary energy ranges at different altitudes and geometries, is a natural path toward cross-experiment consistency.

Uncertainty quantification. The model produces point estimates. Bayesian layers, Monte Carlo dropout at inference, or normalizing flows on the output distributions would provide event-by-event confidence intervals, useful for downstream analyses such as source detection and spectral unfolding; this direction is being explored in a related effort applying Bayesian neural networks to the CONDOR reconstruction problem.

Array geometry. Finally, the robustness results motivate evaluating alternative layouts: hexagonal grids, elliptical footprints, and variable-pitch designs could improve reconstruction for a fixed hardware budget. Any such study requires rerunning the CORSIKA pipeline on the new detector positions, since the processed dataset encodes the current geometry at the level of individual hit aggregations.

References

Abeysekara, A. U., et al. (2023). The High-Altitude Water Cherenkov (HAWC) observatory in México: The primary detector. *Nucl. Instrum. Meth. A*, *1052*, 168253. doi: 10.1016/j.nima.2023.168253

Abreu, P., et al. (2019, 7). The Southern Wide-Field Gamma-Ray Observatory (SWGO): A Next-Generation Ground-Based Survey Instrument for VHE Gamma-Ray Astronomy.

Ackermann, M., Ajello, M., Atwood, W., Baldini, L., Ballet, J., Barbiellini, G., ... others (2012). Fermi-lat observations of the diffuse γ-ray emission: implications for cosmic rays and the interstellar medium. *The Astrophysical Journal*, *750*(1), 3.

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (xai). *IEEE access*, *6*, 52138–52160.

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th acm sigkdd international conference on knowledge discovery & data mining* (pp. 2623–2631).

Alfaro, R., Alvarez, C., Álvarez, J., Camacho, J. A., Arteaga-Velázquez, J., Rojas, D. A., ... others (2022). Gamma/hadron separation with the hawc observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, *1039*, 166984.

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: a corrected feature importance measure. *Bioinformatics*, *26*(10), 1340–1347.

Amenomori, M., Bao, Y. W., Bi, X. J., Chen, D., Chen, T. L., Chen, W. Y., ... others (2019). First detection of photons with energy beyond 100 TeV from an astrophysical source. *Physical Review Letters*, *123*(5), 051101.

Arratia, M., Brooks, W., Huang, J., Muñoz J., G., Navarro F., L., Paul, S. J., ... Viaux M., N. (2025, dec). The condor observatory: a gamma-ray observatory with a 100 gev threshold at 5300 meters above sea level. *Journal of Instrumentation*, *20*(12), P12020. Retrieved from `https://doi.org/10.1088/1748-0221/20/12/P12020` doi: 10.1088/1748-0221/20/12/P12020

Ba, J. L., Kiros, J. R., & Hinton, G. E. (2016). Layer normalization. *arXiv preprint arXiv:1607.06450*.

Bell, A. (2013). Cosmic ray acceleration. *Astroparticle Physics*, *43*, 56–70.

Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyper-parameter optimization. In *Advances in neural information processing systems* (Vol. 24).

Blasi, P. (2013). The origin of galactic cosmic rays. *The Astronomy and Astrophysics Review*, *21*(1), 70.

Cao, Z., della Volpe, D., Liu, S., Editors, :, Bi, X., ... Zuo, X. (2022). *The large high altitude air shower observatory (lhaaso) science book (2021 edition).* Retrieved from `https://arxiv.org/abs/1905.02773`

Caruana, R. (1997). Multitask learning. *Machine learning*, *28*(1), 41–75.

Clevert, D.-A., Unterthiner, T., & Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*, *4*(5), 11.

Collaboration, P. A., et al. (2014). Reconstruction of inclined air showers detected with the pierre auger observatory. *Journal of Cosmology and Astroparticle Physics*, *2014*(08), 019–019.

Conceição, R., González, B., Guillén, A., Pimenta, M., & Tomé, B. (2025). Discriminating sub-tev gamma and hadron-induced showers through their footprints. *Physical Review D*, *111*(4), 043047.

Deligny, O., & Collaboration, P. A. (2025). Searches for ultra-high energy gamma-ray at the pierre auger observatory and implications on super-heavy dark matter. In *Journal of physics: Conference series* (Vol. 3053, p. 012016).

Erdmann, M., Glombitza, J., & Walz, D. (2018). A deep learning-based reconstruction of cosmic ray-induced air showers. *Astroparticle Physics*, *97*, 46–53.

Fesefeldt, H. (1985). *The simulation of hadronic showers: Physics and applications* (Tech. Rep. No. PITHA 85/02). RWTH Aachen.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 770–778).

Heck, D., Knapp, J., Capdevielle, J. N., Schatz, G., & Thouw, T. (1998). Corsika: A monte carlo code to simulate extensive air showers. *Forschungszentrum Karlsruhe Report*, *FZKA 6019*.

Heitler, W. (1984). *The quantum theory of radiation.* Courier Corporation.

Hess, V. F. (1912). *Über beobachtungen der durchdringenden strahlung bei sieben freiballonfahrten.* Kaiserlich-Königlichen Hof-und Staatsdruckerei, in Kommission bei Alfred Hölder.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Huber, P. J. (1992). Robust estimation of a location parameter. In *Breakthroughs in statistics: Methodology and distribution* (pp. 492–518). Springer.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning* (pp. 448–456).

Jain, S., & Wallace, B. C. (2019). Attention is not explanation. *arXiv preprint arXiv:1902.10186*.

Kamata, K., & Nishimura, J. (1958, 02). The lateral and the angular structure functions of electron showers. *Progress of Theoretical Physics Supplement*, *6*, 93-155. Retrieved from `https://doi.org/10.1143/PTPS.6.93` doi: 10.1143/PTPS.6.93

Kawata, K., Sako, T., Ohnishi, M., Takita, M., Nakamura, Y., & Munakata, K. (2017). Energy determination of gamma-ray induced air showers observed by an extensive air shower array. *Experimental Astronomy*, *44*(1), 1–9.

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In *Proceedings of the 3rd international conference on learning representations (iclr)*.

Langner, N., Halim, A. A., Abreu, P., Aglietta, M., Allekotte, I., Cheminant, K. A., ... others (2024). Deep-learning-based cosmic-ray mass reconstruction using the water-cherenkov and scintillation detectors of augerprime. In *38th international cosmic ray conference (icrc2023)-cosmic-ray physics (indirect, cri)*.

LeCun, Y., & Bengio, Y. (1998). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436–444.

Lin, T., Wang, Y., Liu, X., & Qiu, X. (2022). A survey of transformers. *AI open*, *3*, 111–132.

Longair, M. S. (2011). *High energy astrophysics*. Cambridge university press.

Matthews, J. (2005). A heitler model of extensive air showers. *Astroparticle Physics*, *22*(5-6), 387–397.

Meng, Y., Huang, J., Chen, D., Hu, K. Y., Zhang, Y., Zhai, L. M., ... Li, Y. Y. (2025). Machine learning applications to energy reconstruction of gamma-ray showers for the tibet as γ experiment. *Experimental Astronomy*, *59*(2), 23.

Meyes, R., Lu, M., de Puiseau, C. W., & Meisen, T. (2019). *Ablation studies in artificial neural networks.* (arXiv:1901.08644)

Moreira, P., Serrano, J., Wlostowski, T., Loschmidt, P., & Gaderer, G. (2009). White rabbit: Sub-nanosecond timing distribution over ethernet. In *2009 international symposium on precision clock synchronization for measurement, control and communication* (pp. 1–5).

Morrison, P. (1958). On gamma-ray astronomy. *Il Nuovo Cimento (1955-1965)*, *7*(6), 858–865.

Navarro, L., Pezoa, R., Viaux, N., Tapia, S., & Valdivieso, C. (2025). Multi-task deep learning models for particle classification and zenith angle regression in cosmic ray air showers. In *Jornadas chilenas de la computación (jcc)*.

Navarro, L., Pezoa, R., Viaux, N., Tapia, S., Valdivieso, C., Mendizabal, S., ... Brooks, W. (2026a). Deep learning for extensive air showers reconstruction: an interpretable multi-task CNN-Transformer model for the CONDOR observatory. *Astronomy and Computing*, 101173. doi: 10.1016/j.ascom.2026.101173

Navarro, L., Pezoa, R., Viaux, N., Tapia, S., Valdivieso, C., Mendizabal, S., ... Brooks,

W. (2026b). Reconstruction of extensive air showers in the CONDOR project with multi-task deep learning. In *Proceedings of the 43rd international conference on high energy physics (ichep 2026).*

Navarro Farías, L. F. (2025). *Dataset and trained models for CONDOR EAS reconstruction (CNN-Transformer).* Zenodo. doi: 10.5281/zenodo.17717722

Navas, S., et al. (2024). Review of Particle Physics. *Physical Review D*, *110*, 030001. doi: 10.1103/PhysRevD.110.030001

Okukawa, S., Hara, K., Hibino, K., Katayose, Y., Kawata, K., Ohnishi, M., ... others (2024). Neural networks for separation of cosmic gamma rays and hadronic cosmic rays in air shower observation with a large area surface detector array. *Machine Learning: Science and Technology*, *5*(2), 025016.

Pierog, T., & Werner, K. (2009). Epos model and ultra high energy cosmic rays. *Nuclear Physics B-Proceedings Supplements*, *196*, 102–105.

Prahl, J., Eckmann, R., Aharonian, F., Akhperjanian, A., Arqueros, F., Becker, K., ... others (1993). A procedure to improve the angular resolution for extended air showers measured with a scintillator array. In *23rd international cosmic ray conference (icrc23), volume 4* (Vol. 4, p. 734).

Rao, M. V., & Sreekantan, B. V. (1998). *Extensive air showers.* World scientific.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 1135–1144).

Rossi, B., & Greisen, K. (1941, Oct). Cosmic-ray theory. *Rev. Mod. Phys.*, *13*, 240–309. Retrieved from `https://link.aps.org/doi/10.1103/RevModPhys.13.240` doi: 10.1103/RevModPhys.13.240

Schneider, M., Pirke, M., Leitl, F., Glombitza, J., & van Eldik, C. (2026). Deep learning methods for gamma/hadron separation in swgo. In *en. in: Proceedings of 39th international cosmic ray conference—pos (icrc2025)* (Vol. 501, p. 836).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Xiong, R., Yang, Y., He, D., Zheng, K., Zheng, S., Xing, C., ... Liu, T. (2020). On layer normalization in the transformer architecture. In *International conference on machine learning* (pp. 10524–10533).

Appendix A

Supplementary Material

A.1 Atmospheric Depth

The development of an air shower is described as a function of *atmospheric depth* rather than of altitude or of distance travelled, and quantities such as $X_{\max}$ are consequently reported in g cm^{-2}. This section sets out why that variable is the natural one and what the choice implies for the two aspects of this work that depend on it: the reconstruction of the zenith angle and the altitude of the observation site.

The atmosphere is the target in which the cascade develops, and its density falls by orders of magnitude between the top of the atmosphere and the ground. The vertical atmospheric depth at an altitude h is defined as the integral of the air density above that altitude,

$$X_v(h) = \int_h^\infty \rho(h')\,\mathrm{d}h' \ , \tag{A.1}$$

that is, the mass of air per unit area contained in the column above the point in question, which is where the units come from. Because the atmosphere is in hydrostatic equilibrium, this column mass is exactly what a barometer measures: $X_v = P/g$, with P the local pressure and g the gravitational acceleration. The vertical depth at sea level, close to 1,030 g cm^{-2}, is standard atmospheric pressure written as a mass column.

The reason this quantity, and not the geometric path length, is the coordinate of shower development is that the probability of an interaction depends on the amount of matter traversed. Consider a particle crossing a layer of thickness $\mathrm{d}l$ in air of density ρ, with a cross-section σ per target nucleus of mass number A. The number density of targets is $n = \rho N_A/A$, so the interaction probability over the layer is $n\sigma\,\mathrm{d}l = (N_A\sigma/A)\,\rho\,\mathrm{d}l$. The density and the geometric length appear only through

the product $\rho\,\mathrm{d}l \equiv \mathrm{d}X$, so the survival probability after traversing a depth X is

$$P_{\mathrm{surv}}(X) = \exp\left(-X/\lambda\right) \ , \qquad \lambda = \frac{A}{N_A \sigma} \ , \tag{A.2}$$

where the interaction length λ is itself expressed in g cm^{-2}. The same argument holds for radiative processes, whose characteristic scale is the radiation length, $X_0 \approx 36.6$ g cm^{-2} in air (Navas et al., 2024). A particle does not respond to how far it has moved but to how much matter it has crossed, and depth is the variable that counts the latter. Expressing the atmosphere in metres would require carrying the altitude dependence of ρ through every step of the calculation; expressing it in g cm^{-2} absorbs that dependence once and for all.

A direct consequence is that the longitudinal profile of a shower is, to a good approximation, a universal function of X. Two showers of the same primary energy whose first interactions occur at different altitudes pass through the same sequence of generations, and their profiles coincide when expressed in depth even though they do not coincide when expressed in metres. This is why the Heitler model of Section 2.2.1 yields a depth of maximum, Equation 2.3, rather than a height of maximum, and why X_{max} is comparable across showers, experiments, and sites.

Expressed in these terms, the atmosphere is a calorimeter whose thickness is measured in radiation lengths: the vertical column at sea level amounts to $1{,}030/36.6 \approx 28\,X_0$, and less at altitude. That thickness is not a fixed property of the instrument but is set jointly by the observation level and by the arrival direction of each individual shower, which is the sense in which a ground-based array measures every cascade at whatever stage of development it happens to have reached on arrival.

A shower arriving at a zenith angle θ does not traverse the vertical column but the column along its own axis, known as the slant depth. For a plane-parallel atmosphere, the approximation adopted in CORSIKA (Heck et al., 1998), the two are related by

$$X(\theta) = X_v \sec\theta \ . \tag{A.3}$$

An inclined shower therefore crosses more matter before reaching the array than a vertical shower of the same energy: about 6% more at $\theta = 20°$ and about 30% more at $\theta = 40°$, the upper limit of the range simulated in this work. The approximation is accurate over that range and degrades only for nearly horizontal showers, where the curvature of the Earth can no longer be neglected. The physical consequence is that the zenith angle and the effective observation depth are two views of the same variable: a shower observed at larger θ is observed at a later stage of its development, with fewer surviving particles and a wider, more attenuated footprint. That systematic is what the angular reconstruction task exploits, and it is visible directly in the evolution of the mean footprint with θ shown in Figure 2.6.

The same relation accounts for the value of altitude in the design of a ground-based array. Table A.1 lists the vertical depth at several observation levels. Placing the array at 5,300 m removes roughly half of the atmospheric column that a detector at sea level must look through, and about 90 g cm^{-2} relative to HAWC. Since the number of particles surviving to ground falls steeply with the depth traversed beyond the shower maximum, and sub-TeV showers reach their maximum high in the atmosphere, that difference translates into a substantially larger number of secondaries at ground level for a primary of a given energy. This is the origin of the low energy threshold discussed in Section 2.4, and it is a property of the site rather than of the detector technology.

Table A.1: Vertical atmospheric depth at several observation levels, computed from the U.S. standard atmosphere. Values are rounded to the nearest 5 g cm^{-2}.

Observation level	**Altitude**	**Vertical depth**
Sea level	0 m	1,030 g cm^{-2}
HAWC	4,100 m	620 g cm^{-2}
LHAASO	4,410 m	595 g cm^{-2}
CONDOR	5,300 m	530 g cm^{-2}

A.2 Model Architecture Summary

Table A.2 provides a layer-by-layer breakdown of the CNN-Transformer multi-task model and its parameter counts, grouped by functional block. Each Transformer stream is a pre-norm encoder block; the CNN streams operate on the 128-dimensional backbone output and the RAW streams on the six-dimensional masked hit sequence, which is why the former account for roughly twenty times the parameters of the latter. All counts are read directly from the trained model.

Table A.2: Layer-by-layer summary of the CNN-Transformer model, with parameter counts grouped by functional block.

Component	**Layer**	**Parameters**
CNN backbone	Conv1D (128 filters, k=7) + BN + ELU	6,016
	Conv1D (128 filters, k=7) + BN + ELU	115,328
	Conv1D (128 filters, k=7) + BN + ELU	115,328
	MaxPool1D (stride=2)	0
	Dense(128) + ELU	16,512
	Dense(256) + ELU + Dropout(0.4)	33,024
	Dense(128) + ELU	32,896
Transformer × 6	3 CNN streams (LN + MHA + FFN)	150,432
	3 RAW streams (LN + MHA + FFN)	7,692
Output heads	Classification: Dense(128) + Dense(1, sigmoid)	18,049
	Angle: Dense(128) + Dense(1, linear)	18,049
	Energy: Dense(128) + Dense(1, linear)	18,049
Total parameters		**531,375**
Trainable		530,607
Non-trainable (BN statistics)		768

A.3 Global Feature Definitions

Table A.3 specifies the exact computation and ordering of the five global features as implemented in the preprocessing pipeline.

Table A.3: Global feature vector: computation, ordering, and normalization as implemented in the code.

Index	**Name**	**Computation**	**Normalization**
0	`total_particles`	$\sum_i \texttt{count}_i$	Min-max
1	`total_energy`	$\sum_i \varepsilon_i$	Min-max
2	`active_detectors`	Number of unique detector IDs	Min-max
3	`duration`	$t_{\max} - t_{\min}$	Min-max
4	`central_energy`	$\sum_{i \in \mathcal{C}} \varepsilon_i$	Min-max

$\mathcal{C}$: set of 16 central detector modules closest to the array origin.

A.4 Quality Factor Definition

The quality factor Q is a standard figure of merit in gamma-ray astronomy that quantifies the signal-to-noise improvement achieved by a classifier. For a given classification threshold, the gamma-ray efficiency ε_γ is the fraction of true gamma events correctly classified, and the proton survival rate ε_p is the fraction of proton events misclassified as gamma. The quality factor is defined as

$$Q = \frac{\varepsilon_\gamma}{\sqrt{\varepsilon_p}} \, . \tag{A.4}$$

A larger Q indicates a better signal-to-noise ratio after classification. For the model presented in this thesis, at the default threshold of $P(\gamma) = 0.5$, the quality factor evaluates to $Q = 5.48$ with $\varepsilon_\gamma = 0.971$.

A.5 Angular Resolution: PSF_{68} Definition

The angular resolution of a ground-based gamma-ray observatory is conventionally reported as a containment radius rather than as a mean error. The reason is that the distribution of angular residuals is not Gaussian: it is asymmetric and carries a tail of poorly reconstructed events, produced by showers whose core falls near the edge of the array or whose front is sampled by few detectors, which would dominate a mean-based figure of merit while affecting only a small fraction of the data. Characterizing the resolution through the shape of the residual distribution rather than through a single moment is long-established practice for scintillator arrays (Prahl et al., 1993), and containment radii are the form in which contemporary ground-based arrays such as HAWC report their angular performance (Abeysekara et al., 2023).

For each test event, the *angular residual* is the difference between the reconstructed and true zenith angles, $\hat{\theta} - \theta_{\mathrm{true}}$, and the absolute angular error is $\Delta\theta = |\hat{\theta} - \theta_{\mathrm{true}}|$. The point spread function at 68% containment, PSF_{68}, is the 68th percentile of the distribution of $\Delta\theta$ over the test set:

$$\mathrm{PSF}_{68} = \inf \left\{ r \geq 0 \ : \ P\left(\Delta\theta \leq r\right) \geq 0.68 \right\} \, . \tag{A.5}$$

In words, PSF_{68} is the angular radius around the true direction within which 68% of reconstructed events fall. The 68% level is chosen because it corresponds to the fraction of a one-dimensional Gaussian distribution contained within one standard deviation, which makes the quantity directly comparable to a resolution quoted as σ for instruments whose residuals happen to be Gaussian, without assuming that they are. For the model presented in this thesis, $\mathrm{PSF}_{68} = 0.50°$.

Scope of this definition. An observatory normally quotes PSF_{68} over the *space angle* ψ, which combines the zenith and azimuth residuals into a single separation on the sky. Here the azimuth is fixed at $\phi = 0°$ (Section 2.5) and is not reconstructed, so the value above is a zenith-only containment radius: a lower bound on the space-angle PSF_{68} that the same architecture would reach with the azimuth left free. Comparisons with observatories that reconstruct both angles should be read with that restriction in mind.