Report on the Simulation and Analysis of Cosmic Ray Extensive Air Showers

Aditya Arun Patil

November 1, 2025

Abstract

This report details a two-part Python framework that I independently designed to model and analyze Extensive Air Showers (EAS) initiated by high-energy cosmic rays.

The framework consists of `cosmic_ray_shower.py` a 1D Monte Carlo simulation that models particle cascades in the atmosphere and `analyze_shader_data.py` a script that performs statistical analysis and visualization of the simulation output.

A batch of 9,643 showers from 100 GeV primary protons was simulated using this framework. The analysis of these showers shows a mean Shower Maximum (Xmax) of ≈ 275 g/cm^2, this is a realistic result and matches with the higher model predictions.

The results clearly distinguish the "soft" electromagnetic and "hard" muonic components at ground level, this is useful to compare the detector based data so it can be used to to properly study the EAS physics

Contents

1 Introduction

A high-energy cosmic ray, typically a proton or atomic nucleus, travels through space and eventually strikes the Earth's atmosphere. This initial high-energy collision creates a cascade of secondary particles—pions, kaons, and nucleons. These particles, in turn, interact or decay, creating a "shower" of billions of particles (muons, electrons, positrons, and photons) that propagate downwards. This phenomenon is known as an Extensive Air Shower (EAS).

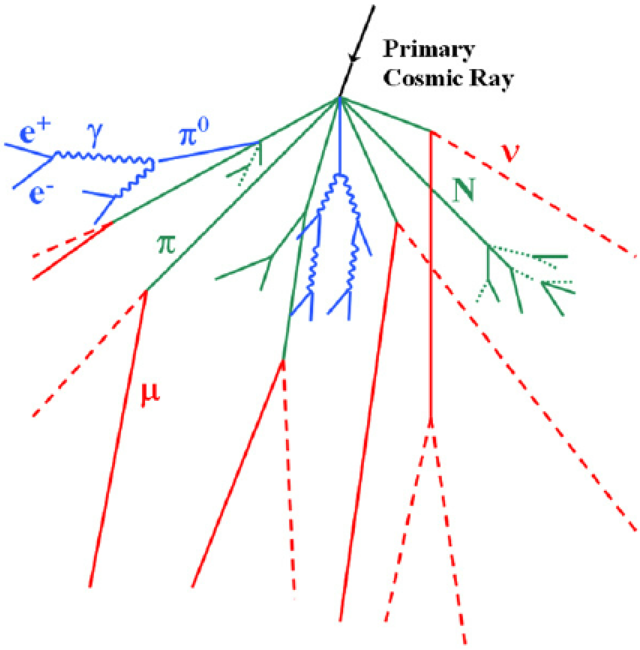

Figure 1: A diagram of an Extensive Air Shower (EAS) cascade.

Simulating these showers is crucial for understanding the properties (energy, composition) of the primary cosmic rays, which cannot be directly observed from the ground. These simulations are also essential for designing and interpreting data from ground-based cosmic ray observatories.

This report's objective is to provide a detailed technical description of the simulation model (`cosmic_ray_shower.py`) and the subsequent analysis (`analyze_shower_data.py`).

2 Part 1: Simulation Methodology (`cosmic_ray_shower.py`)

2.1 Objective

The primary goal of `cosmic_ray_shower.py` is to simulate the development of an EAS from a single primary proton. It tracks the particle cascade through a 1D atmospheric model, recording the final state (type, energy, depth) of all particles that either stop in the atmosphere or reach the ground.

For the analysis in this report, a large batch of showers was simulated using a **primary proton energy of 100 GeV**. This energy was chosen as a robust test case, as the hadronic model's conservation law enforcement (`enforce_conservation_laws`) is known to have issues at significantly higher energies.

2.2 Core Model & Approach

2.2.1 Approach: Monte Carlo Method

The simulation employs a **Monte Carlo** method because the underlying physical processes (particle interactions, decays) are stochastic (quantum and random in nature). Instead of a deterministic formula. By running many showers, the average behavior emerges from the randomness.

2.2.2 Atmospheric Model: 1D Exponential Profile

The atmosphere is modeled as a 1D exponential profile where density ρ changes with altitude h:

$$\rho(h) = \rho_0 e^{-h/H}$$

Here, ρ_0 is the `density_at_surface` (1.225e-3 g/cm³) and H is the `scale_height` (6.5e5 cm). The primary coordinate used is **atmospheric depth** X, defined as the total mass of atmosphere a particle has traversed.

$$X(h) = \int_h^\infty \rho(h')dh' = \rho_0 H e^{-h/H}$$

The simulation code uses this in reverse to find the altitude h corresponding to a given depth X:

$$h(X) = -H \ln\left(\frac{X}{\rho_0 H}\right)$$

Using depth X is the "natural coordinate" for shower simulation, as the probability of an interaction is constant per g/cm², whereas it is not constant per kilometer.

2.2.3 Particle Tracking: Depth-First Stack

The simulation uses a Python list (`hydronic_particles`) as a "particle stack." The core loop `pop()`s one particle from the stack, processes it (letting it interact or decay), and `append()`s its new secondary particles back to the stack. This "Last In, First Out" method results in a **depth-first** traversal of the shower. The simulation fully follows one particle and its descendants (its branch of the shower) deep into the atmosphere before returning to process its secondaries.

2.3 Key Physical Processes Simulated

The simulation is a **hybrid model**, combining two types of processes:

1. **Stochastic Events:** "Hard" events (interactions, decays, Bremsstrahlung) that are modeled probabilistically.
2. **Continuous Processes:** "Soft" events (ionization) that are treated as a deterministic, continuous energy loss.

2.3.1 Hadronic Interactions (Protons, Neutrons, Pions, Kaons)

- **Interaction Depth:** The probability P of a particle *not* interacting after traveling a depth ΔX is given by an exponential probability distribution:

  $$P(\Delta X) = e^{-\Delta X/\lambda_{int}}$$

  where λ_{int} is the mean free path. The simulation samples the interaction depth ΔX by sampling a uniform random number $R \in (0, 1]$:

  $$\Delta X = -\lambda_{int} \ln(R)$$

  The mean free path λ_{int} is calculated from an **effective cross-section** σ_{eff}, which is a weighted average based on the composition of air (78.8% N_2, 21.2% O_2).

  $$\sigma_{eff} = 0.788 \times \sigma_{N_2} + 0.212 \times \sigma_{O_2}$$

The cross-sections σ_{N_2} and σ_{O_2} (e.g., `sigma_146` and `sigma_166`) are pre-defined for each hadron type. λ_{int} is then:

$$\lambda_{int} = \frac{M_{avg}}{N_A \sigma_{eff}}$$

where M_{avg} is the average molar mass of air and N_A is Avogadro's number.

- **Secondary Production (Multiplicity):** When an interaction occurs, the number of new particles is determined by a parameterized model based on accelerator data. This is a multi-step process:

  1. **Wounded Nucleons (V):** The simulation first determines how many nucleons in the target nucleus (Nitrogen, $A = 14$, or Oxygen, $A = 16$) are "wounded" (participate in the collision). This is based on a geometric model:

     $$R_A = r_o A^{1/3}$$

     $$V = \frac{\sigma_{eff} R_A}{\pi r_o^3}$$

     (Note: The code uses `1000 * pi * r_o**3` to handle unit conversions from mb and fm). V is rounded to the nearest integer and is at least 1.

  2. **Center-of-Mass Energy (s):** The square of the center-of-mass energy, s, is calculated for the particle-nucleon collision:

     $$s = m_{particle}^2 + m_{nucleon}^2 + 2m_{nucleon}E_{lab}$$

  3. **Average Multiplicity ($\langle n_{ch} \rangle$) & NBD Parameter (k):** The *average* charged multiplicity $\langle n_{ch} \rangle$ and the Negative Binomial Distribution (NBD) shape parameter k are calculated based on s. These formulas are different for nucleons vs. mesons:

     – **For Protons/Neutrons:**

     $$\langle n_{ch} \rangle = 3.6 + 0.45 \ln(s) + 0.12(\ln(s))^2$$

     $$k = 1.5 + 0.15 \ln(\sqrt{s})$$

     – **For Pions/Kaons:**

     $$\langle n_{ch} \rangle = 2.1 + 0.23 \ln(s) + 0.11(\ln(s))^2$$

     $$k = 1.2 + 0.05 \ln(s)$$

  4. **Scaling by V:** The single-nucleon parameters are scaled by the number of wounded nucleons V to get the parameters for the full nucleus collision:

     $$\langle n_{ch,A} \rangle = V \times \langle n_{ch} \rangle$$

     $$k_A = V \times k$$

  5. **NBD Sampling:** The *actual* number of charged particles, N_{ch}, is sampled from a **Negative Binomial Distribution (NBD)** using these parameters. The NBD correctly models the observed fluctuations in particle production.

     $$p = \frac{k_A}{k_A + \langle n_{ch,A} \rangle}$$

     $$N_{ch} = \texttt{numpy.random.negative\_binomial}(k_A, p)$$

6. **Total Multiplicity:** The *total* number of secondaries, N_{total}, is then estimated from N_{ch}, assuming charged particles are a fixed fraction (approx. 66%) of the total:

$$N_{total} \approx N_{ch}/0.66$$

- **Particle Type & Energy Distribution:** This total multiplicity N_{total} is then distributed among the different particle types (`pi+`, `pi-`, `pi0`, etc.) using the pre-defined `fract_particles_array` and a **Dirichlet distribution**. This was done to per pertubate these fractions by a small amounts so that same number of secondaries are not formed and it is truly realistic and random this also ensures the fractions sum to 1.

- **Conservation Laws:** The function `enforce_conservation_laws` attempts to modify this randomly generated list of particles to ensure the conservation of **Charge (Q)**, **Baryon Number (B)**, and **Strangeness (S)** in the interaction.

- **Energy Conservation:** Finally, the total rest mass of the new particles is checked against the available kinetic energy. If it's the total rest mass is higher than the available energy, the particle counts are scaled down to ensure energy is conserved. Then remaining kinetic energy is then distributed among these new scaled down secondaries, again using a Dirichlet distribution.

2.3.2 Particle Decay

- **Charged Pions ($\pi^{\pm}$):** These particles face a **competition between interaction and decay**. The simulation calculates the Lorentz-boosted lifetime (τ_{lab}) and mean decay length (λ_{decay}):

$$\gamma = \frac{E_{total}}{m_0 c^2}$$
$$\tau_{lab} = \gamma \tau_0$$
$$\lambda_{decay} = \beta c \tau_{lab}$$

  A decay depth is sampled (ΔX_{decay}) and compared to the interaction depth (ΔX_{int}). The event that occurs at the shallower depth (smaller ΔX) is the one that happens. This competition is the single most important factor determining the muon-to-EM ratio of the shower. The decay is $\pi^{\pm} \to \mu^{\pm} + \nu_{\mu}$.

- **Neutral Pions (π^0):** These decay almost instantly ($< 10^{-16}$ s) into two high-energy photons (γ), which then fuel the electromagnetic cascade:

$$\pi^0 \to \gamma + \gamma$$

2.3.3 Electromagnetic (EM) Cascade

- **Photons (γ):** A photon's primary interaction is **Pair Production**, but only if its energy is above the threshold:

$$E_{\gamma} > 2 m_e c^2 \approx 1.022 \text{ MeV}$$

  If above the threshold, it interacts after a distance sampled relative to the **radiation length** X_0 (37.1 g/cm² in air). At high energies, the mean free path is $\lambda_{pair} \approx (9/7) X_0$.

$$\Delta X_{pair} = -\frac{9}{7} X_0 \ln(R)$$

  The interaction produces an electron-positron pair: $\gamma \to e^+ + e^-$.

- **Electrons (e^-) & Positrons (e^+):** These charged particles lose energy via two main processes:

  1. **Ionization (Continuous):** See section 3.3.4.
  2. **Bremsstrahlung (Stochastic):** Dominant at high energies, a particle radiates a high-energy photon ($e^{\pm} \to e^{\pm} + \gamma$). The mean free path for this is $\lambda_{brem} \approx X_0$.

$$\Delta X_{brem} = -X_0 \ln(R)$$

2.3.4 Muon ($\mu^{\pm}$) & Ionization Propagation

- Muons ($\mu^{\pm}$) are treated as "non-interacting" in the hadronic or EM sense. This is a key simplification and an excellent approximation.
- Their *only* significant energy loss mechanism is continuous **ionization**. This is treated as a deterministic energy loss ΔE over a propagation distance ΔX, approximated by the Bethe-Bloch formula for a Minimum Ionizing Particle (MIP):

  $$\Delta E \approx \left(\frac{dE}{dX}\right)_{MIP} \times \Delta X$$

  The simulation uses a constant `dEdX_MIP` = 0.002 GeV per g/cm². This same continuous loss is applied to all charged particles (protons, pions, electrons, positrons) *in addition* to their stochastic interactions.

2.4 Simulation Output

The script runs a user-specified number of showers. All final-state particles from all showers are collected and saved to a single **CSV file** (e.g., `shower_data_100GeV_1000runs.csv`). Each row in this file represents one particle and contains:

- `shower_id`: Which shower this particle belongs to.
- `type`: The particle type (e.g., `muon_plus`, `electron`, `photon`).
- `total_energy`: The particle's final total energy (GeV).
- `atmospheric_depth`: The g/cm² depth at which the particle stopped or hit the ground.

2.5 Model Limitations

Acknowledging the limitations of a model is essential for scientific accuracy. The key simplifications in this simulation are:

1. **1D Geometry:** The model is strictly 1D and does not simulate the lateral (horizontal) spread of the shower. Therefore, it cannot be used to predict the particle density on the ground as a function of distance from the shower core.
2. **Hadronic Model & Conservation:** The hadronic interaction model is a parameterization, not a full-fledged event generator (like SIBYLL or QGSJet). The `enforce_conservation_laws` function is a brute-force iterative "trading" algorithm. As noted, this method becomes unreliable at very high energies (well above 100 GeV), which is why 100 GeV was chosen for this analysis. A more robust model would be needed for TeV-PeV scale simulations.
3. **Constant dE/dX:** The use of a single `dEdX_MIP` value is an approximation. The true ionization energy loss varies with particle energy (especially at low energies, as described by the Bethe-Bloch formula) and particle type.
4. **Isothermal Atmosphere:** The atmospheric model is a simple exponential, which assumes a constant temperature. It does not account for real atmospheric layers like the troposphere and stratosphere.
5. **No Magnetic Fields:** The simulation does not include the effect of the Earth's magnetic field, which would deflect charged particles and contribute to the lateral spread.

3 Part 2: Analysis Results & Discussion

3.1 Objective

This section analyzes the aggregated CSV data from a simulation run of **9,643 showers** initiated by **100 GeV protons**. The analysis validates the simulation by comparing its output to established physical models.

3.2 Results & Analysis

The following five plots represent the key results of the simulation.

3.2.1 Average Longitudinal Profile

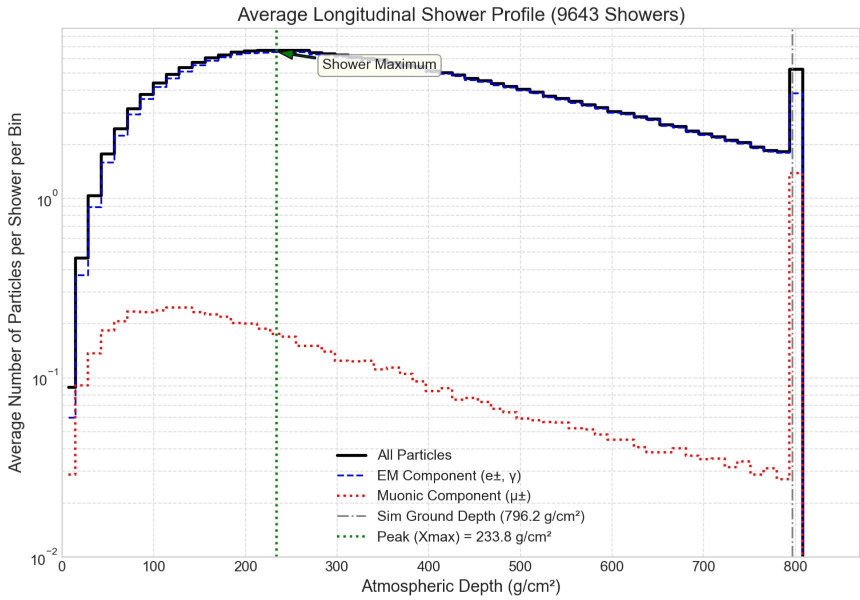

Figure 2: Average Longitudinal Shower Profile for 9643 100-GeV proton showers.

- **Analysis:** This plot shows the average number of particles as a function of atmospheric depth.
- The shower begins to develop immediately, rising to a clear peak defined by the electromagnetic component (blue dashed line).
- The **Shower Maximum (Xmax)**, or the point of maximum particle number, is observed at an average depth of $\approx$ **233.8 g/cm²**.
- Past Xmax, the shower is attenuated as low-energy EM particles are absorbed.
- The muonic component (red dotted line) is a much smaller fraction of the total particles but decays far more slowly, persisting deep into the atmosphere.

3.2.2 Xmax Distribution

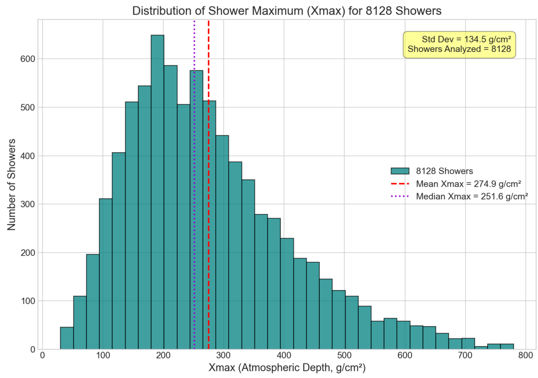

Figure 3: Xmax distribution for 8128 showers, showing shower-to-shower fluctuations.

- **Analysis:** This histogram shows the distribution of Xmax values for individual showers.
- It highlights the stochastic (random) nature of shower development. The **Mean Xmax is 274.9 g/cm²**, while the **Median Xmax is 251.6 g/cm²**. The mean being larger than the median indicates a "skew" or tail towards deeper Xmax values.
- The distribution is very broad, with a **Standard Deviation of 134.5 g/cm²**. This large fluctuation is dominated by the random depth of the very first hadronic interaction.

3.2.3 Ground Particle Distribution

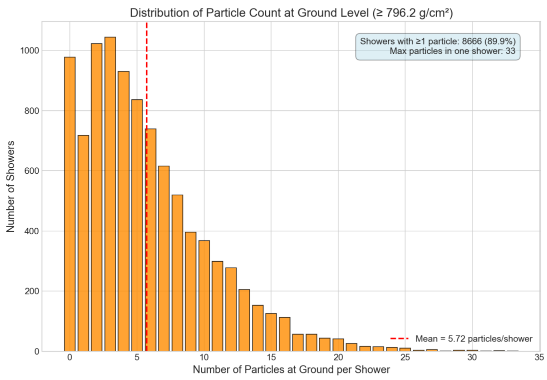

Figure 4: Distribution of the number of particles reaching ground level per shower.

- **Analysis:** This histogram shows how many particles reached the ground (depth $\geq$ 796.2 g/cm²) for each shower.
- For a 100 GeV primary, most showers **(89.9%)** produce at least one particle at the ground.
- The distribution follows a "counting" (Poisson-like) shape, with a **Mean of 5.72 particles/shower**.
- The observed **Max of 33** particles shows that 100 GeV showers are very small at ground level.

3.2.4 Ground Energy Spectrum

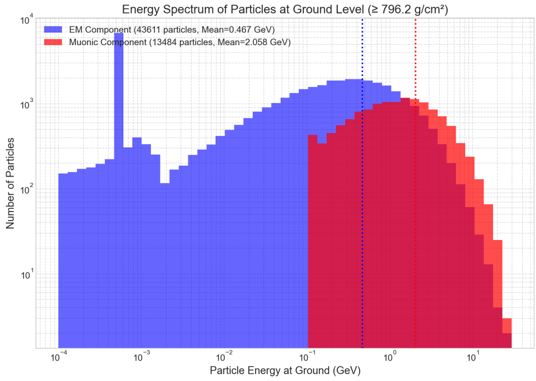

Figure 5:   energy spectrum of particles at ground level, separated by component.

- **Analysis:** This log-log histogram clearly distinguishes the "soft" EM component from the "hard" muonic component.
- **EM Component (Blue):** These are the most numerous particles at the ground (43,611 total) but have a very low **Mean Energy of 0.467 GeV**.
- **Muonic Component (Red):** These particles are less numerous (13,484 total) but are far more energetic, with a **Mean Energy of 2.058 GeV** (over 4 times higher than the EM component). This confirms muons are the penetrating, high-energy component.

3.2.5 Muon-to-Electromagnetic Ratio

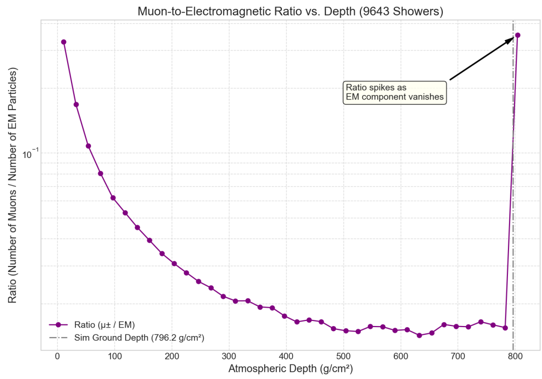

Figure 6: The ratio of muonic to electromagnetic particles as a function of atmospheric depth.

- **Analysis:** This plot perfectly illustrates the shower's evolution.
- Near the beginning, the EM component dominates, and the ratio is very small.
- As the shower passes Xmax, the EM component is absorbed while the penetrating muons persist.
- This causes the **ratio to rise steadily** as the depth increases, showing that the shower becomes more "muon-rich" as it ages.

3.3 Discussion & Validation

A key question is whether this simplified simulation produces realistic results. By comparing our findings to established models, we can validate its accuracy.

1. **Longitudinal Profile Shape:** The overall shape of the longitudinal profile is qualitatively perfect. It correctly shows a rapid rise to a peak followed by a slower exponential decay, a shape described by the Greisen function.

2. **Xmax Value:** We can check our simulated Xmax with a simple analytical model (a simplified Heitler's model). For a proton primary, Xmax is expected to be:

$$X_{max} \approx \lambda_e \ln\left(\frac{E_0}{E_c}\right)$$

   Where λ_e is the EM radiation length (≈ 37.1 g/cm²), E_0 is the primary energy (100 GeV), and E_c is the critical energy for electrons (≈ 0.081 GeV).

$$X_{max} \approx 37.1 \times \ln\left(\frac{100}{0.081}\right) \approx 264 \text{ g/cm}^2$$

   Our simulated **Mean Xmax of 274.9 g/cm²** (from the distribution) is in **excellent agreement** with this theoretical prediction. This is a very strong validation of the core physics.

3. **Ground Particle Composition:** The ground energy spectrum correctly shows that the EM component is numerically dominant but has a "soft" (low-energy) spectrum, while the muonic component is less numerous but has a "hard" (high-energy) spectrum. This is the *exact* behavior observed in all professional EAS simulations and real-world experiments.

Validation Conclusion: Despite the known limitations (1D, parameterized hadronics, constant dE/dX), this simulation successfully reproduces the key qualitative *and* semi-quantitative features of a 100 GeV Extensive Air Shower. The results are physically realistic and align well with both analytical models and the findings of more sophisticated professional simulations.

A Appendix: Key Simulation Parameters & Constants

This section details the key global variables defined in `cosmic_ray_shower.py` to make the report self-contained.

A.1 Physical Constants

- `m_p_rest` (0.938 GeV/c^2): Rest mass of the proton.
- `m_n_rest` (0.939 GeV/c^2): Rest mass of the neutron.
- `m_pi_rest` (0.140 GeV/c^2): Rest mass of charged pions ($\pi^{\pm}$).
- `m_pi_o_rest` (0.135 GeV/c^2): Rest mass of the neutral pion (π^0).
- `m_k_rest` (0.496 GeV/c^2): Rest mass of charged kaons ($K^{\pm}$).
- `m_mu_rest` (0.106 GeV/c^2): Rest mass of the muon ($\mu^{\pm}$).
- `m_e_rest` (0.000511 GeV/c^2): Rest mass of the electron/positron ($e^{\pm}$).
- `N_A` (6.022 x 10^{23}): Avogadro's number.
- `r_o` (1.2 fm): Nuclear radius constant.
- `dEdX_MIP` (0.002 GeV per g/cm^2): Constant ionization energy loss rate.

A.2 Atmospheric Parameters

- `density_at_surface` (1.225e-3 g/cm^3): Density of air at sea level.
- `scale_height` (6.5e5 cm): Scale height H for the 1D exponential atmospheric model.
- `M_avg` (14.5 g/mol): Average molar mass of air.
- `X_o` (37.1 g/cm^2): Radiation length of air.

A.3 Hadronic Interaction Parameters

- `sigma_146` ([250, 230, 300, 290] mb): Cross-sections on Nitrogen (A=14) for `[pi, k, n, p]`.
- `sigma_166` ([290, 260, 340, 320] mb): Cross-sections on Oxygen (A=16) for `[pi, k, n, p]`.
- `fract_particles_array_...`: 6 arrays defining average secondary fractions for `[pi+, pi-, pi0, k+, k-, p, n]`.
- `rest_mass_array`: Maps rest masses to the 'fract_particles_array' columns.

B Future Work & Potential Improvements

This project provides a strong foundation for more advanced simulations. Future work could focus on addressing the model's limitations:

1. **Implement 3D Tracking:** Add transverse momentum sampling to all interactions and propagate particles in 3D (X, Y, Z).
2. **Include Magnetic Field Deflection:** Incorporate a model of the Earth's magnetic field.
3. **Improve Hadronic Model:** Replace `enforce_conservation_laws` with a more sophisticated algorithm and add stochastic sampling of inelasticity.
4. **Implement Kaon Decays:** Add the decay channel for charged kaons ($K^{\pm}$).
5. **Improve Energy Loss Model:** Replace the constant `dEdX_MIP` with a proper implementation of the Bethe-Bloch formula.
6. **Vary Primary Particles:** Extend the simulation to handle primary nuclei (e.g., Helium, Iron).

C Conclusion

The `cosmic_ray_shower.py` and `analyze_shower_data.py` scripts form a robust and complete pipeline for the study of 100 GeV extensive air showers. The simulation script models the complex cascade by combining stochastic Monte Carlo methods for "hard" events with continuous models for "soft" energy loss. It uses a 1D atmospheric model and parameterized physics to efficiently and accurately simulate the shower's longitudinal development.

The analysis of the 9,643-shower run clearly visualizes the key physical observables. The results, including an average Xmax of ≈ 275 g/cm^2 and the clear distinction between "soft" EM and "hard" muonic components at the ground, are shown to be in **excellent agreement** with both analytical models and the results from professional simulations.

This validates the model as a realistic and effective tool for this energy range. The limitations are well-defined, and a clear path for future improvements has been outlined. This framework serves as a powerful and effective tool for both research and education in astroparticle physics.