The air shower simulation framework CORSIKA 8: Development and first applications to muon production

zur Erlangung des akademischen Grades

DOKTOR DER NATURWISSENSCHAFTEN

der KIT-Fakultät für Physik des
Karlsruher Instituts für Technologie
vorgelegte

DISSERTATION

von

M. SC. MAXIMILIAN REININGHAUS

Tag der mündlichen Prüfung: 12. November 2021
Referent: Prof. Dr. Ralph Richard Engel
Korreferent: Dr. Diego Gabriel Melo
Betreuer: Dr. Ralf Matthias Ulrich

The air shower simulation framework CORSIKA 8: Development and first applications to muon production

Tesis presentada para optar por el título de

DOCTOR EN ASTROFÍSICA

del Instituto de Tecnología *Prof. Jorge A. Sábato* de la
Universidad Nacional de San Martín
por

M. SC. MAXIMILIAN REININGHAUS

Fecha de la defensa oral: 12 de Noviembre de 2021
Director: Prof. Dr. Ralph Richard Engel
Co-director: Dr. Diego Gabriel Melo
Colaborador: Dr. Ralf Matthias Ulrich

Abstract

Tools to accurately simulate extensive air showers are a key asset for the understanding of ultra-high energy cosmic rays. In this thesis, the Monte Carlo air shower simulation framework CORSIKA 8 is presented. CORSIKA 8 constitutes a next-generation code that aims to combine new functionality with a high level of flexibility and modularity. Notable aspects include the ability to freely combine an arbitrary number of physical processes and to setup simulation environments consisting of several media, including custom atmospheric models. A special feature is the possibility to inspect the complete lineage of particles, which allows linking particles on ground with any of their preceding generations.

After describing the foundations of Monte Carlo shower simulations, I explain the architecture of CORSIKA 8 in depth. Focusing on the hadronic and muonic shower components, results obtained with CORSIKA 8 and other simulation codes are compared with each other. Even when using the same hadronic interaction models, a number of differences are observed, in particular regarding low-energy interactions, which have a considerable impact on the lateral distribution of muons at kilometre-scale distances up to a factor of two and more.

Making use of the lineage technique, I study the phase space of hadronic interactions in order to quantify the importance for muon production and compare the results with the Heitler-Matthews toy model. At high energies ($\sqrt{s} \gtrsim 500\,\mathrm{GeV}$) particle production in the forward region is confirmed to be especially important, while the central region becomes relevant at low energies ($\sqrt{s} \lesssim 50\,\mathrm{GeV}$) in particular for muons at large distances.

Additionally, I study the impact of modified hadronic interactions on air shower observables. Modified hadron-air cross-sections mainly affect the longitudinal development, causing a larger shift of the maximum muon production depth than of the shower maximum. Artificially increased ρ^0 production, on the other hand, can greatly increase the number of muons with only small impact on other observables.

Finally, I also consider the possibility of large multiplicity boson production in the first interaction and study its phenomenology in air showers with a simple toy model.

Within the scope of this thesis, I developed the foundations of the CORSIKA 8 framework. Based on the studies that have become possible with CORSIKA 8, I point out some new opportunities towards an improved understanding of muons in air showers.

Zusammenfassung

Werkzeuge zur akkuraten Simulation von Luftschauern sind ein Schlüsselfaktor für das Verständnis ultrahochenergetischer kosmischer Strahlung. In dieser Arbeit wird das Monte-Carlo-Luftschauer-Simulationsframework CORSIKA 8 präsentiert. CORSIKA 8 stellt einen neuen Code dar, der das Ziel hat, neue Funktionalität mit einem hohen Grad an Flexibilität und Modularität zu vereinen. Bemerkenswerte Aspekte beinhalten die Fähigkeit, eine beliebige Anzahl physikalischer Prozesse zu kombinieren sowie Simulationsumgebungen aus mehreren Medien, darunter auch benutzerdefinierte Atmosphärenmodelle, zusammenzusetzen. Ein Spezifikum ist die Möglichkeit, die vollständige Abstammungslinie von Teilchen zu inspizieren, was es ermöglich, Teilchen auf dem Boden mit jedem ihrer Vorgängergenerationen in Verbindung zu bringen.

Nach der Beschreibung der Grundlagen von Monte-Carlo-Schauer-Simulationen im Allgemeinen erkläre ich eingehend die Architektur von CORSIKA 8. Anschließend werden Ergebnisse von Simulationen mit CORSIKA 8 und anderen Simulationscodes in Bezug auf hadronische und myonische Schauerkomponenten miteinander verglichen. Selbst bei Nutzung der gleichen hadronischen Wechselwirkungsmodelle werden einige Unterschiede sichtbar, insbesondere die niederenergetischen Wechselwirkungen betreffend, welche einen deutlichen Einfluss auf die Lateralverteilung von Myonen im Kilometer-Bereich bis zu einem Faktor zwei und mehr haben.

Mit der Technik der Abstammungslinie untersuche ich den Phasenraum hadronischer Wechselwirkungen, um die Wichtigkeit für die Myonenproduktion zu quantifizieren, und vergleiche die Ergebnisse mit dem Heitler-Matthews-Spielzeugmodell. Während bei hohen Energien ($\sqrt{s} \gtrsim 500\,\mathrm{GeV}$) die besondere Wichtigkeit der Teilchenerzeugung in Vorwärtsrichtung herausgestellt wird, gewinnt bei niedrigen Energien ($\sqrt{s} \lesssim 50\,\mathrm{GeV}$)) auch die Zentralregion an Bedeutung insbesondere für Myonen bei großen Abständen.

Außerdem untersuche ich den Einfluss veränderter hadronischer Wechselwirkungen auf Luftschauerobservablen. Veränderte Hadron-Luft-Wirkungsquerschnitte beeinflussen hauptsächlich die longitudinale Entwicklung, wobei beim Maximum der Myonenproduktionstiefe eine größere Verschiebung verursacht wird als beim Schauermaximum. Auf der anderen Seite erhöht eine künstlich vergrößerte ρ^0-Produktion die Myonenzahl beträchtlich, ohne andere Observablen stark zu beinflussen.

Zuletzt betrachte ich die Möglichkeit einer Produktion von Bosonen mit hoher Multiplizität in der ersten Wechselwirkung und studiere die Phänomenologie in Luftschauern mittels eines einfachen Spielzeugmodells.

Im Rahmen dieser Arbeit habe ich die Grundlagen des CORSIKA-8-Framworks entwickelt. Basierend auf den Studien, die durch CORSIKA 8 möglich wurden, zeige ich neue Möglichkeiten auf, die zu einem besseren Verständnis von Myonen in Luftschauern beitragen.

Resumen

Herramientas para simular con precisión lluvias atmosféricas extendidas son clave para la comprensión de los rayos cósmicos de ultra alta energía. En esta tesis se presenta el marco de simulación Monte Carlo de lluvias atmosféricas extendidas CORSIKA 8. CORSIKA 8 constituye un código de nueva generación que tiene como objetivo combinar nuevas funcionalidades con un alto nivel de flexibilidad y modularidad. Aspectos destacables son la capacidad de combinar libremente un número arbitrario de procesos físicos y de configurar entornos de simulación que combinan diferentes medios. Inclusive son posibles modelos atmosféricos personalizados. Una característica especial es la posibilidad de inspeccionar el linaje completo de partículas, lo cual permite conectar partículas que alcanzan el nivel del suelo con cualquiera de sus generaciones anteriores.

Luego de describir los fundamentos de las simulaciones Monte Carlo de lluvias atmosféricas extendidas se explica en detalle la arquitectura de CORSIKA 8. Con foco en los componentes hadrónico y muónico de las lluvias, se comparan los resultados obtenidos con CORSIKA 8 con los de otros códigos de simulación. Incluso cuando se utilizan los mismos modelos de interacción hadrónica, se observa una serie de diferencias, en particular con respecto a las interacciones de baja energía. Estas tienen un impacto considerable en la distribución lateral de muones a distancias del órden del kilómetro de hasta un factor de dos o más.

Haciendo uso de la técnica del linaje, se estudia el espacio de fase de las interacciones hadrónicas con el fin de cuantificar la importancia para la producción de muones y comparar los resultados con el modelo de juguete de Heitler-Matthews. A altas energías ($\sqrt{s} \gtrsim 500\,\text{GeV}$) la producción de partículas en la dirección hacia delante es especialmente importante, mientras que la región central se vuelve relevante a bajas energías ($\sqrt{s} \lesssim 50\,\text{GeV}$), en particular para muones a altas distancias.

Además se estudia el impacto que tienen interacciones hadrónicas modificadas en las observables de las lluvias atmosféricas extendidas. Modificaciones en la sección eficaz de hadrón-aire afectan principalmente al desarrollo longitudinal, generando un mayor corrimiento de la profundidad de máxima producción de muones que del punto de máximo desarrollo de la lluvia. Un aumento artificial de la producción de ρ^0, por otro lado, puede aumentar en gran medida el número de muones con un impacto pequeño en otras observables.

Finalmente, también se considera la posibilidad de una multiplicidad alta de bosones en la primera interacción y se estudia su fenomenología en lluvias atmosféricas extendidas haciendo uso de un modelo de juguete sencillo.

Dentro del alcance de esta tesis, se desarrollaron los fundamentos del marco CORSIKA 8. Basado en los estudios que han sido posibles con CORSIKA 8, se señalan algunas oportunidades nuevas para mejorar la comprensión de los muones en lluvias atmosféricas extendidas.

Contents

1 Introduction

Earth is perpetually hit by high-energy particles pervading the cosmos. A small fraction of these, so-called ultra-high energy cosmic rays (UHECRs), are the most energetic particles known in the Universe. They are accelerated to values much higher than any man-made accelerator could ever achieve. The study of UHECR promises to provide insight into the most violent regions of space, e.g. active galactic nuclei or certain classes of supernovae, where they are believed to originate from [1], and at the same time into particle physics at extreme energies [2].

Due to their scarcity, UHECR can be measured virtually only indirectly via the formation of extensive air showers (EAS). In these, a plethora of aspects of particle physics at energy scales ranging from a few MeV (10^6 eV) up to ZeV (10^{21} eV) play a role, making them highly complex objects [3]. They are the objects that connect microscopic particle physics with the astrophysics of UHECR. Not surprisingly, research in the field of UHECR detection and reconstruction is heavily dependent on accurate simulations of EAS [4].

Currently, the most widely used Monte Carlo simulation code for EAS is CORSIKA (Cosmic Ray Simulations for KASCADE [5]) [4]. Despite being a supporting pillar of the field for decades and still being updated regularly, it has become evident that a continued development cannot adequately satisfy the growing requirements of the community [6]. For UHECR-related applications, a major drawback is the inability to simulate showers in multiple media, e.g. passing through air into ice. Furthermore, handles on the internal mechanisms of such simulations, which are necessary to study systematic effects in particle production, are lacking. The code of CORSIKA, however, is fundamentally monolithic and has outgrown its original use-case by far. It is becoming an unbearable burden to extend, let alone to entirely overhaul, the legacy codebase. The CORSIKA 8 collaboration has formed to address these challenges by developing a new C++ framework [6]. With its modular and flexible design, CORSIKA 8 aims to be a future-proof successor of the legacy version. The work reported here focuses on the development of a first version capable of simulating the hadronic and muonic shower components and its usage to study muon production in EAS.

In the first part of this thesis, important groundwork is established. After a brief introduction to the physics of UHECR and EAS in chapter 2 and an overview of EAS simulations in chapter 3, we consider the mathematical details of particle propagation in chapter 4. A particular emphasis lies on the interplay between the equations of motion that govern the evolution of the particle state and the probability distributions of stochastic events. We present a new sampling method that offers several advantages over existing approaches. The architecture of CORSIKA 8 is explained in chapter 5, with the most significant original contributions to the project in sections 5.3.1 to 5.3.3, 5.4, 5.7 and 5.8.

In many regards, CORSIKA 8 allows performing simulations with a level of detail on par

with existing tools. A number of aspects of CORSIKA 8, however, go beyond what was possible before, including a great deal of flexibility regarding the atmospheric models, and more generally the medium in which showers evolve, as well as the physical processes considered. A particularly noteworthy and unique feature of CORSIKA 8 developed as a cornerstone of this thesis is the ability to conserve and inspect the complete history ("lineage") of particles in the shower.

The further focus of this thesis lies on hadronic interactions in air shower simulations and their role in the production of muons. Models of hadronic interactions suffer from the fact that the bulk of them cannot be calculated from first principles and must rely largely on phenomenological approaches [7]. The phase space particularly relevant for the development of the hadronic cascade is difficult to access even with dedicated accelerator experiments [8]. This lack of constraints results in significant modelling uncertainties in these regions, also affecting important air shower observables [9]. Studies performed, among others, at the Pierre Auger Observatory show that at the highest energies simulations using any of the up-to-date hadronic interaction models show a deficit regarding the total number of muons by more than 40 % compared to data [10].

With CORSIKA 8 ready to simulate hadron-muon cascades with several hadronic interaction models, extensive comparisons with a number of other state-of-the-art air shower simulation codes are performed in chapter 6. Differences in their predictions are identified and possible explanations given. Especially the treatment of low-energy interactions and their imprint on muons at radial ranges in the km scale are highlighted with important potential implications on the muon deficit.

To showcase what is possible with the newly gained capabilities of inspecting the cascade history, we study the phase space of hadronic interactions with respect to their relevance for muon production in a quantitative way ("muon genealogy"), considering especially low-energy interactions for laterally separated muons (section 7.1).

Additionally, in section 7.2 we investigate two scenarios of modified hadronic interactions and their impact on air shower observables, whose implementation is greatly facilitated with CORSIKA 8. In the first scenario, we consider a modification of hadron-air cross-sections. Novelties compared to previous work are that pions, kaons and nucleons are treated independently and that the impact on the muon production depth is taken into account. In the second scenario, we examine the production of ρ^0 instead of π^0 mesons.

Looking also at exotic interaction scenarios, we investigate the phenomenology of hypothetical electroweak interactions with large multiplicities of Higgs bosons in the first interaction of UHECR air showers in chapter 8, before summarizing the main results obtained in this thesis in chapter 9.

2 Cosmic rays and extensive air showers

2.1 Cosmic rays

Cosmic rays (CR), atomic nuclei travelling at the speed of light, permeate the Universe. Their energies as measured in the reference frame of Earth, cover many orders of magnitude. Figure 2.1 shows the CR energy spectrum, multiplied by energy $E^{2.6}$ to pronounce spectral features. The spectrum can be well described by a broken power-law with an exponent varying around -3 up to $\sim 10^{20}$ eV, where it seems to be cut off. The steepness makes it increasingly laborious to measure the spectrum with high precision: The integral CR flux above the energy of 3×10^{15} eV, the so-called *knee*, is in the order of $1\,\mathrm{m}^{-2}\,\mathrm{yr}^{-1}$. Above the *ankle*, located at 5×10^{18} eV, the integral flux is already diminished to $1\,\mathrm{km}^{-2}\,\mathrm{yr}^{-1}$. Studying UHECRs, i.e. having energies $\geq 10^{18}\,\mathrm{eV} = 1\,\mathrm{EeV}$ [1], in a reasonable timeframe therefore requires detectors capable of monitoring huge areas in the order of $1000\,\mathrm{km}^2$.

Despite being studied for several decades with a lot of progress (see e.g. the historical review by Kampert and Watson [4]), the nature of UHECRs remains not well understood. Open questions pertain in particular their sources, generally thought to be extragalactic, and the mechanisms to accelerate particles to energies far beyond the reach of any artificial accelerator. Although a number of classes of astrophysical objects are considered as potential UHECR sources (see e.g. the recent review ref. [1]), no single object has been identified as such.

Moreover, it is not clear what causes the flux suppression at the highest energies. One possible explanation is simply that the acceleration mechanism in unable to accelerate cosmic rays further. Another explanation invokes effects during the propagation from the source to Earth that effectively slow down UHECR above a threshold energy. At $E = 5 \times 10^{19}$ eV, protons start scattering off cosmic microwave background photons (with cross-sections of a few $100\,\mu\mathrm{b}$) to produce a Δ^+ resonance, which subsequently decays into a nucleon and a pion:

$$p + \gamma_{\mathrm{CMB}} \rightarrow \Delta^+ \begin{array}{l} \nearrow\ p + \pi^0 \\ \searrow\ n + \pi^+. \end{array} \tag{2.1}$$

The proton loses about 20 % of its energy via this process. A rough estimate of the mean free path of this photohadronic process is 6 Mpc. Effectively, protons will undergo this type of interactions repeatedly until their energy drops below the threshold. This so-called *Greisen-Zatsepin-Kuzmin (GZK) effect* [22, 23] limits the distance an cosmic ray proton detected with above-threshold energy could have travelled to within 50 Mpc to 100 Mpc around Earth. For heavier nuclei, similar photo-disintegration effects exist.

Complementary to the search for UHECR sources is the determination of the *mass composition* (or *nuclear composition* [2]) of UHECR. Precise measurements provide strong constraints not only on acceleration models but also on propagation models [8]. Especially in energy ranges around spectral features, mass composition data can provide valuable insights on their origin. Measurements of mass composition at ultra-high energies, however, are among the most challenging, owing to the indirect nature of the detection technique and large systematic uncertainties associated to them. This is exemplified in fig. 2.2, which shows the composition fractions inferred from the X_{max} observable described further below. The results are strongly dependent on the choice of hadronic interaction models (here EPOS-LHC and SIBYLL 2.3) necessary to derive the composition fractions from actual data. To date, no high quality mass composition measurements exist at the highest energies [1].

2.2 Extensive air showers

Besides these *astrophysical* aspects, UHECR also provide opportunities to study *particle physics* via their interactions with the terrestrial atmosphere. For (lab-frame) energies E_{lab} above 10^{17} eV, the centre-of-mass energy per target nucleon, $\sqrt{s_N} \simeq (2m_p E_{\mathrm{lab}})^{1/2}$, exceeds the Large Hadron Collider (LHC), running at 14 TeV. At 10^{19} eV, $\sqrt{s_N}$ reaches a value of 140 TeV, which may be matched only by the proposed Future Circular Collider (FCC), which will not be operational for at least two more decades [25].

These vast amounts of energy are the driver of particle cascades, so-called *extensive air showers*, that start to develop after the *primary* cosmic ray has interacted inelastically with a nucleus of the atmosphere (consisting dominantly of 78 % nitrogen, 21 % oxygen, 0.9 % argon). In this hadronic collision a large number of high-energy secondary particles are created, collimated in the direction of the primary momentum. These typically interact again themselves with air nuclei after travelling macroscopically long distances. This process ends when particles either are sufficiently low energetic to be below the production threshold of new particles, or when their interaction cross-sections are negligibly small (muons, neutrinos).

The shower development is a highly complex process: Many particle species are involved; strong, weak and electromagnetic forces play a role; energy scales from MeV to EeV are equally relevant. Quantitative predictions usually require detailed simulations, which is the topic of the subsequent chapter. Nevertheless, the behaviour of extensive air showers and important observables can be understood at least qualitatively to a large degree, on which we focus here.

Let us begin the discussion with some preliminaries. EAS develop mainly longitudinally, i.e. in the direction of the momentum of the primary particle. The imaginary straight line that continues the trajectory of the primary particle in its direction of motion is called the *shower axis* and the point where it impinges on the ground is known as *shower core*. The angle between the shower axis and the vertical direction pointing upwards at the shower core is the *zenith angle* ϑ. The longitudinal development is measured along the shower axis,

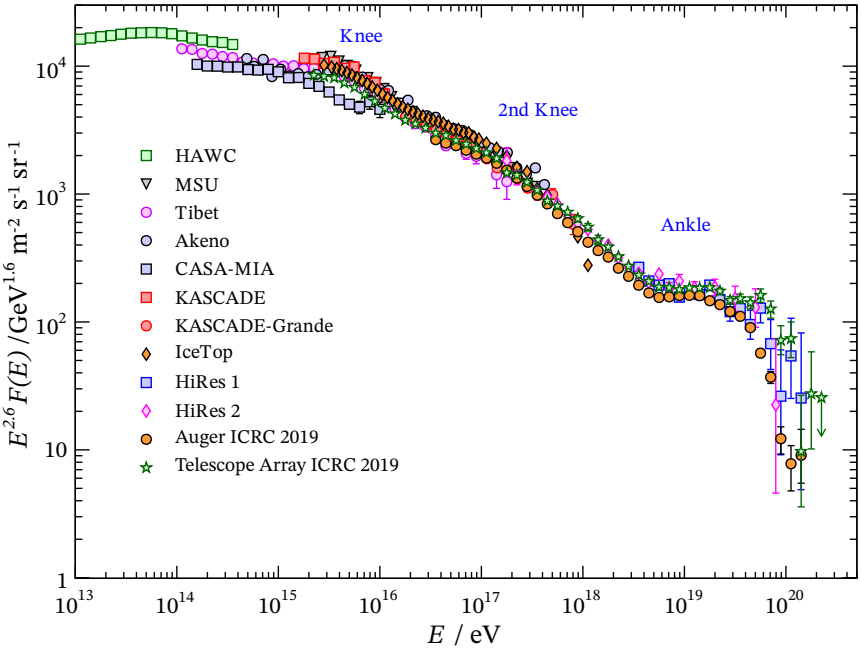


Figure 2.1: All-particle energy spectrum of cosmic rays; adapted from ref. [11]. Data from refs. [12–21]

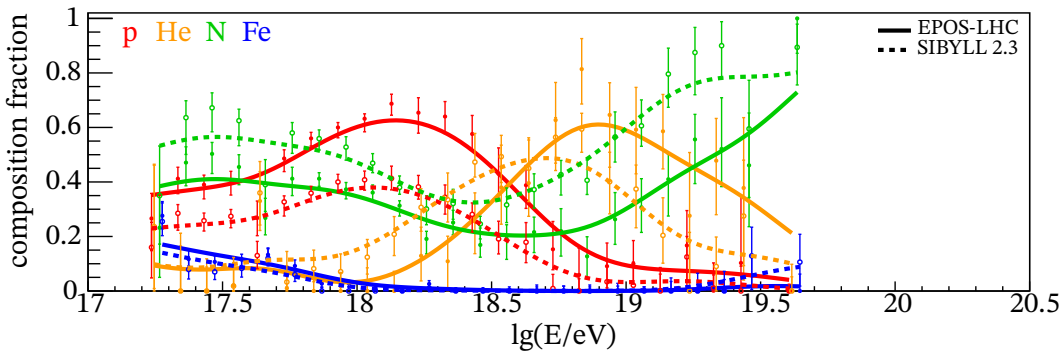


Figure 2.2: Composition fractions inferred from $X_{\max}$ distributions [24]. Error bars indicate statistical uncertainties. Adapted from ref. [1].

however, not in units of length but in *slant depth*

$$X = \int_{\text{shower axis}} \varrho \, \mathrm{d}s, \tag{2.2}$$

i.e., integrated mass density ϱ along the shower axis. Slant depth is the natural scale as it quantifies how many target atoms a particle travelling through matter passes, which is the relevant quantity for the mean free path between to interactions. A more detailed discussion will follow in chapter 4. For vertical showers, the total slant depth from the top of the atmosphere down to sea level, around 100 km in distance, is about $1000\,\mathrm{g\,cm^{-2}}$.

The lateral spread is much smaller than the longitudinal extent, albeit of great importance for ground-based EAS detectors.

The particles in a cosmic-ray induced EAS can be categorized in three groups (or components) which are closely linked to, but develop almost independently of each other: electromagnetic (EM) particles ($\gamma, e^{\pm}$), hadrons (in particular long-lived ones), and muons.

2.3 Electromagnetic cascade

EM processes in air showers can be calculated perturbatively in the framework of quantum electrodynamics (QED) with high precision. The EM component is therefore believed to be well understood and modelled. The characteristic scale of EM processes is the *radiation length* X_{rad}, having a value of $X_{\mathrm{rad}} = 37\,\mathrm{g\,cm^{-2}}$ in air.

2.3.1 Interaction processes

Photons passing through matter are mainly subject to three competing electromagnetic processes, the photoelectric effect, the Compton effect, and pair creation, of which the last is by far the most relevant at high energies. In this process, the incident photon gets converted to an $e^{\pm}$ pair. The cross-sections of pair creation were first calculated by Bethe and Heitler [26]: In the limit where the total screening approximation is valid, i.e. at energies $\gtrsim 40\,\mathrm{MeV}$ in air, the interaction length of pair production $\lambda_{\mathrm{int}}^{(\mathrm{pair})} = 9/7 \times X_{\mathrm{rad}}$ is energy-independent.

Electrons and positrons are subject to energy losses, $\mathrm{d}E/\mathrm{d}X$, due to ionization, which we treat as continuous process, as well as bremsstrahlung. Ionization losses rise only logarithmically with the electron energy, while losses due to bremsstrahlung increase almost linearly with energy. The point at which both contributions are equal, the *critical energy* E_{c}, has a value of $E_{\mathrm{c}} = 87\,\mathrm{MeV}$ in air. For $E \gg E_{\mathrm{c}}$, the energy of electrons decreases exponentially as $E \sim \exp(-E/X_{\mathrm{rad}})$. Since the bremsstrahlung cross-section, calculated, too, by Bethe and Heitler [26], is infrared-divergent, however, the number of photons emitted is infinite, and their energy distribution of the soft photons behaves like E_γ^{-1}, where E_γ is the energy of the radiated photon. In Monte Carlo simulations this necessitates introducing a cut-off energy. Only photons above that energy are treated explicitly, while the losses by soft photons below the cut are added to the continuous losses.

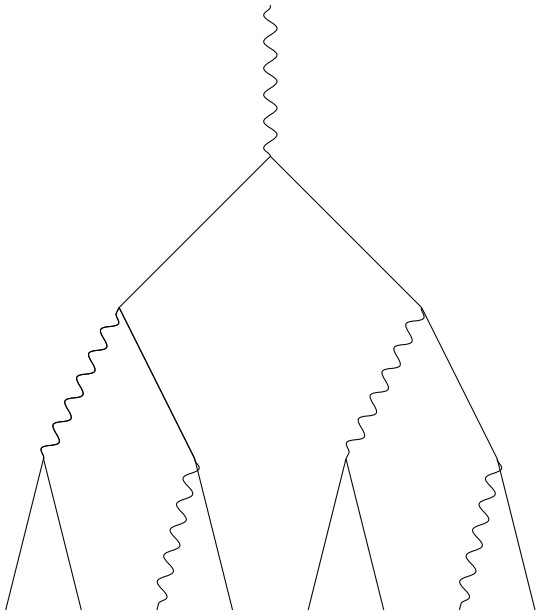


Figure 2.3: Heitler model with a primary photon

2.3.2 Heitler model

Because of the very similar length scales of pair production and bremsstrahlung that differ only by a factor of 7/9, electrons (including positrons) and photons well above the critical energy behave almost symmetric. Heitler [27] introduced a toy model of EM showers which can qualitatively describe certain features. In this model, particles travel a fixed depth $d = \log(2)X_{\text{rad}}$, after which they are split into two; photons create a $e^{\pm}$ pair and electrons radiate off a photon. Each split particle carries half of the incident particle's energy. This branching process continues down to the critical energy, where the particles are considered to be absorbed and the cascade stops. Figure 2.3 illustrates this process. After each splitting, the total number of particles doubles, leading to an exponential growth. The maximum number of particles, $N_{\text{max}} = E_0/E_{\text{c}}$, is proportional to the primary energy E_0. The depth at which the maximum occurs,

$$X_{\text{max}} = X_{\text{rad}} \log\left(\frac{E_0}{E_{\text{c}}}\right), \tag{2.3}$$

depends logarithmically on E_0 and the *elongation rate* $\mathrm{d}X_{\text{max}}/\mathrm{d}\log_{10} E_0$, is constant. Both behaviours are found also in detailed simulations [28] when considering average values. Furthermore, an interesting observation is that the number of each particle species quickly equilibrates after a few steps (or *generations*) irrespective of the primary particle.

2.3.3 LPM effect and dielectric suppression

The results of the previous sections are based on cross-section calculations with a single target atom. As it turns out, this description becomes inaccurate under certain conditions [29]:

From the uncertainty principle it becomes evident that the interaction is not point-like but spread out over a certain region. The size of this *formation zone* is inversely proportional to the momentum transfer. If it is sufficiently large to contain multiple target atoms (which introduces a density dependence), this gives rise to interference effects. Landau and Pomeranchuk [30], Migdal [31] and Ter-Mikaelian [32] calculated these, now termed *LPM effect* and *dielectric suppression*, leading to a suppression of the Bethe-Heitler cross-sections. The cross-sections for bremsstrahlung and pair creation become dependent on energy as well as density. Moreover, the dielectric suppression removes the infrared-divergence, rendering a cutoff unnecessary [33].

Especially the LPM effect cannot be neglected in EM showers at ultra-high energies as it leads to significantly shifted shower maxima [34]. The elongation rate is found to increase significantly under these conditions [28].

2.4 Hadronic interactions

Hadrons and their interactions air showers are among the most complex and least well understood aspects of air shower physics. Hadrons in air showers are usually not measured themselves but they serve as main source of both muons and EM particles.

The properties of strong interactions make it practically impossible to make solid predictions with ab initio calculations. Perturbative calculations in quantum chromodynamics (QCD), the established mathematical theory of strong interactions, are possible only in a regime where the coupling constant is small and only for interactions among partons (quarks and gluons). Due to the – still not well understood [35] – confinement, however, hadrons are always composite objects made of multiple partons. For interactions between hadrons, some phenomenological modelling is necessary to first go from hadron-level to parton-level, and later, after the scattering scattering of individual partons, go back to final-state hadrons. This final stage is called *hadronization*.

2.4.1 Hadronic interaction models

Therefore, hadronic interactions comprise several different aspects and energy (or length) scales. In practice, one relies on specialized codes, so called *hadronic interaction models* or *event generators*, which generate random lists of final-state particles for two initial-state hadrons. A number of such codes exist, some of which are focused on applications in collider physics, others more targeted at interactions in air showers. For the application in air shower simulations, these need to support interactions of the relevant hadrons (at least nucleons, pions, kaons) with light nuclei. Furthermore, they should provide reasonable predictions up to the highest energies, say $\sqrt{s} \lesssim 1000\,\text{TeV}$, which cannot be taken for granted in collider-oriented event generators.

State-of-the-art models frequently used in air shower physics are [7, 36]:

SIBYLL 2.3d [37, 38] is the simplest and therefore also the fastest (in terms of runtime required per event) of the high-energy models, developed specifically for air showers.

It is the only model in which the production of charmed particles is considered, which contribute significantly to the flux of very high energy muons and neutrinos.

QGSJet-II.04 [39] has the least number of free parameters of the models considered here, which reduces the extrapolation uncertainty to the highest energies. As a consequence, it features a less detailed description of the final-state hadrons, in which only nucleons, light mesons and $\Lambda/\bar{\Lambda}$ occur.

EPOS-LHC [40] is a model useful for both air shower applications and heavy ion physics. It aims at describing a plethora of data from different systems in great detail. This makes EPOS-LHC also the slowest of the models.

More details about the different physical concepts and approaches are given in the review by Engel et al. [7]. It is also worth to mention PYTHIA 8 [41], an event generator very popular in high energy physics. Only very recently features necessary for usage in air shower physics (most importantly, nuclear targets) have been implemented [42], but are not yet publicly released. It remains to be seen how well it performs in shower simulations since its "traditional" applications do not require extrapolations to the highest energies. With an earlier version, PYTHIA 6, air showers have been simulated in a proton atmosphere [43].

The traditional high-energy hadronic interaction models for air showers are not suited to describe hadronic interactions at low energies. SIBYLL, for instance, cannot be used at energies $\sqrt{s} < 10\,\mathrm{GeV}$ ($E_{\mathrm{lab}} < 50\,\mathrm{GeV}$. As a remedy, in air shower simulations they are complemented by specialized interaction models for this energy regime (typically down to kinetic energies of $\sim 100\,\mathrm{MeV}$), which themselves do not aim at describing high-energy interactions accurately. Common low-energy models include:

FLUKA [44, 45] contains an interaction model that treats nuclear effects in great detail.

UrQMD [46, 47] is a microscopic model used mainly in heavy ion physics. The (not up-to-date) version 1.3 is in use in several air shower simulation codes.

HSA [48] The *Hillas splitting algorithm* is a simple toy model of hadronic interactions that approximately reproduces their main features. In the HSA, the available energy is split randomly and recursively into many pieces. It does not, however, incorporate transverse momentum and does not predict cross-sections. An extended version is the default low-energy model in the air shower simulation code AIRES.

2.4.2 Kinematic variables

The kinematical phase-space of particles produced in (hadronic) interactions can be parameterized in several ways, that shall be briefly introduced here. For more pedagogical treatments, see e.g. refs. [49, 50]. The *rapidity* y is a replacement for velocities in special relativity. For a particle of energy E, mass m and momentum $\vec{p}$, split into components $p_{\parallel}$ parallel and $p_{\perp}$ transversal to the beam axis, it is defined as

$$y = \frac{1}{2}\log\left(\frac{E + p_{\parallel}}{E - p_{\parallel}}\right) = \operatorname{atanh}\left(\frac{p_{\parallel}}{E}\right). \tag{2.4}$$

In contrast to velocity, rapidity is an additive quantity under Lorentz transforms. If E and p are defined in the centre-of-mass frame of the collision, the kinematically possible range covered by (massive) secondaries, $[-y_{\max}, +y_{\max}]$, is given by

$$y_{\max} \simeq \log\left(\frac{\sqrt{s}}{m_{\mathrm{b}}}\right) + \log\left(\frac{m_{\mathrm{b}}}{m}\right), \tag{2.5}$$

where m_{b} is the mass of the beam particle.

Measurements of rapidity require independent determinations of energy and momentum or, equivalent, particle identification, which is often not possible. An experimentally much easier to handle quantity is *pseudorapidity*

$$\eta = \frac{1}{2}\log\left(\frac{p + p_\perp}{p - p_\perp}\right) = \operatorname{atanh}\left(\frac{p_\parallel}{p}\right) = -\log\tan\frac{\theta}{2}. \tag{2.6}$$

The last identity shows that a determination of η only requires to know the scattering angle θ with respect to the beam axis. It can in principle reach arbitrarily high values.

Another variable, *Feynman-x* x_{F}, is a measure of the longitudinal momentum relative to its maximum possible value $p_{\max} \simeq \sqrt{s}/2$, both given in the centre-of-mass frame:

$$x_{\mathrm{F}} = \frac{p_\parallel}{p_{\max}}. \tag{2.7}$$

It was introduced by Feynman [51], hypothesizing that at very high energies the inclusive cross-section of a particle

$$\frac{\mathrm{d}^3\sigma}{\mathrm{d}x_{\mathrm{F}}\mathrm{d}^2p_\perp} \equiv F(s, x_{\mathrm{F}}, p_\perp) \xrightarrow{s\to\infty} F(x_{\mathrm{F}}, p_\perp) = f(x_{\mathrm{F}})g(p_\perp) \tag{2.8}$$

becomes energy-independent. A related quantity in the lab frame is

$$x_{\mathrm{L}} = \frac{E_{\mathrm{lab}}}{E_{\mathrm{lab}}^{(\mathrm{proj})}} \simeq x_{\mathrm{F}} - \frac{m^2 + p_\perp^2}{2mE}, \tag{2.9}$$

where $E_{\mathrm{lab}}^{(\mathrm{proj})}$ is the lab-frame energy of the projectile.

2.5 Muon production

2.5.1 Heitler-Matthews model

The basic phenomenology of hadrons and muons and important scaling relations can be derived in an extension of the Heitler model to hadron cascades, the Heitler-Matthews model [52]. In this model, hadronic interactions produce m equal-energy pions, out of which a fraction R is taken to be long-lived (having a decay length of $c\tau = 7.8\,\mathrm{m}$) charged pions. The remaining ones are π^0, which decay quasi immediately ($c\tau = 26\,\mathrm{nm}$) into two photons and feed the electromagnetic cascade. If we consider only the π^0 of the first interaction (which is

a good approximation for large values of m, say > 10) happening at a slant depth of X_0, we can derive

$$X_{\max}^{(p)} = X_0 + X_{\text{rad}} \log\left(\frac{E_0}{2mE_{\text{c}}}\right) \tag{2.10}$$

for proton primaries. Comparing this with eq. (2.3), we find that the elongation rate of proton showers is smaller than that of EM showers.

The charged pions re-interact after travelling one interaction length λ_{int}, which is assumed to be constant, splitting the energy further, until they reach the critical energy $E_{\text{c}}^{(\pi)} \sim 100\,\text{GeV}$. It is defined to be the energy at which a decay becomes as probable as an interaction. At that point, which happens after

$$N_{\text{gen}} = \frac{\log\left(E_0/E_{\text{c}}^{(\pi)}\right)}{\log m} \tag{2.11}$$

generations, the pions are assumed to decay into muons. The total number of muons $N_\mu^{(p)}$ increases as

$$N_\mu^{(p)} = \left(\frac{E_0}{E_{\text{c}}^{(\pi)}}\right)^\beta, \tag{2.12}$$

where $\beta = \log(Rm)/\log(m) \simeq 0.9$ for realistic values of R and m. The sub-linear growth of $N_\mu^{(p)}$ with E_0 stems from the energy fraction lost to the EM cascade in each generation.

For heavier nuclei as primary particles these formulas can be extended if the first interaction is treated within the frame of the superposition model. Here, the interaction of a nucleus with A nucleons is replaced with A interactions of protons, which create A independent proton-induced showers of energy E_0/A. Taking these assumptions into account, the new relations read

$$X_{\max}^{(A)}(E_0) = X_{\max}^{(p)}(E_0/A) = X_{\max}^{(p)}(E_0) - X_{\text{rad}} \log(A), \tag{2.13}$$

$$N_\mu^{(A)}(E_0) = A N_\mu^{(p)}(E_0/A) = A^{1-\beta} N_\mu^{(p)}(E_0). \tag{2.14}$$

These two relations are the foundation of mass composition analyses based on $X_{\max}$ and N_μ measurements.

A more detailed discussion of the Heitler-Matthews model, its limitations and several improvements can be found in ref. [53].

2.5.2 Features of hadronic interactions and their relevance for air shower observables

Clearly, hadronic interactions in reality are not as simple as assumed in the Heitler-Matthews model and to obtain quantitatively meaningful predictions one has to rely on interaction models. Although all of them have been retuned to better describe LHC data taken in proton-proton (pp) collisions[1], which reduced differences e.g. of their pp cross-section predictions, many relevant aspects still remain rather unconstrained. The phase-space of interactions

[1] The retuned versions are often termed *post-LHC models*.

in air showers is much more diverse than what is (currently) possible to be measured at colliders. From the available data, hadronic interaction models need not only to extrapolate in energy but also to different projectile species (mainly pions and kaons) and nuclear targets. Additionally, as will be explicitly proven in chapter 7, at high energies especially particles produced in the difficult-to-measure forward direction have a high impact on muon production. A step forward in this direction will be the planned proton-oxygen run of the LHC [54, 55].

Ulrich et al. [9] investigate the impact of a number of "high-level" observables of hadronic interactions on air shower observables. An energy-dependent modification factor,

$$F(E) = \left(1 + \frac{f_{19} - 1}{\log \frac{E_{\mathrm{th}}}{10^{19}\,\mathrm{eV}}} \max\left(1, \log \frac{E}{E_{\mathrm{th}}}\right)\right), \tag{2.15}$$

that sets in at $E_{\mathrm{th}} = 10^{15}\,\mathrm{eV}$ is introduced with the parameter f_{19} steering how large the modification shall be at $10^{19}\,\mathrm{eV}$. The modification factor is used to change

- the inelastic cross-section (irrespective of the species),
- the multiplicity,
- the π^0 fraction ($1 - R$ in the Heitler-Matthews model),
- the elasticity, defined as fraction of energy of the most-energetic secondary (so-called *leading particle*) of the projectile energy.

Some results, summarized in fig. 2.4 for proton showers at $3 \times 10^{19}\,\mathrm{eV}$, show that X_{max} is very sensitive to the hadronic cross-section while leaving the number of muons almost unchanged. Conversely, the muon number changes most drastically with the π^0 fraction without influencing X_{max}.

2.5.3 Muon production depth

Besides the total number of muons it is also possible to measure the height at which they were created. Assuming they are produced close to the shower axis they travel essentially in straight lines with the speed of light. Depending on whether the production happens early (high altitude) or late (low altitude) in the shower development, they arrive earlier or later on ground at a certain distance from the shower core [56]. The distribution obtained this way is the so-called *apparent muon production depth* (AMPD). Since muons may decay in flight, which is further enhanced by energy losses, it is not equal to the real MPD. The total number of muons on ground is equal to the integrated AMPD along X. This is illustrated in fig. 2.5. Here, the average AMPD of vertical proton-induced showers (simulated with CORSIKA 8 with a fixed height of first interaction $h_0 = 21.6\,\mathrm{km}$ and an observation level at $875\,\mathrm{g\,cm^{-2}}$) is shown. Besides the standard scenario (solid line), the two artificial cases of stable muons (dashed line) and stable muons without energy losses (dash-dotted line) are depicted, considering three cutoff energies. Only in the case of stable muons without energy losses, the AMPD is equal to the real MPD. The greatest impact is seen for muons

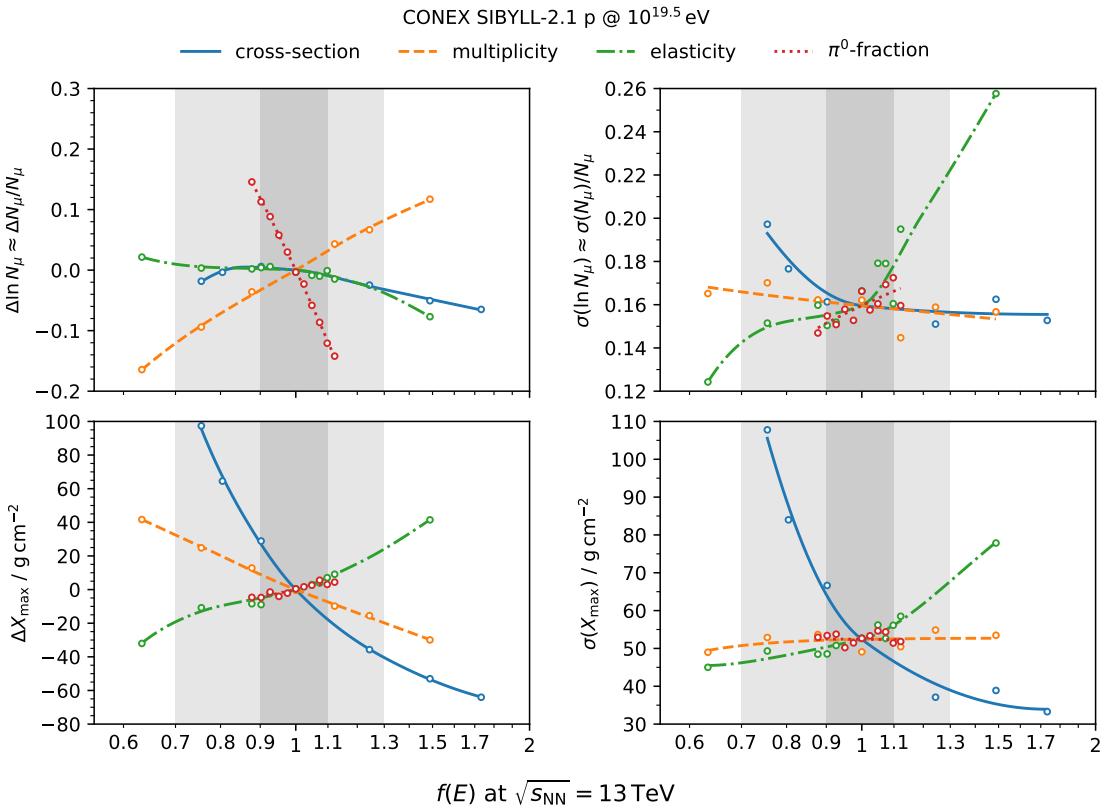


Figure 2.4: Impact of changes of hadronic interaction features on air shower observables; taken from ref. [53], data from ref. [9]

above 1 GeV. In this case, disabling energy losses together with decay has a considerably larger impact than only disabling muon decays. For muons above 10 GeV both effects are of similar magnitude but have a smaller impact combined since decays are suppressed by the higher energy. For $E_\mu \geq 100$ GeV the effect is even further suppressed. It is also interesting to note that the solid and dashed lines converge towards the observation level. The shorter the distance to ground, the smaller is the probability of decaying or being absorbed in flight. The dash-dotted line lies a bit higher than the others even at ground since energy losses also affect other particles, in particular the low-energy pions which are the parents of the muons.

Like $X_{\max}$ and N_μ, the maximum $X_{\max}^{(\mu)}$ is a composition-sensitive observable. A more detailed account of the AMPD and its dependencies is given in ref. [57].

2.5.4 The muon deficit

With $X_{\max}$, N_μ and $X_{\max}^{(\mu)}$ several observables are at hand that are composition-sensitive. At the same time, however, the interpretation of such data relies on predictions made with hadronic interaction models. Ideally, reconstructing the mass composition from any of these observables should yield comparable results, which allows testing the consistency of the models. Moreover, even if only a single of these observables is considered, tensions arise when reconstructions yield values outside the region bounded by proton and iron ($A = 56$) primaries.

Since the year 2000, evidence has been growing that muon-based reconstructions suggest a heavier (larger values of $\log A$) composition than those based on $X_{\max}$. After the first report of a discrepancy by the HiRES and MIA collaborations [58], their findings were confirmed by the Pierre Auger Observatory in a nearly model-independent way [59]. Even though the post-LHC models have larger muon production compared to pre-LHC versions, a *muon deficit* still remains in simulations. The Working Group on Hadronic Interactions and Shower Physics founded by eight experimental collaborations has conducted a meta-analysis of their partially conflicting results [60–62]. As each experiment has different means of counting muons and different energy thresholds, the *z scale* is introduced to allow combining these measurements into a unified frame. Based on some quantity $\hat{N}_\mu$ assumed to be proportional to the total number of muons, it is defined as

$$z = \frac{\log\left(\hat{N}_\mu^{(\text{meas.})}/\hat{N}_\mu^{(\text{p})}\right)}{\log\left(\hat{N}_\mu^{(\text{Fe})}/\hat{N}_\mu^{(\text{p})}\right)}, \tag{2.16}$$

where $\hat{N}_\mu^{(\text{meas.})}$ is the measured value and $\hat{N}_\mu^{(\text{p/Fe})}$ are the corresponding values from simulations (including detector simulation). Figure 2.6 shows the compiled data interpreted with EPOS-LHC simulations. Here, the value z_{mass} has been subtracted, which is the z value of a muon content as expected from a composition according to the *global spline fit* (GSF) model. With a consistent hadronic interaction model, data are expected to lie in the grey band expected from $X_{\max}$ measurements. The GSF composition model is derived mainly from $X_{\max}$ data itself, so that the z value expected from $X_{\max}$ is close to z_{mass}. The black line shows the result of a fit, indicating an excess in data (or deficit in simulations) increasing

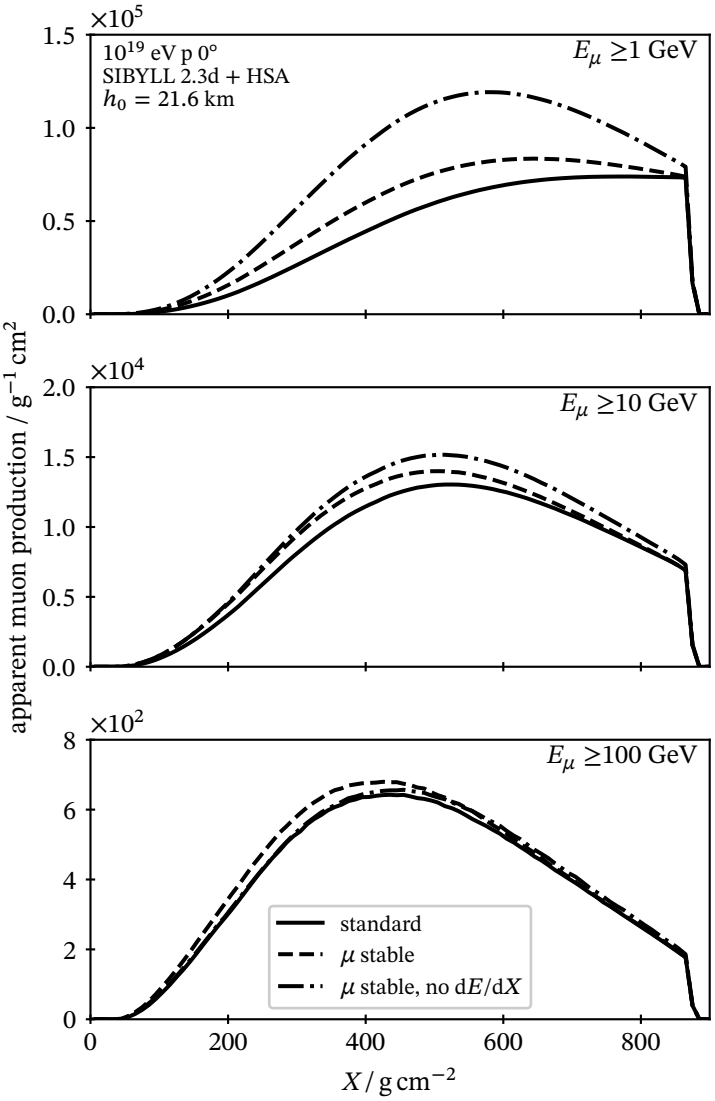


Figure 2.5: Effect of decay and energy losses on apparent muon production

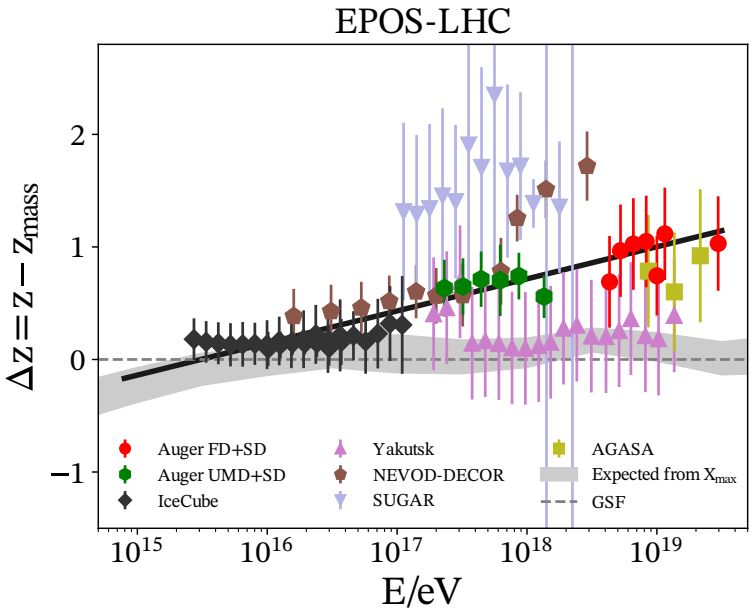


Figure 2.6: Compilation of z scale measurements. Adapted from ref. [62]; data from refs. [10, 59, 63–70]

logarithmically with energy. The slope of the line is non-zero with a significance of 8σ in the case of EPOS-LHC and above 10σ with QGSJet-II.04.

On the other hand, a recent measurement of the fluctuations of N_μ at the Pierre Auger Observatory shows good agreement between simulations and data [67], which puts strong constraints on (modified) hadronic interaction models at ultra-high energies. Considering that this observable is dominated by the first few interactions in the shower, one can reason that the cause of the muon deficit is likely the effect of accumulating slight deviations over several generations, covering many orders of magnitude in energy, rather than severely mismodelled or new physics at the highest energies. Due to the cascade effect, even small effects get amplified by the number of generations.

Besides the muon number on ground, also measurements of $X_{\mathrm{max}}^{(\mu)}$ are inconsistent with X_{max} [71]. For EPOS-LHC the incompatibility amounts to at least 2.5σ. Both $X_{\mathrm{max}}^{(\mu)}$ and N_μ suggest a heavier composition than X_{max}.

2.5.5 Possible solutions

As the muon deficit became more and more evident, efforts were put to figure out ways to increase the muon production. Important aspects are the production of baryon/anti-baryon production [72] and the role of forward ρ^0 production in pion charge exchange reactions, which will be further explained in section 7.2.2. Nevertheless, even in the up-to-date versions, the post-LHC models still produce too few muons. It has been suggested that the

only plausible way to reconcile the $X_{\max}/N_\mu$ inconsistency is to decrease the production of π^0 (and to a minor extent η) [53], which can be achieved with a modified hadronization model. In fact, recent LHC measurements of pp collisions indicate the onset of formation of a quark-gluon plasma, in which a statistical hadronization takes places instead of the conventional string fragmentation (see ref. [73] and references therein). Among other things, enhanced strangeness production is expected (i.e. more kaons) under these conditions, which in turn suppresses π^0. The *core-corona* effect and its incorporation into hadronic models is a promising possible solutions but still requires more accelerator measurements such as with the upcoming proton-oxygen run [53].

2.6 Principles of EAS detection

The fact that UHECR initiate EAS extending over many kilometres in size makes it possible to detect them with a comparatively low amount of instrumentation. Compared to direct detection techniques which become inefficient around $\sim 10^{14}$ eV due to the tiny CR flux, this indirect measurement exploits the atmosphere as a calorimeter. Today, three techniques are commonly employed to detect signals of EAS.

2.6.1 Arrays of particle detectors

By placing particle detectors (e.g. scintillators or Cherenkov detectors) on ground, one can detect the shower footprint when it reaches ground. Because of the lateral spread it is for many purposes sufficient to cover an area only sparsely with such detectors. The largest array to date is the surface detector (SD) of the Pierre Auger Observatory in operation near the town of Malargüe, Argentina, consisting of about 1600 SD stations placed on a a grid with 1.5 km spacing [74].

Especially useful are detectors which can discriminate well between EM particles and muons, so that muon-based observables can be exploited. An example is the underground muon detector of the Pierre Auger Observatory [75] consisting of scintillators buried 2.3 m under ground next to SD stations in a subset of the array of 30 km^2 size. These detect only muons with an energy threshold of $\sim$ 1 GeV, while the water-Cherenkov detector above ground is sensitive also to electromagnetic particles.

2.6.2 Fluorescence telescopes

A fraction of the energy deposited into the air by ionization is re-emitted as fluorescence light isotropically. This ultra-violet light can be seen with fluorescence telescopes, yielding a time-resolved image of the shower. Properly calibrated, the signal can be used not only to measure $X_{\max}$ but also the total energy deposit of the shower in the atmosphere, which provides a viable method to reconstruct the shower energy insensitive to details of hadronic interactions.

Fluorescence telescopes can be ground-based, as e.g. in the Pierre Auger Observatory [76], where they overlook the atmosphere above the surface array, or space-based to cover huge

areas. The main downside is the limited uptime of $\sim 15\,\%$ because they need favourable conditions (dark nights, no clouds).

2.6.3 Radio antennas

Since a considerable fraction of particles in a shower are electrons and positrons, their creation, acceleration and absorption results in time-varying currents which emit classical EM radiation that happens to be in the MHz to GHz frequency range [77]. The dominant contribution stems from synchrotron-like emission due to transversal acceleration in the geomagnetic field, but to a minor extent also excess of electrons over positron is relevant. Radio emission can be measured with standard radio antennas, whose signal traces are read out and digitized.

Together with the underlying modelling of the EM cascade, radio emission is widely considered to be almost free of systematic uncertainties. Additionally, the radio waves are barely attenuated, which makes the technique attractive to measure horizontal showers, whose particle components usually do not reach ground.

3 Simulations of air showers

3.1 Cascade equations

The *cascade equations* (CE) in their canonical form describe the average longitudinal development of a particle cascade, neglecting any lateral spread. The variables of interest are the energy spectra $n^{(i)}(E,X) = \mathrm{d}N^{(i)}/\mathrm{d}E$ of all relevant particle species i. They take the form [78]

$$\frac{\partial n^{(i)}}{\partial X}(E,X) = -\left(\frac{1}{\lambda_{\mathrm{int}}^{(i)}(E)} + \frac{1}{d^{(i)}(E)}\right) n^{(i)}(E,X) \tag{3.1a}$$

$$+ \sum_j \int_E^\infty \left(\frac{Y_{\mathrm{int}}^{(j\to i)}(E',E)}{\lambda_{\mathrm{int}}^{(j)}(E')} + \frac{Y_{\mathrm{dec}}^{(j\to i)}(E',E)}{d^{(j)}(E')}\right) n^{(j)}(E',X)\,\mathrm{d}E' \tag{3.1b}$$

$$+ \frac{\partial}{\partial E}\,\beta_{\mathrm{ion}}^{(i)}(E) n^{(i)}(E,X). \tag{3.1c}$$

The first line describes a decrease of the particle number due to interactions (first term) and decays (second term), which produces new secondaries, possibly of another species, with reduced energy. Here, $d^{(i)}(E) = \rho c \tau_0^{(i)} E/m^{(i)}$ is the decay length of a particle with lifetime $\tau_0^{(i)}$ at rest and mass $m^{(i)}$. Since the CE are formulated in X as independent variable, $d^{(i)}(E)$ is expressed as grammage.

Correspondingly, the second line describes a gain stemming from interactions and decays of particles at higher energies. The yield functions $Y_{\mathrm{dec/int}}^{(j\to i)}(E',E)$ describe how many particles of species i and energy E are produced on average in a decay or interaction of a j particle of energy E'. They are derived from event generators and therefore render the solutions of the CE dependent on the chosen hadronic interaction model. The interaction yield $Y_{\mathrm{int}}^{(j\to i)}$ also depends on the target atom (in case of a mixture, an average over all components is assumed). The lower limit of the integral expresses energy conservation in the meaning that new particles of a certain energy can only be created from particles with higher energy, neglecting a small contribution from the target. The upper limit can be replaced by the primary energy.

The third line describes the influence of ionization energy losses, which can act both as loss and gain term. Here, $\beta^{(i)} = \mathrm{d}E/\mathrm{d}X$ is the continuous loss of a single particle. This last term is relevant virtually only for low-energy particles ($\lesssim 10\,\mathrm{GeV}$).

The boundary conditions for an air shower with a single primary particle of energy E_0 and species P entering the top of the atmosphere ($X = 0$) read

$$n_0^{(i)}(E, X=0) = \delta_{i,P}\delta(E - E_0). \tag{3.2}$$

The CE can also be used for inclusive fluxes. In that case the boundary conditions resemble the primary cosmic-ray flux, which also results in a change of units. An extensive overview over analytic solutions for both conditions is given by Gaisser et al. [3]. The CE for air shower conditions can be solved purely analytically only in few scenarios, most notably electron-photon cascades. It is more practical to solve them numerically, which is done by discretizing the energy domain into bins (typically with a logarithmic binning). The integrals in eq. (3.1b) then turn into sums and one is left with a system of ordinary differential equations in X for the number of particles in each energy bin, which can then readily be integrated with standard methods with little computational effort.

The main downside of a purely CE-based shower simulation is, however, that it by construction cannot yield any insight about distributions of observables since describes only an average shower. Detailed simulations that resemble a real shower with all its (random) features are possible only with the Monte Carlo method.

3.2 Monte Carlo simulations

In the broad sense, the term *Monte Carlo* refers to a technique to calculate integrals by interpreting them as expectation value [79]. Considering a probability density $p(x)$ on some domain Ω, the expectation value of a function $f(x)$ is

$$E_p[f] = \int_\Omega f(x)p(x)\,\mathrm{d}x. \tag{3.3}$$

If we have access to a sequence of (independent) samples $\{x_i\}$ distributed according to $p(x)$, the empirical value

$$\hat{E}_p[f] = \frac{1}{N}\sum_{i=1}^{N} f(x_i) \tag{3.4}$$

is an estimate of $E_p[f]$ and will converge to the true value with increasing number of samples N by the strong law of large numbers. The error incurred by using a finite N, which can be estimated from the sample variance, decreases as $\mathcal{O}(N^{-1/2})$. In practical terms this means a reduction of the statistical fluctuations of $\hat{E}_p[f]$ by a factor of 10 requires a 100-fold greater sample size. Nevertheless, for high dimensional domains, Monte Carlo often outperforms numerical integration methods based on quadrature formulas. For a given accuracy, these require a number of evaluations growing exponentially with the dimension of Ω – a feature known as *curse of dimensionality*. The $N^{-1/2}$-law for the Monte Carlo method, however, is universal and valid in any dimensions.

Moreover, the Monte Carlo method itself does not need any closed-form expression for $p(x)$ or evaluations thereof, as long as some way of sampling from that distribution is available. In the context of particle cascades this is of particular relevance. While it is manageable to handle the distributions of single particles (how far they propagate, what secondaries they produce, etc.), describing a probability distribution of all possible microscopic configurations of the particles in a shower is prohibitively difficult (see also the discussion by Metropolis and Ulam [80]).

Instead, Monte Carlo shower simulations try to mimic real showers as close as possible by treating each particle individually, taking into account both deterministic and probabilistic aspects. Whenever randomness comes into play, e.g. when determining the outcome of a decay, it is sampled from the single-particle distributions just like in reality one possible outcome out of many is realized. This way, one ends up with a sample shower out of the distribution of all possible showers and one may evaluate any observable of interest from it.

More concretely, what happens in a Monte Carlo shower simulation is the following: Given the initial conditions (species, four-momentum, etc.), the primary particle is propagated (i.e., its equations of motion are solved, usually numerically) up to the randomly determined point where a stochastic event, i.e. an interaction with the medium or a decay, happens. At this point, the primary particle ceases to exist and secondary particles are created. They are sampled from the distributions given by models of such events. These secondaries are themselves propagated, each one independently, and events are generated for the sub-showers they initiate. This procedure is repeated until all particles of the shower have been processed. For this to happen, some additional criteria are introduced which may terminate the propagation. These terminating conditions include:

energy cut It makes sense to track particles only if they contribute in a non-negligible way to observables one is interested in. For example, particles that have energies below the detection threshold of a particular detector do not need to be simulated.

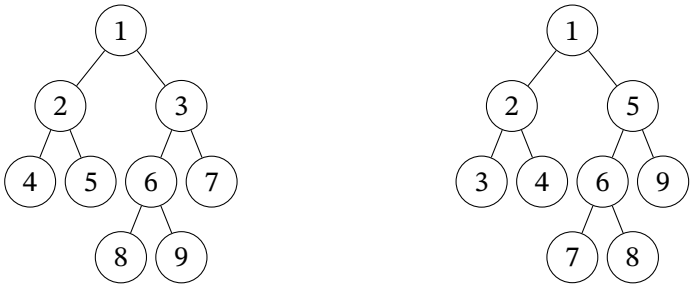
observation level In many applications one is interested in the footprint of the shower on ground, where detectors are located. In the simulations, a planar or curved surface is taken into account. Particles crossing this *observation level* are recorded into a file on disk and discarded.

time cut Stray particles lagging significantly behind the shower front are discarded.

geometric cut Particles that leave the volume of interest are discarded.

3.2.1 Shower traversal

How can a shower, containing particles in a number that can easily reach values of 10^{10} and more, be simulated on a computer? Clearly, it is hardly possible to keep the complete particle content with a 7-dim. state (position, time, momentum) associated with each particle in memory: Even with representation of the particle state with single-precision floating point numbers, amounting to $\sim 30\,\mathrm{B}$ per particle, the total memory required exceeds 100 GB. The solution is based on the insight that, from the point of view of an "algorithmist" [81], a particle shower is essentially a *tree*: Particle trajectories form the edges between nodes, interactions and decays, at which splittings happen. Each node has one incoming edge (the projectile) and several outgoing edges (the secondaries). Owing to this correspondence, we can think of a Monte Carlo shower simulation as traversing a tree [82]. Because of the randomness involved in the process, the tree is not known in advance and only unfolds itself while it is being traversed, with the outcome dependent on the order of traversal.



(a) Breadth-first (b) Depth-first

Figure 3.1: Fundamental tree traversal orders. The numbers indicate in which order the nodes are visited.

Trees can be traversed in various ways, the most common of which are *breadth-first* and *depth-first*, referring to the order in which the nodes are visited (see fig. 3.1). A breadth-first algorithm visits all nodes at the current depth before advancing to the next level down. In an air shower simulation this means that all particles of a certain generation are processed, followed by the particles of the subsequent generation. In a depth-first traversal, all child nodes of a given node are visited before continuing with its siblings. A sketch of a recursive implementation of a depth-first shower generation in pseudo-code is given in Algorithm 3.1. The algorithm would be called with a set data type as argument containing only the primary particle. Most real-world shower simulation codes do not implement this recursive algorithm directly. Instead, they reformulate the algorithm as an iterative one, which requires introducing an auxiliary data structure to store particles waiting to be propagated. For a depth-first traversal, it must operate "last in, first out", which is the defining property of a *stack*. The iterative version is outlined in algorithm 3.2. If one were to replace the stack with a queue, which operates "first in, first out", the algorithm would turn into a breadth-first traversal.

Let us now consider the memory footprint of both approaches in a simple Heitler-type toy model with multiplicity m. In the breadth-first approach, the number of particles in the queue grows exponentially with the generation number currently being processed. In the last stage just before the critical energy is reached, the number of particles in memory reaches

Algorithm 3.1 Recursive depth-first shower generation

```
1: procedure RecDFShower(set of particles S)
2:     for all p_i in S do
3:         propagate p_i
4:         if p_i survives cuts then
5:             T ← GenerateEvent(p_i)          ▷ store secondaries in set T
6:             call RecDFShower(T)
7:         end if
8:     end for
9: end procedure
```

Algorithm 3.2 Iterative depth-first shower generation with a stack

```
1:  procedure ITERDFSHOWER(primary particle p_0)
2:      push p_0 to stack                          ▷ stack contains only primary now
3:      while stack not empty do
4:          p ← pop stack                          ▷ take top particle from stack
5:          propagate p
6:          if p survives cuts then
7:              for all s in GENERATEEVENT(p) do   ▷ push secondaries onto stack
8:                  push s to stack
9:              end for
10:         end if
11:     end while
12: end procedure
```

$N = E_0/E_c$.

Going depth-first, the highest memory consumption in the toy model happens when the first particle reaches the critical energy. At that point, $N = (m-1)\log_m(E_0/E_c)$ particles reside in memory. This $\mathcal{O}(\log E_0)$-behaviour also holds on average in real simulations. It shows that even at the highest energies, memory consumption is very small. It can still be reduced further by sorting the particles by energy and propagating the lowest-energy particle first.

3.2.2 Hybrid simulations

Another promising technique to reduce the runtime of shower simulations, especially for UHECR, are *hybrid simulations*, which aim to combine the advantages of detailed Monte Carlo simulations with the speed of CE-based solutions [78, 83]. At high energies close to the primary energy, a full Monte Carlo treatment is performed, which ensures that the largest contributions to shower-to-shower fluctuations are modelled correctly. Particles that drop below a certain threshold, are then filled in an (E, X) grid. After the high-energy part of the shower has been fully simulated, the CE are solved numerically on the grid with the low-energy particles acting as sources in the CE. The CE are solved down to a second threshold energy, below which particles are again subject to Monte Carlo treatment. This final Monte Carlo stage is introduced to properly account for the lateral spread of the shower in the energy range where boost factors are small enough that transverse momentum becomes important.

3.2.3 Runtime and optimizations

The most important parameter governing the runtime of a Monte Carlo shower simulation is the primary energy as it determines the total number of particles and interactions in the shower. The average runtime scales linearly with E_0 [84]. A weak dependence on the zenith angle can be also be observed. Vertical showers are cut off if the shower maximum is not

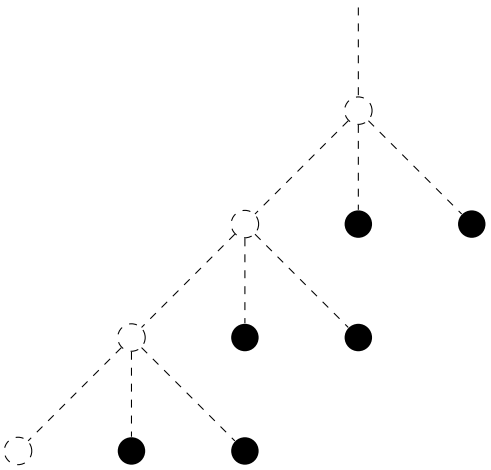


Figure 3.2: Snapshot of a depth-first traversal of a shower. Solid black particle nodes are still to be processed. Dashed particles have already been processed.

too far away from the ground. Inclined showers, however, can develop fully, leading to more particles and propagation steps required, thus increasing the runtime.

Although the generation of events by hadronic interaction models costs a multiple of time compared to electromagnetic interactions, their contribution to the total runtime is insignificant, owing to the vast number of electromagnetic particles, which outnumbers hadrons by a factor of 1000 [7]. The runtime of a vertical 10^{18} eV proton shower, simulated with CORSIKA 7 and QGSJet-II.04 amounts to approximately a week with electromagnetic cut-off energies set to 3 MeV. Turning electromagnetic interactions off, the runtime reduces to a few hours.

The fact that a shower can be split into individual, independent sub-showers (which themselves can be split again) makes it conceptually simple to devise a parallelization scheme: A particle initiating a sub-shower can be sent to a different CPU, possibly even on a different node, and processed there. Virtually no access to shared resources is required, which leads to almost perfectly scaling linear speedup. Subtleties arise when the shower shall be reproducible, especially with a different number of CPUs. Careful bookkeeping of the random-number streams and their seeds is required for this purpose.

Another, orthogonal technique for boosting the performance is vectorization. Modern CPU architectures allow applying the same operation on a number of values (typically a small power of 2, e.g. 8) at the same time if a suitable memory layout is chosen. This is also referred to as ”single instruction, multiple data” (SIMD). In the context of Monte Carlo shower simulations, this feature can be exploited especially if a certain operation can be applied to multiple particles at the same time. Attempts to rewrite algorithms in shower simulations to make use of vectorization have been applied to EGS4 [82] and, more recently, to Geant [85].

3.2.4 Thin sampling

Parallelization is a viable way of reducing the runtime of a single shower by distribution the load onto multiple CPUs, but does not reduce the total resources required. Most often, one is interested in generating a shower library with many, say several hundreds or thousands, showers. They can be generated independently of each other on different CPU. Here, parallelization already happens at a higher level (library-level instead of shower-level parallelism), so that further parallelization does not offer any advantage.

For situations like this, where good statistics are needed, a technique called *thin sampling* (or simply *thinning*) has been proposed by Hillas [48], which belongs to the more general category of *variance reduction* techniques. The idea is not to follow all particles individually but only a small (random) subset of them. The discarded particles are accounted for by assigning weight factors to the actually followed ones.

Above a certain energy threshold $E_{\mathrm{th}} = E_0$, usually expressed as a fraction $\varepsilon = E_{\mathrm{th}}/E_0$ of the primary energy, the simulation works as usual. Typical values are $\varepsilon \leq 10^{-6}$. Below the threshold, each time secondaries are created in a stochastic event, a random selection takes place, in which a secondary i survives only with a probability p_i. Depending on the details of the implementation, one may either select a single particle out of all secondaries, or decide for each secondary independently whether it is kept or not [86]. The weight of the surviving particle(s) is $w_i = w_{\mathrm{p}}/p_i$, where w_{p} is the weight of the parent particle ($w_{\mathrm{p}} = 1$ if the parent had not been subject to thinning yet). This definition ensures that expectation values stay unaffected by the thinning procedure. The probabilities should be defined in a way that predominantly those particles survive which are "important" for the shower development and contribute most to observables. For this reason, the standard prescription is to assign probabilities $p_i = E_i/E_{\mathrm{p}}$ based on the secondary energy E_i and parent energy E_{p}.

A downside of thinning is that only mean values stay unbiased but artificial fluctuations are introduced by the additional randomness. These can be minimized by modifying the algorithm in a way that the distribution of weights is narrow. A way to achieve that is to introduce a maximum weight w_{max} above which thinning is not applied anymore [86]. Additionally one may introduce separate ε and w_{max} parameters for hadronic and electromagnetic interactions.

3.3 Historical overview and current status

The first mention of the Monte Carlo method applied to particle cascades is in the (1949) seminal paper by Metropolis and Ulam [80], where the fundamental idea is proposed. A few years later, Wilson [87] describes the first implementation, carried out not with a computer but an analogue "Monte Carlo machine", a rotating cylinder with curves drawn onto it. The rotating cylinder is randomly stopped, triggered by a Geiger counter, and the fate of a particle is read off from the curve at the current position. Obviously, this manual procedure taking at least several seconds per step is very laborious.

A number of programs with relevance for cosmic rays have been developed over time. Pure Monte Carlo simulations include:

MOCCA is a program developed by Hillas [48], written in Pascal. MOCCA contained the original implementations of thinning and the HSA and was in use since the 1980s.

AIRES Originally a reimplementation of MOCCA in FORTRAN by Sciutto [88], AIRES was first released in 1997 [89] and is still being updated, with the latest version released in 2021. With the TIERRAS extension [90] it is possible to simulate showers not only in the atmosphere but also their development below ground. The ZHAires extension [91] allows to simulate the radio emission of particles.

Cosmos is another FORTRAN code, whose development started in the 1970s by Kasahara [92, 93]. It is mainly used in the Japanese cosmic ray community and still being maintained. Recent developments include combining Cosmos with EPICS to allow propagation in further media [94].

ASICO was developed during the 1970s and 1980s by Grieder [95] in FORTRAN. It was then handed over to the Karlsruhe group where it became the foundation of CORSIKA.

EGS is a FORTRAN code for pure electromagnetic showers. EGS4 [96] is included for electromagnetic interactions in CORSIKA with a number of modifications. More recent versions of EGS are maintained by two different groups under the names EGSnrc [97] and EGS5 [98].

CORSIKA The development of CORSIKA [5] began in the late 1980s [99, 100], initially just for usage in the KASCADE experiment in Karlsruhe. The code was made publicly available and soon became the most widely used air shower Monte Carlo simulation and can now be considered the de facto standard in the community. Extensions of the physics capabilities of CORSIKA include CoREAS [101] to simulate radio emission and Cherenkov light [102]. The COAST extension, summarized in ref. [103], among other features, allows writing custom plugins to process particles and trajectories at runtime to add functionality not provided by CORSIKA itself. To better distinguish the long-standing FORTRAN version better from CORSIKA 8, we will refer to it as CORSIKA 7.

Well-known programs for numerical treatment of the CE, either pure or for hybrid simulations, are:

SENECA is a hybrid air shower simulation code written in FORTRAN [83]. The hadronic CE are solved numerically and solutions for EM sub-cascades are pre-tabulated. Its development has been discontinued.

CONEX is a FORTRAN hybrid code with full numerical treatment of the CE [78]. CONEX is now integrated in CORSIKA [103].

MCEq is a Python code mainly targeted at calculating inclusive fluxes [104], but it can also solve the CE with the air shower boundary conditions of eq. (3.2). In its present version it does not handle the electromagnetic cascades. MCEq is currently being extended to treat the lateral shower development [105].

EmCa is a CE solver for pure electromagnetic cascades written in Python [33].

3.4 Atmospheric models

A key ingredient to all air shower simulations, Monte Carlo and CE-based alike, are models of the atmosphere, especially its density $\varrho(h)$ as a function of height h above a reference, e.g. sea level. CE-solvers like MCEq, which work only one-dimensionally along the shower axis, can easily work with tabulated values, e.g. obtained from measured real-world data. For 3D Monte Carlo simulations this is often unfeasible and simple, analytical models are needed.

A simple model of the atmosphere, assuming an ideal gas of constant temperature T in a linear gravitational potential characterized by the acceleration g, yields the barometric formula [106]

$$\varrho(h) = \varrho_0 \exp\left(-\frac{mg}{k_\mathrm{B}T}h\right) \equiv \varrho_0 \exp\left(-\frac{h}{h_\mathrm{sc}}\right) \tag{3.5}$$

with reference density ϱ_0, mass of a gas molecule m and k_B denoting the Boltzmann constant. Typical values of the *scale height* h_sc on Earth are in the range 5 km to 10 km.

A more realistic model can be obtained by splitting the atmosphere in several layers with separate scale heights and reference densities. A well-known model in the field of cosmic ray physics is Linsley's model[1], consisting of four exponential and one top-most layer with constant density, as a parameterization of the 1976 *U.S. standard atmosphere* [108]

Linsley's parameterization, however, features discontinuities in the density profile at the boundaries between two layers, which can cause unphysical artefacts in simulations. These can be seen in the muon production distribution, where the discontinuity causes a sudden change of the critical energy. An example AMPD featuring an unphysical "step" is shown in fig. 3.3. Keilhauer et al. [109] provides an updated parameterization in which density varies continuously with height at the boundaries, so that these unphysical artefacts are not present.

[1] unpublished; more details can be found in ref. [107]

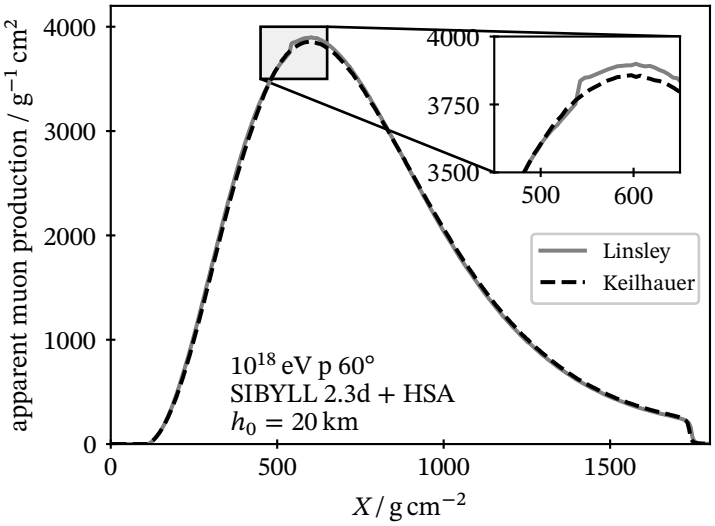


Figure 3.3: Apparent muon production distribution for Linsley's and Keilhauer's parameterizations of the US standard atmosphere

4 Numerical and probabilistic aspects of particle propagation in shower simulations

Shower simulations consist of two aspects: propagation of particles and stochastic event generation (decays and interactions). In this chapter we will deal with the first aspect, the numerics of particle propagation. In general, the problem addressed here is solving the equations of motion, taking into account all influencing forces such as electromagnetic fields and continuous energy losses, while at the same time sampling the point of stochastic events. As we will see, their distributions depend on the solutions to these equations of motion, so that both cannot be treated independently.

4.1 Equations of motion

The particle state is given by its four-vectors in coordinate space $x^\mu = (t, \boldsymbol{x})$ and momentum space $p^\mu = (E, \vec{p})$, with the constraint that particles are on-shell, i.e. $E^2 - \vec{p}^2 = m^2$ for particles of mass m. We treat t as independent and define $s(t) = (\boldsymbol{x}, \vec{p})$ as dependent variable. The evolution of $s(t)$ is governed by the equations of motion

$$\begin{aligned} \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} &= \frac{\vec{p}}{E}, \\ \frac{\mathrm{d}\vec{p}}{\mathrm{d}t} &= \sum_i \vec{F}_i(s). \end{aligned} \tag{4.1}$$

The force terms $\vec{F}_i$ usually comprise continuous energy losses and the Lorentz force in the presence of an electromagnetic field (in case of charged particles). For the sake of generality we will not make any assumptions regarding the force terms.

The task of a propagation routine of a shower simulation is to integrate these equations of motion, usually numerically, from the initial state up to the point when the particle shall not be propagated any further. One of these terminating conditions is the occurrence of a stochastic event.

4.2 Survival functions for stochastic events

Stochastic events happen randomly somewhere on the particle trajectory. The distributions of these points are characterized by ordinary differential equations (ODE) that govern the

evolution of the corresponding *survival probability*[1], i.e. the probability to not undergo a stochastic event.

The survival probability for interactions, $P_{\text{int}}(l)$, of a particle to travel a path of length l without undergoing an interaction fulfills

$$\frac{\mathrm{d}P_{\text{int}}}{\mathrm{d}l} = -n\sigma P_{\text{int}} \tag{4.2}$$

with the initial value $P_{\text{int}}(0) = 1$, where $n = n(\boldsymbol{x}(l))$ denotes the position-dependent number density of scattering targets, in our context atoms, and $\sigma = \sigma(s(l))$ the sum of the cross-sections of all considered interaction processes. It is convenient to introduce the concept of *grammage*[2] defined as integrated mass density ϱ along the path

$$X = \int \varrho \, \mathrm{d}l. \tag{4.3}$$

With this definition and the relation $\varrho = m_{\text{t}} n$ (m_{t} denotes the mass of a single target atom), eq. (4.2) reads

$$\frac{\mathrm{d}P_{\text{int}}}{\mathrm{d}X} = -\frac{\sigma}{m_{\text{t}}} P_{\text{int}} \equiv -\frac{1}{\lambda_{\text{int}}} P_{\text{int}}. \tag{4.4}$$

If we consider the *interaction length* λ_{int} as constant during the propagation, the solution of eq. (4.4) is simply an exponential distribution with scale parameter λ_{int}, which is easy to sample from. The complexity of the problem though is shifted into calculating $X(l)$, or rather its inverse $l(X)$, from eq. (4.3), which can be performed analytically only in a very limited number of cases, the most important of which we treat in section 4.3.

On the other hand, unstable particles are subject to decays. For a particle with lifetime τ_0 in its rest frame, the analogue of eq. (4.2) for the survival probability of decays $P_{\text{dec}}(t)$ reads

$$\frac{\mathrm{d}P_{\text{dec}}}{\mathrm{d}t} = -\frac{1}{\gamma\tau_0} P_{\text{dec}} = -\frac{m}{E\tau_0} P_{\text{dec}}. \tag{4.5}$$

The gamma factor given by the ratio of energy E of the particle to its mass m is a consequence of time dilation in the lab frame. If E does not change during the propagation, eq. (4.5) describes an exponential distribution. This is, however, often not the case as charged particles are subject to energy losses. In specific setups, also electric fields may be considered, which accelerate or decelerate the particle.

4.3 Grammage integration

4.3.1 Homogeneous density

The simplest case is a homogeneous density, i.e.

$$\varrho = \text{const.}, \tag{4.6}$$

[1] The terminology follows ref. [110].

[2] also referred to as *traversed matter* or *column density*

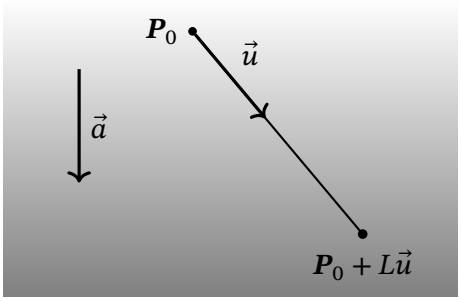


Figure 4.1: Visualisation of quantities used for grammage integration in the flat exponential model described by eq. (4.9)

and we have

$$X = \varrho L \tag{4.7}$$

for an arbitrary path of arclength L. The well-known GEANT 4 toolkit [111] for instance can handle only homogeneous media, motivated by the fact that essentially all man-made detectors in high-energy physics consist of piecewise-homogeneous materials.

4.3.2 Exponential models

Flat atmosphere

Simplyfing the Earth as flat, we can generalize eq. (3.5) as

$$\varrho(\boldsymbol{P}) = \varrho_0 \exp\left((\boldsymbol{P} - \boldsymbol{P}_{\text{ref}}) \cdot \frac{\vec{a}}{h_{\text{sc}}}\right). \tag{4.8}$$

Here $\boldsymbol{P}_{\text{ref}}$ is a reference point at which $\varrho(\boldsymbol{P}_{\text{ref}}) = \varrho_0$. $\vec{a}$ denotes a unit vector pointing in direction of the gradient, i.e. vertically downwards.

For a straight particle trajectory starting at point $\boldsymbol{P}_0$, pointing in (normalized) direction $\vec{u}$ and having length L (see fig. 4.1), eq. (4.3) yields

$$X = \varrho(\boldsymbol{P}_0)\frac{h_{\text{sc}}}{\vec{a}\cdot\vec{u}}\left(\exp\left(\vec{a}\cdot\vec{u}\frac{L}{h_{\text{sc}}}\right) - 1\right), \tag{4.9}$$

In the limit $\vec{a}\cdot\vec{u} = 0$, the trajectory lies in a plane of constant density and eq. (4.7) has to be used instead.

The flat-Earth approximation becomes more and more inaccurate for increasing zenith angles (the angle between $\vec{u}$ and $\vec{a}$) [112]. In particular, for horizontal showers with zenith angle 90°, the total grammage (*slant depth*) from infinite distance to sea level diverges.

Spherical atmosphere

A more accurate model assumes a spherical Earth with radius R. For a straight trajectory starting at height h_0 above sea level going in direction $\vec{u}$ and enclosing an angle ϑ with the

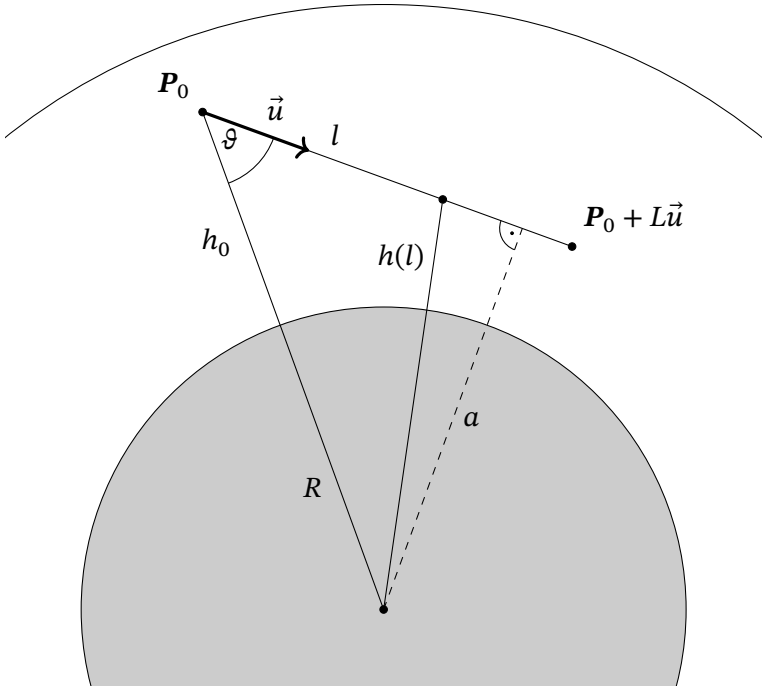


Figure 4.2: Straight trajectory starting at $\boldsymbol{P}_0$ with direction $\vec{u}$ and length L in spherical atmosphere. Furthermore, we define $s = R + h_0$ and $a = s|\sin\vartheta|$.

plumb-line direction, the relation between height and distance l along the trajectory is a radical function. From fig. 4.2 we find

$$(R+h)^2 = (R+h_0)^2 + l^2 - 2l(R+h_0)\cos\vartheta. \tag{4.10}$$

Solving this for $h(l)$ and inserting the result together with eq. (3.5) into eq. (4.3) yields

$$X(L) = \varrho_0 \int_0^L \exp\left(-\frac{1}{h_{\text{sc}}}\left[\sqrt{(R+h_0)^2 - 2(R+h_0)l\cos\vartheta + l^2} - R\right]\right)\mathrm{d}l \tag{4.11}$$

$$= \varrho(\boldsymbol{P}_0)\int_0^L \exp\left(-\frac{1}{h_{\text{sc}}}\left[\sqrt{s^2 - 2sl\cos\vartheta + l^2} - s\right]\right)\mathrm{d}l, \tag{4.12}$$

where we have used $\varrho(\boldsymbol{P}_0) = \varrho_0 \mathrm{e}^{-h_0/h_{\text{sc}}}$ and defined $s = R + h_0$. There is no known closed-form expression for the solution in general. It is instructive to consider two special cases: For $\cos\vartheta = \pm 1$ the situation simplifies to the flat atmosphere case. Furthermore it is possible to solve the definite integral in the limit $L \to \infty$ to determine the slant depth X_{hor} for horizontal showers ($\cos\vartheta = 0$) thanks to the identity [113, p. 364]

$$\int_0^\infty \exp\left(-a\sqrt{x^2 + b^2}\right)\mathrm{d}x = bK_1(ab), \quad \mathrm{Re}(a) > 0, \mathrm{Re}(b) > 0, \tag{4.13}$$

where $K_1(x)$ refers to the modified Bessel function of second kind. Applied to eq. (4.12) we find

$$X_{\text{hor}} = \varrho(\boldsymbol{P}_0)\, s\, \mathrm{e}^{s/h_{\text{sc}}} K_1\left(\frac{s}{h_{\text{sc}}}\right). \tag{4.14}$$

Equation (4.14) as it stands is numerically almost impossible to evaluate since $s/h_{\text{sc}} \sim 10^3$ for typical values. Expanding K_1 asymptotically and keeping only the first term,

$$K_1(z \gg 1) \approx \sqrt{\frac{\pi}{2z}}\mathrm{e}^{-z}, \tag{4.15}$$

yields

$$X_{\text{hor}} = \varrho(h_0)\sqrt{\frac{\pi}{2} h_{\text{sc}} s}. \tag{4.16}$$

This purely analytical formula can be convenient for some back-of-the-envelope calculations but not useful in air shower simulations.

In more general situations, eq. (4.12) can be straightforwardly integrated numerically. For numerical inversion, i.e. finding L for a given X_0, one may combine the numerical integration with a root-finding algorithm which iteratively determines a sequence of lengths $L^{(k)}$ and performs the integration $X(L^{(k)})$ until $X(L^{(k)}) \approx X_0$. Alternatively, one may formulate the problem as initial value problem of the ODE

$$\frac{\mathrm{d}L}{\mathrm{d}X} = \frac{1}{\varrho}, \tag{4.17}$$

and use standard numerical ODE solvers to integrate eq. (4.17) from $X = 0$ to X_0. These numerical procedures have the advantage that they can be applied not only to the density function considered in this section but also to very generic densities.

For application in air shower Monte Carlo codes, however, numerical integrations are prohibitively computationally expensive as the whole procedure has to be applied at least once for each particle in the shower. Therefore, approximations have been developed based on the fact that $h_{\text{sc}} \ll R$ and consequently only heights $h(l) \ll R$ contribute to the integral eq. (4.12) without exponential suppression.

Chernatkin [114] describes a technique similar to the saddlepoint approximation to approximate the integral (4.12) with a Gaussian by expanding the exponent to second order in $x \equiv l/s$.[3] In our notation, the result reads

$$X = \varrho(h_0)\sqrt{\frac{\pi}{2}\frac{h_{\text{sc}} s}{\sin^2\vartheta}}\exp\left(\frac{s}{2h_{\text{sc}}\tan^2\vartheta}\right)\operatorname{erf}\left(\sqrt{\frac{s}{2h_{\text{sc}}}}\left[x\sin\vartheta - \frac{1}{\tan\vartheta}\right]\right)\Bigg|_{x=0}^{L/s}. \tag{4.18}$$

For $\vartheta = \pi/2$ and $L \to \infty$ we recover the result for X_{hor} of eq. (4.16). This approximation is used in CONEX [78].

[3] This result is also presented in ref. [115].

A different approximation is explored by Glaser [116, pp. 165ff.], where eq. (4.12) is rewritten as

$$X = \varrho(h_0) \int_{h_0}^{h(L)+h_0} \mathrm{e}^{-\tilde{h}/h_{\mathrm{sc}}} \frac{\mathrm{d}l}{\mathrm{d}\tilde{h}} \, \mathrm{d}\tilde{h} \tag{4.19}$$

with $\tilde{h} \coloneqq h - h_0$ and

$$\frac{\mathrm{d}l}{\mathrm{d}\tilde{h}} = \pm \frac{\tilde{h} + s}{\sqrt{(\tilde{h} + s)^2 - a^2}}. \tag{4.20}$$

Here, we have defined the *impact parameter* $a \coloneqq s|\sin\vartheta|$ as minimum distance between the (infinitely extended) trajectory and the centre of the Earth. The point of closest approach is referred to as *perigee*. The sign depends on whether the considered part of the trajectory is approaching (−) or receding from (+) the perigee. The former case is the typical situation in downward-going showers, the latter in upward-going ones. For trajectories containing the perigee the method is not applicable since $\mathrm{d}l/\mathrm{d}\tilde{h}$ diverges. The method continues by expanding eq. (4.20) in $x \coloneqq \tilde{h}/s$ to any desired order, so that the remaing integrals to be solved are of the form

$$\int_0^{x'} \mathrm{e}^{-sx/h_{\mathrm{sc}}} x^n \, \mathrm{d}x = n! \left(\frac{h_{\mathrm{sc}}}{s}\right)^{n+1} \left(1 - \mathrm{e}^{-sx'/h_{\mathrm{sc}}} \sum_{k=0}^{n} \frac{(sx'/h_{\mathrm{sc}})^k}{k!}\right). \tag{4.21}$$

With this, the result reads

$$\begin{aligned} X = \varrho(h_0) s \Bigg[& \pm \frac{1}{|\cos\vartheta|} \frac{h_{\mathrm{sc}}}{s} \left(1 - \mathrm{e}^{-\tilde{h}(L)/h_{\mathrm{sc}}}\right) \\ & \mp \frac{\sin^2\vartheta}{|\cos^3\vartheta|} \left(\frac{h_{\mathrm{sc}}}{s}\right)^2 \left(1 - \mathrm{e}^{-\tilde{h}(L)/h_{\mathrm{sc}}} \left(1 + \frac{\tilde{h}(L)}{h_{\mathrm{sc}}}\right)\right) \\ & \pm 3 \frac{\sin^3\vartheta}{|\cos^5\vartheta|} \left(\frac{h_{\mathrm{sc}}}{s}\right)^3 \left(1 - \mathrm{e}^{-\tilde{h}(L)/h_{\mathrm{sc}}} \left(1 + \frac{\tilde{h}(L)}{h_{\mathrm{sc}}} + \frac{1}{2}\left(\frac{\tilde{h}(L)}{h_{\mathrm{sc}}}\right)^2\right)\right) \\ & + \ldots \Bigg]. \end{aligned} \tag{4.22}$$

Clearly, inverting this equation to obtain $\tilde{h}(L)$ from a given X is not possible in closed form and requires numerical root finding.

Sliding planar atmosphere

The implementation of the spherical atmosphere in AIRES, which was later also ported to CORSIKA [117], is the so-called *sliding planar atmosphere*. Here, the flat-Earth approximation is used locally while it is accurate. When it becomes inaccurate for long, inclined trajectories, a step-size limitation is imposed.

4.4 Sampling

We now turn to the question of how to sample from the distributions described by eqs. (4.4) and (4.5). As both exhibit the same structure, namely being linear, homogeneous ODEs of first order, it is useful to employ a more general notation. For simplicity we use time as independent variable, although other variables like arclength, grammage, or any other quantity that behaves monotonically along the path are equally valid. The evolution of the survival probability P_s is governed by

$$\frac{\mathrm{d}P_\mathrm{s}}{\mathrm{d}t} = -\alpha(s(t))P_\mathrm{s}(t). \tag{4.23}$$

The non-negative *hazard function* α depends on the local state of the particle, typically its energy, since cross-sections and decay time depend on it. In our original starting point of the discussion for interaction probabilities, eq. (4.2), also local properties of the medium occured, which have been integrated out by the change of variables from l to X, as long as density effects as described in section 2.3.3 do not play a role. The solution of eq. (4.23) is known:

$$P_\mathrm{s}(t) = \exp\left(-\int_0^t \alpha(s(t'))\,\mathrm{d}t'\right). \tag{4.24}$$

To perform the integration, however, the solution $s(t)$ of the equations of motion of the particle must be known since α depends on it. Depending on the hazard function at hand, the occurence of a stochastic event in finite time is either inevitable ($P_\mathrm{s}(t \to \infty) = 0$) or there is a chance such an event does not happen at all ($P_\mathrm{s}(t \to \infty) > 0$). The latter is the case if

$$\int_0^t \alpha(s(t'))\,\mathrm{d}t' < \infty. \tag{4.25}$$

We furthermore define

$$F(t) = 1 - P_\mathrm{s}(t) \tag{4.26}$$

and

$$p(t) = \frac{\mathrm{d}F}{\mathrm{d}t} = \alpha(s(t))P_\mathrm{s}(t). \tag{4.27}$$

In case of an inevitable stochastic event, $F(t)$ and $p(t)$ are a cumulative distribution function (c.d.f.) and probability density function (p.d.f.) in the strict sense. In the other case we have $F(t \to \infty) < 1$. For our purposes here, this is only of minor importance and we will still use the same nomenclature. The consequence for the sampling algorithms described below that work iteratively is that they may be non-terminating. Since in shower simulation codes usually also other conditions (cuts) may terminate the propagation, this does not result in infinite runtime.

How can $p(t)$ be sampled? If α varies only slowly one can assume it to be constant so that just a simple exponential distribution with $\alpha_0 = \alpha(t = 0)$ needs to be sampled from. With a suitable step-size limitation this condition can always be enforced and in case a sample exceeds the maximum permissible value, the particle is only propagated to that point without

undergoing a subsequent stochastic event and procedure is repeated. This method is employed e.g. in CRPropa [118]. Efficiency and accuracy of this method severely rely on the flatness of $\alpha(t)$ as well the magnitude of the error one is willing to tolerate.

4.4.1 Rejection sampling

An improvement of this method makes use of a form of rejection sampling: Instead of sampling from an exponential with parameter α_0, one chooses some upper bound on the codomain of $\alpha(s(t))$, $\alpha_{\max} \geq \alpha(s(t))$, and samples a value t_1 from the p.d.f.

$$\tilde{p}(t) = \alpha_{\max} \mathrm{e}^{-\alpha_{\max} t}. \tag{4.28}$$

The sample t_1 is then accepted with probability $p_{\mathrm{acc}} = \alpha(s(t_1))/\alpha_{\max}$, which requires the particle to be propagated to t_1 first, so that $\alpha(s(t_1))$ can be evaluated. If accepted, a stochastic event takes place. If rejected, no event takes places and the procedure is repeated, i.e. a new sample $t_2 > t_1$ is drawn (possibly with an updated $\alpha_{\max}$) until a t_k is accepted. Formal proofs that this method in fact yields samples correctly distributed according to eq. (4.24) are given by Stanev et al. [119] and Kawrakow [120]. Instead of repeating them here, we will simply provide some intuitive reasoning based on the implementation of this method in EGS4 [96, 120]. For simplicity, let us consider interacting particles as described by eq. (4.2) in a homogeneous medium, so that the hazard function is essentially given by the (non-constant) interaction cross-section $\sigma_{\mathrm{real}}(E)$. For the sampling, we artificially introduce a *fictitious* cross-section $\sigma_{\mathrm{f}}(E) \geq 0$ that corresponds to an "interaction" which leaves the projectile unchanged. We may essentially consider this a forward elastic scattering with vanishing momentum transfer, intistinguishable from a non-interaction. We define $\sigma_{\mathrm{f}}(E)$ such that

$$\sigma_{\mathrm{real}}(E) + \sigma_{\mathrm{f}}(E) = \sigma_0 = \mathrm{const} \tag{4.29}$$

for some arbitrary σ_0. If an interaction point is sampled using σ_0, the particle is propagated to that point and a "real" interaction is performed with a probabiliy corresponding to the branching ratio $\sigma_{\mathrm{real}}(E')/\sigma_0$, using the updated energy E'. If, however, the "fictitious" interaction is selected, propagation simply continues after sampling another interaction point.

While this method guarantees correctly distributed samples (assuming negligibly small error in the numerical e.o.m.-integration), its efficiency, i.e. the average p_{acc}, depends on the choice of $\alpha_{\max}$ as well as the shape of the hazard function. Both have an impact on how many intermediate samples are rejected before the final sample is accepted – desirably as few as possible: Each of the intermediate steps requires two additional random numbers to be drawn. Furthermore, these intermediate steps also require additional steps of the e.o.m.-integration in order to evaluate $\alpha(s(t_i))$. As it requires more than just local information about the hazard function and the particle state, $\alpha_{\max}$ is often not easy to estimate. It depends on the particular problem at hand how much knowledge about the global behaviour of the hazard function can be assumed. Moreover, large variations of α away from its maximum value cause a higher rejection rate. Put differently, the efficiency is high if α stays almost constant and a value for $\alpha_{\max}$ can be given that is close to $\max \alpha(t)$. If $\alpha(t)$ decreases during the propagation, $\alpha_{\max} = \alpha(s(t = 0))$ is the optimal choice.

If the e.o.m.-integration alone requires a rather dense stepping compared to the stochastic steps, the computational costs of inefficient sampling are negligible for the propagation as a whole (e.o.m.-integration plus sampling). If, however, the lengths of rejected stochastic steps become comparable to or smaller than the step-lengths of the pure e.o.m.-integration, the sampling efficiency becomes a limiting factor.

4.4.2 Inverse transform sampling

Another class of sampling methods builds on the transformation properties of probability distributions when functions are applied on individual samples. If we draw a sample x from a distribution with p.d.f. $p_X(x)$ and apply a strictly monotone function φ to obtain a value $y = \varphi(x)$, then its p.d.f. $p_Y(y)$ fulfills

$$p_X(x) = p_Y(y)\left|\frac{\mathrm{d}y}{\mathrm{d}x}\right| = p_Y(y)|\varphi'(x)|\,. \tag{4.30}$$

We obtain a particularly important statement for the special case that the applied function is the c.d.f. of x, $F(x)$. In that case we have $\mathrm{d}F/\,\mathrm{d}x = p_X(x)$, so that

$$p_Y(y) = 1. \tag{4.31}$$

Phrased in words, this means that the c.d.f. of any probability distribution is itself distributed uniformly on the unit interval $[0, 1]$. This fact can be exploited to sample from a distribution by inverting its c.d.f.: Given a uniformly distributed variate $u \in [0, 1)$, the variate

$$x = F^{-1}(u), \tag{4.32}$$

where $F^{-1}(u)$ is called *quantile function*, is distributed with c.d.f. $F(x)$. Applied to our problem, this means we need to solve

$$u = F(t) = 1 - \exp\left(-\int_0^t \alpha(s(t'))\,\mathrm{d}t'\right) \tag{4.33}$$

for t. This principle is followed in the code MMC [121] as well as its successor PROPOSAL [122]. They propagate leptons one-dimensionally through homogeneous media with continuous energy losses but without (electro-)magnetic fields. Since under these special cicrumstances the hazard function depends solely on energy, in eq. (4.33) the substitution $t \to E$ is made. Defining $f(E) = \mathrm{d}E/\mathrm{d}X$ as continuous energy losses per unit of grammage and combining the hazard functions of interaction and decay, we obtain

$$u = 1 - \exp\left(-\int_{E_\mathrm{i}}^{E_\mathrm{f}} \left(\frac{\sigma(E)}{m} + \frac{m}{E\tau_0}\frac{1}{\varrho_0 v}\right)\frac{1}{f(E)}\,\mathrm{d}E\right). \tag{4.34}$$

Given the initial energy E_i of the particle, the equation needs to be solved for the final energy E_f at the time of the stochastic event. In this representation the integration can be readily performed numerically and the result tabulated in dependence of $E_{\mathrm{i/f}}$. This allows for a fast lookup of the final-state energy (and other final-state quantities like displacement) for a given u.

4.4.3 ODE-based sampling for CORSIKA 8

For a full 3D particle propagation in inhomogeneous media this pre-calculation is impossible. It is nevertheless conceivable to perform the integration in eq. (4.33) on-the-fly during the propagation and trigger a stop of the propagation as soon as the value of the integral crosses the threshold defined by the value of u drawn at the beginning of the propagation. This will, however, always overshoot the exact value to a certain degree since the integral grows in discrete amounts determined by the steplength.

For usage in CORSIKA 8, none of the methods described above fully suits the needs and I have devised a new method that combines the idea of inverse transform sampling, the ODE governing the evolution of the distribution to be sampled from, and the e.o.m. The key point is to change the independent variable of the e.o.m. from t to u and extend the e.o.m. with an attional equation for $\mathrm{d}t/\,\mathrm{d}u$. Using eqs. (4.23) and (4.33), we can derive

$$\frac{\mathrm{d}t}{\mathrm{d}u} = \left(\frac{\mathrm{d}u}{\mathrm{d}t}\right)^{-1} = -\left(\frac{\mathrm{d}P_\mathrm{s}}{\mathrm{d}t}\right)^{-1} = \frac{1}{\alpha P_\mathrm{s}} = \frac{1}{\alpha(1-u)}. \tag{4.35}$$

With this, the e.o.m. can then be rewritten as

$$\frac{\mathrm{d}s}{\mathrm{d}u} = \frac{\mathrm{d}s}{\mathrm{d}t}\,\frac{\mathrm{d}t}{\mathrm{d}u}\,. \tag{4.36}$$

The procedure is straight-forward: A flat random number u^* is drawn and the IVP defined by eq. (4.36) is solved from $u = 0$ to u^*. The final state at the end of the step, just before the interaction, is then given by $s(t(u^*))$.

The main advantage of this method is that it can be applied to essentially any non-zero hazard function since only local information is needed. Since the grammage calculation is no longer necessary, arbitrary density functions can be used, not limited to, for example, (stepwise) exponential models. This remedies a long-standing shortcoming that has non-negligible effects on the systematic uncertainties of the energy scale of the CTA observatory [123]. Furthermore, density effects on cross-sections can be handled naturally and no rejection sampling is necessary as it is the case e.g. in CORSIKA for the LPM effect [84]. Furthermore, decay and interaction can be combined into a single hazard function so that only one sample is necessary. In the conventional approach both are treated independently from each other and the one happening earlier is selected.

For stable, non-interacting particles (most notably neutrinos) the method cannot be applied since the hazard function vanishes. Under these circumstances the e.o.m. either need to be solved in their original form or one introduces fictitious cross-sections as described above.

Example: Minimum ionizing atmospheric muon

We consider a muon with (total) energy $E_0 = 20m_\mu$ injected at height $h_0 = 15\,\mathrm{km}$ a.s.l. at an angle of $\vartheta = 60°$. The atmosphere is spherical with a density at sea level of $\varrho_0 = 1.2\,\mathrm{kg\,m^{-3}}$ and a scale height of $h_\mathrm{sc} = 8\,\mathrm{km}$. The muon is propagated one-dimensionally with continuous energy loss modeled with a constant value of $\mathrm{d}E/\,\mathrm{d}X = -2\,\mathrm{MeV\,g^{-1}\,cm^2}$. The particle state

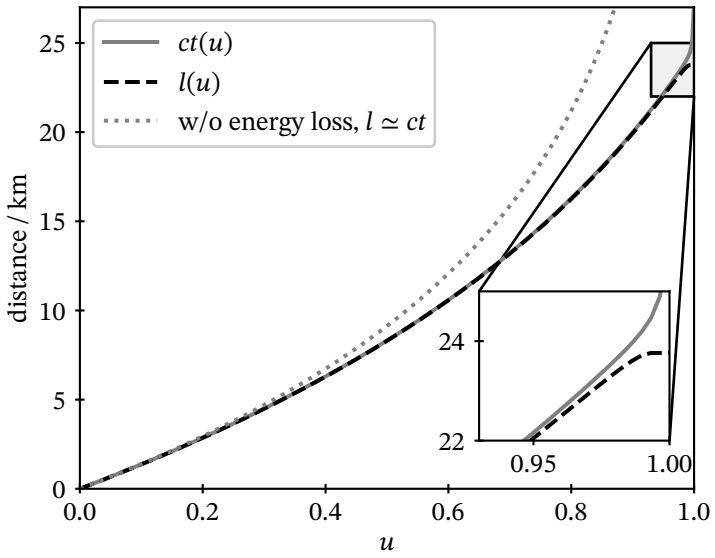


Figure 4.3: Quantile functions of time ct and displacement l until decay

$s = (ct, l, E)$ is defined by elasped time (in units of length) ct, propagated distance l and total energy E. We consider only decay as stochastic event. The e.o.m. of this scenario read

$$\begin{aligned} \frac{\mathrm{d}ct}{\mathrm{d}u} &= \frac{m_\mu}{Ec\tau}\frac{1}{1-u}, \\ \frac{\mathrm{d}l}{\mathrm{d}u} &= \frac{\mathrm{d}l}{\mathrm{d}ct}\frac{\mathrm{d}ct}{\mathrm{d}u} = \frac{1}{E}\sqrt{E^2 - m_\mu^2}\,\frac{\mathrm{d}ct}{\mathrm{d}u}\,. \\ \frac{\mathrm{d}E}{\mathrm{d}u} &= \frac{\mathrm{d}E}{\mathrm{d}X}\,\varrho(h(l))\,\frac{\mathrm{d}l}{\mathrm{d}u}\,. \end{aligned} \tag{4.37}$$

Figure 4.3 shows the solutions $ct(u)$ and $l(u)$. They coincide over a large fraction of the u range due to the ultrarelativistic motion. Only at $u \gtrsim 0.9$ deviations arise as the muon slows down and eventually comes to rest so that $l(u)$ no longer increases. On the other hand, $ct(u)$ diverges logarithmically as $u \to 1$. Additionally, the analytic solution disregarding energy losses, $l \simeq ct = -m_\mu c\tau \ln(1-u)/E$, is shown.

The survival probability is depicted in fig. 4.4. For small times it decreases exponentially with its initial (boosted) lifetime. At large times, when the Lorentz factor has approached unity, the behaviour is exponential again with the lifetime at rest. Figure 4.4 is obtained from the numerical solution essentially by inverting the axes of fig. 4.3.

Finally, fig. 4.5 shows the obtained probability densities. In terms of the numerical solutions, they can be expressed as

$$p(s_i) = \left(\frac{\mathrm{d}s_i}{\mathrm{d}u}\,(u, s(u))\right)^{-1}. \tag{4.38}$$

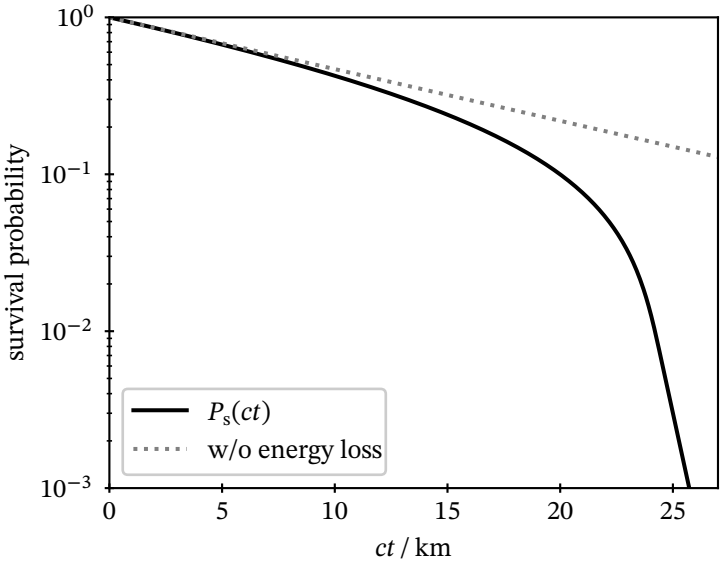


Figure 4.4: Survival function $P_s(ct)$

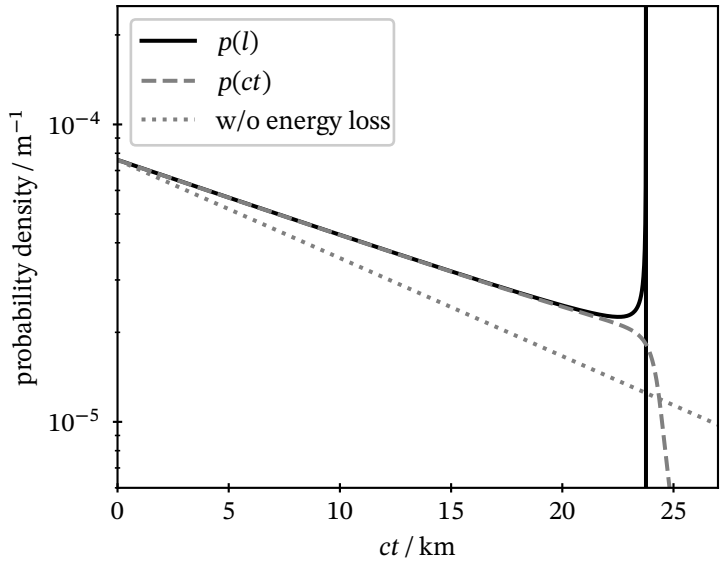


Figure 4.5: Probability densities of propagated distance l and elapsed time ct before decay

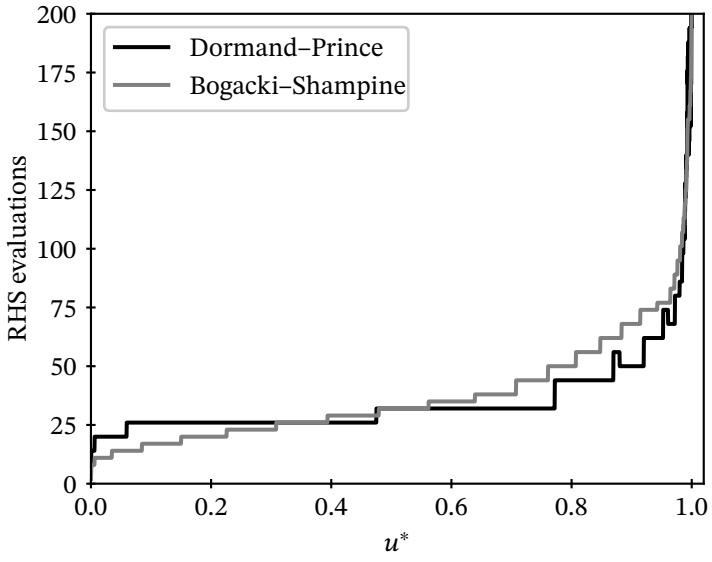


Figure 4.6: Number of right-hand side evaluations of the ODE system

Also here the exponential behaviour of $p(ct)$ for the two limits $E \simeq E_0$ and $E \simeq m_\mu$ is visible. On the other hand, $p(l)$ diverges at the point where the muon has come to rest, but in a way that the area under the curve appoaches unity.

The performance depends on the integration scheme one employs to solve the system. Typically one will use an adaptive scheme that determines a suitable step-length based on the desired accuracy of the solution. A measure of the performance is the number of evaluations of the right-hand side (RHS) of the ODE system. In fig. 4.6 the number of RHS evaluations of eq. (4.37) as a function of the upper limit of integration u^* is shown exemplarily for two different integration schemes, Dormand-Prince [124], an explicit Runge-Kutta method of order (4)5, and Bogacki-Shampine [125], an explicit Runge-Kutta method of order (2)3, both of which as implemented in SciPy [126]). The relative tolerance is set to 10^{-4}. We find that for small values of u^*, Bogacki–Shampine needs fewer evaluations, whereas Dormand-Prince performs better at large values. Since u^* is uniformly distributed, the differences avarage out.

Conclusions and outlook

The example demonstrates that the proposed method is a convenient way to solve the problem of sampling and propagation at the same time. Arbitrary forces can be included in the e.o.m. and all aspects are treated on the same footing. In particular, it is possible to combine both hazard functions of decay and interaction into a single one. In the conventional approach, one typically samples interaction and decay points separately and chooses the one that happens first, which requires converting t, l and X to a common variable. This is not easy to do as velocity is not constant (at least when the ultrarelativic approximation

becomes invalid), trajectories are curved (with changing curvature due to energy losses) and X is already difficult to calculate for straight trajectories.

At the time of writing, the new method is not yet implemented in CORSIKA 8 and the overall performance in a full shower simulation with particles at all energies remains to be seen.

5 Architecture of CORSIKA 8

This chapter is devoted to the architecture of CORSIKA 8. Starting with the motivation to develop a new code, the outline follows a bottom-up approach: After the description of low-level fundamentals, the more high-level building blocks and concepts are explained. Many parts have been published in several publications: The motivation for the project and early design considerations are described in the CORSIKA 8 whitepaper [6]. The unit and geometry systems (sections 5.3.1 and 5.3.2) are discussed also in ref. [127]. The random-number generation technique used in CORSIKA 8 (section 5.3.4) is explained in detail in ref. [128]. An early account of the environment system (section 5.4) as well as the overall program structure (section 5.7) is given in ref. [129]. Section 5.8 has been published in ref. [130], which also contains a description of the output format covered in section 5.10.

Significant original contributions of this thesis to CORSIKA 8 are sections 5.3.1 to 5.3.3, 5.4 and 5.8, the histogram serialization described in section 5.10 as well as a large part of section 5.7.

5.1 Motivation

Although serving as a major workhorse for a large fraction of the astroparticle physics community for more than 30 years, it has become more and more evident that the existing codebase of CORSIKA can hardly be maintained for a long term period in a way that meets the ever increasing demands. Implementing new features necessary to tackle current and future challenges requires detailed in-depth knowledge of the complex and monolithic source code. At the same time, Fortran, once considered the lingua franca of high-energy physics, has largely been replaced by C++ for performance critical code [131] so that only few physicists still learn Fortran. Both aspects together pose a hurdle to potential developers. A number of issues have been identified

- Propagation is possible only in air. The atmospheric model is fixed to a Linsley-like parameterization consisting of four exponential and one constant-density layers.
- The shower geometry is to a large degree tailored to standard, down-going showers with a well-defined point of first interaction. Upward-going showers require a special geometry to be selected at compile-time [132] and only the longitudinal profile is available as output.
- The particle history (available with EHISTORY option [133]) is limited to the mother and grandmother generations. Information about previous ancestor generations is unavailable.

- Physics modules like radio or Cherenkov emission cannot influence the particle simulation. In particular, the step-length cannot be shortened when it might be desirable for higher accuracy.

It was therefore decided to design and develop a new code that later has been named CORSIKA 8. It is envisaged to replace the Fortran version as soon as it has reached a sufficient level of maturity and covers all relevent aspects of EAS simulations accurately. It is designed as a framework to accommodate different user communities' needs, who can build their individual applications on top of it. Since its inception CORSIKA 8 has been developed as a community-driven open source project.

5.2 Design Principles

The design of CORSIKA 8 is based on a number of aspects. The overarching principle is *modularity*: Individual aspects of the simulation and its underlying algorithms are grouped in logical units (modules) responsible for only that specific task. The functionality they provide is exposed via well-defined interfaces. No knowledge about the internal workings of other modules shall be imposed (separation of concerns). This makes it possible to develop individual modules independently from each other. More concretely, the modular approach in CORSIKA 8 allows one to freely combine and exchange certain physics modules depending on the actual requirements, providing a great degree of *flexibility*.

Furthermore, *performance* is of high importance. For some purposes, air shower simulations are needed in vast numbers — either to cover large parameter spaces or to obtain high statistics. For instance, large scale simulation campaigns of the CTA observatory consumed more than 2×10^8 h of CPU time in the year 2018 [134]. Optimizations therefore can have significant impact both in terms of monetary costs as well as energy consumption. Designing efficient code requires awareness of modern CPU and memory architecture and the techniques are employed to achieve high performance like branch prediction, vectorization and cache hierarchies, to name a few. Therefore, static design patterns are preferred over dynamic ones in CORSIKA 8 wherever useful and appropriate data structures are chosen. Nevertheless, at the time of writing performance assessments of CORSIKA 8 and corresponding optimization are still at an early stage many details might be revised in the future.

To prohibit the introduction of bugs and ensure *correctness*, most functionality implemented in CORSIKA 8 is accompanied by unit tests, i.e. sanity checks that compare the actual result e.g. of function calls with the expected behaviour. These unit tests are executed in a continuous integration system on the whenever a commit is made on the development platform[1]. Also, a policy of the project requires new code to have corresponding unit tests in place before it can be considered to be merged.

[1] `https://gitlab.ikp.kit.edu/`

5.3 Fundamentals

5.3.1 Unit system

The incorrect usage of physical units of measure is regarded as one of the three most common errors in scientific computing [135]. The perhaps most illustrative example is the loss of the Mars Climate Orbiter during its orbital insertion maneuver, the cause of which is largely attributed to the mixture of imperial and metric units [136]: One part of the orbiter's onboard software, provided by an external contractor, supplied data in imperial units, while another system, developed by NASA, interpreted these data as being in SI units.

Especially in a collaborative environment such as CORSIKA 8, a strategy to avoid these kinds of errors is a key asset, also because their presence may render the simulation wrong in an inconspicuous manner. A possible solution, annotating data types with units, was proposed already more than 40 years ago [137]. However, hardly any programming language today provides built-in support so that usually custom libraries are required [138]. Umrigar [139] descibes the first application of the technique of template metaprogramming in C++ to provide dimensional analysis at compile time. A conceptually very similar solution is employed in CORSIKA 8, based on the *PhysUnits C++11* library [140] with some custom modifications.

The basic idea is to introduce new data types for each dimensionful quantity occuring in the code (length, time, density, etc.) and enforce their usage instead of plain floating point numbers. Multiplication and division of dimensionful quantities yields quantities with different dimensions, so that an arbitrary number of such data types needs to be defined. With C++ this can readily be achieved with template metaprogramming:

A templated type `dimensions<N1, N2, N3, N4, N5, N6, N7, N8>` is introduced in order to keep track of the dimensions of a quantity. The first seven integers $N_{1,\ldots,7}$ represent the exponents of the basic physical dimensions *length*, *mass*, *time*, *electric current*, *absolute temperature*, *amount of substance* and *luminous intensity*. The final integer represents the exponent of a custom *HEP energy dimension* whose purpose will be explained below.

The actual type for quantities, `quantity<Dim, Rep=double>`, is likewise templated. The first template argument is a `dimensions` type, the second selects the underlying floating point type that is used to store the numerical value of the quantity, by default **`double`**, in its base unit. For better readability type aliases like `TimeType` or `GrammageType` are defined, so that one usually does not need to handle the dimension indices explicitly.

The laws of quantity calculus (see e.g. ref. [141]) are encoded in C++. For instance, only quantities of the same dimension can be added and multiplying a `quantity<dimensions<N1, N2, N3, ...>>` with a `quantity<dimensions<M1, M2, M3, ...>>` returns a `quantity<dimensions<N1+M1, N2+M2, N3+M3, ...>>`. This way, dimensional analysis of all calculations involving quantities is conducted during compilation and any violation of the laws of quantity calculus results in a compiler error.

Besides the dimensional analysis, also the conversion of units to common base units are performed. A number of pre-defined constants as well as *user-defined literals* for convenience are provided to initialize a quantity. Writing `l = 1.450_km` (where `l` is a `quantity` for lengths) or `B = 0.48 * gauss + 5_uT` (where `B` is a `quantity` for magnetic flux density) converts the

quantity from the value in the stated unit to the base unit.

The strict dimensional analysis based on the SI becomes inconvenient when one switches to the natural ("god given" [142]) units of particle physics, where the convention $c = \hbar = 1$ is employed, which reduces the number of dimensions by two. The unit system as described so far prevents the use of intentional "sloppy" expressions like $E^2 = p^2 + m^2$ in code. This is remedied by the introduction of the spurious eighth HEP energy dimension which we treat as independent. Quantities of that type (`HEPEnergyType`) are intended for particle masses, energies and momenta, given as multiples of electronvolts. Only a few situations in CORSIKA 8 exist where these microscopic units come into contact with the macroscopic SI units. For these cases conversion functions are provided to convert back and forth between SI and natural units when possible, which mainly relies on the identity $\hbar c = 197.327\,\mathrm{MeV\,fm}$ that relates energy with length. To convert quantity q whose dimensions are length$^l \times$ time$^t \times$ mass^m from SI to natural units, the conversion reads

$$\frac{q}{\mathrm{eV}^{m-t-l}} = \left(\frac{\hbar c}{\mathrm{eV\,fm}}\right)^{m-t-l} \cdot \left(\frac{\hbar}{\mathrm{kg\,m^2\,s^{-1}}}\right)^{-m} \cdot \left(\frac{c}{\mathrm{m\,s^{-1}}}\right)^{m+t} \times \frac{q}{\mathrm{m}^l\,\mathrm{s}^t\,\mathrm{kg}^m}, \tag{5.1}$$

so that the final dimensions are energy^{m-t-l}. The inverse conversion of a quantity in natural units with dimension energy^e reads

$$\frac{q}{\mathrm{m}^l\,\mathrm{s}^t\,\mathrm{kg}^m} = \left(\frac{\hbar c}{\mathrm{eV\,fm}}\right)^{-e} \cdot \left(\frac{\hbar}{\mathrm{kg\,m^2\,s^{-1}}}\right)^{m} \cdot \left(\frac{c}{\mathrm{m\,s^{-1}}}\right)^{t-m} \times \frac{q}{\mathrm{eV}^e}, \tag{5.2}$$

with the constraint $m - l - t = e$.

It should be noted that none of the features described here have any negative impact on runtime performance as long as the most basic compiler optimizations (inlining) are enabled. A short non-trivial test, compiling a function for calculating the scalar product of two arrays with g++ 8.4.1, resulted in identical assembly code for `quantity` as for **`double`**.

5.3.2 Geometry

The second cornerstone of CORSIKA 8 are the geometry classes that deal with points, vectors and coordinate systems. The design borrows ideas from the *Offline* software framework of the Pierre Auger Collaboration [143]. Points and vectors are modeled not just as 3-dim. tuples of their components/coordinates but are always defined with respect to a specific coordinate system (CS), of which there can be multiple. Having several CS are useful for instance when interfacing event generators, which usually follow the convention to align the momenta of the projectile particle along the z-axis. One unique root CS is predefined and new CS are defined in terms of a reference CS and a transformation matrix that relates the the new CS with its reference. Supported transformations are elements of the special Euclidean group SE(3), i.e. translations and rotations, represented by a 4×4 transformation matrix. They come into play when operations involve two vectors/points defined in different CS. In this case the components/coordinates of one of the objects are temporarily transformed into the CS of the other object. Since the CSes form a tree structure (see fig. 5.1), this is achieved by

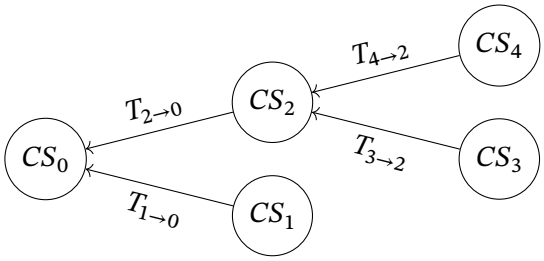


Figure 5.1: Example coordinate system tree. CS_0 is the root coordinate system.

multiplying the transformation matrices along the shortest path connecting the two CS. For example the transformation required to go from CS_3 to CS_1 is $T_{3\to1} = T_{3\to2}T_{2\to0}T_{1\to0}^{-1}$. These transformations happen completely transparent to the developer and the explicit usage of CS is barely necessary. Most often, expressions can be very close to the symbolic notation used in equations as well. For instance, eq. (4.8) can be expressed as `rho0 * exp(axis.dot (P - Pref)/ scaleHeight)`. Explicit usage of coordinates is required mostly during the initialization phase of a simulation, when the geometry is set up.

For the actual linear algebra computations the geometry system relies on the Eigen3 library [144], which is highly optimized.

We distinguish clearly between points and vectors as they appear in *affine space* (see e.g. ref. [145]): Points are subject to rotations and translations. Vectors can be thought of as displacements between two points (or multiples thereof) and are not affected by translations. Moreover, vector components can also carry arbitrary dimensions (”units”) as described in section 5.3.1, while point coordinates are necessarily lengths. The allowed operations follow the rules of the affine structure of Euclidean space: Points and length vectors can be added to return another point, subtracting two points yields a length vector. Scalar and cross-products are defined for vectors, respecting their dimensions.

5.3.3 Partice ID and physical properties

To distinguish different particle species, an integer *particle ID* is introduced. It is not only used to label a certain particle but also serves as an index to lookup tables for masses, lifetimes, electric charges, etc. These data are read from the `ParticleData.xml` file of the Pythia 8 distribution during the compilation of CORSIKA 8. At the same time, the CORSIKA-8-internal particle IDs are generated for each particle simply by incrementing a counter, resulting in numbers currently in the range 1 to 175. The numeric values never need to be used directly, though, as they are exposed via human-readable **`enum class`**es like `Code::MuPlus` or `Code::SigmaMinusBar` for higher expressiveness.

Conversion functions between CORSIKA 8 IDs and the *Monte Carlo Particle Numbering Scheme* of the Particle Data Group (PDG codes) [11] are provided for interoperability with other software. These cannot conveniently be used directly as internal particle IDs because they sparsely cover a large range of numbers, making them unsuitable as lookup indices.

Nuclei, which are identified by their mass and atomic numbers (A, Z), are treated differ-

ently as explained in section 5.6.

5.3.4 Random-number generation

Being a Monte Carlo code, CORSIKA 8 heavily relies on (pseudo-)random numbers, which are consumed in abundance. Recently, the random number generator of CORSIKA 8 was changed to a counter-based random number generator (CBRNG) [128]. The defining propery of CBRNG is that the advancement of their internal state is simply incrementing a counter, which allows advancing the state by an arbitrary number of steps at once. This makes them attrative for use in parallel algorithms in need of random numbers: A CBRNG-based stream of random numbers can be broken into a number of independent sub-streams by dividing the counter range into many distinct intervals, each still containing more numbers than needed by the simulation.

CORSIKA 8 follows the same idea as CORSIKA 7 of employing a distinct stream of random numbers for each module, each with its own seed.

5.4 Environment

CORSIKA 8 is designed to allow propagation of particle cascades in user-definable environment setups which allows a great deal of flexibility in three aspects:

- The medium of propagation is freely selectable, considering the limitations set by the physics modules.
- Worlds can be modeled consisting of different media. Example use-cases are air showers that continue their propagation below ground, e.g. in soil, water or ice, or τ-induced showers emerging from mountains. Transitions between two media are also considered.
- The environment/medium properties can be customized. Depending on the use-case simulations may e.g. require querying (electro-)magnetic field strengths or refractive index as a function of position.

5.4.1 Worldbuilding with the volume tree

To simulate particle cascades in several media it is necessary to set up a data structure and corresponding algorithms to map spatial regions to media in a way that offers flexibility and at the same time does not impact performance too much. In CORSIKA 8 this is achieved with the *volume tree*. The nodes of this tree structure are geometric primitives (volumes) like spheres and their placement in the tree represents containment, i.e. a parent node contains its children. Whis this arrangement it is possible to perform a lookup of the node to which a certain point belongs. More importantly, the search space for calculating possible entry/exit points of particle trajectories crossing volume boundaries is small because only as subset of all nodes needs to be considered.

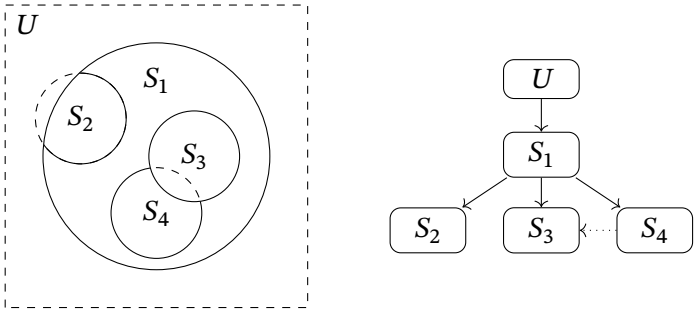


Figure 5.2: Example volume tree

An example world consisting of a number of spheres S_i together with its corresponding volume tree are depicted in fig. 5.2. The root node U, called *Universe*, is equivalent to a sphere with infinite radius. It exists to make sure each point in space can b mapped to a well-defined volume node. Complications arise when volumes have intersections with each other so that the idea of containment needs rethinking. We can distinguish between two classes of such intersections:

- An intersection of a child volume with its parent ($S_1 - S_2$ in fig. 5.2). In this case the outer volume (parent) clips the inner volume (child).

- An Intersection of two child nodes at the same level requires additional information. For these cases nodes contain a (by default empty) list of pointers to *excluded* nodes. A point does not belong to a node if it is contained by any of the excluded nodes. In fig. 5.2 this is the case for the overlap between S_3 and S_4. S_4 has S_3 in its exclusion list (indicated by the dotted arrow) so that points in the overlapping region belong to S_3.

5.4.2 Dressing volumes with models

The individual volumes alone do not constitute any description of the media and their physical properties. Therefore, they have to be furnished with models. For this purpose we make use of dynamic polymorphism: An abstract class defines the interface and serves as base class. Implementations of this interface happen in classes which derive from the interface class. A special challenge arises in CORSIKA 8 because the medium interface is not fixed and depends on the physics modules one wants to use for a particular simulation: Cherenkov and radio emission modules need to query the index of refraction, electromagnetic interaction models, in particular PROPOSAL, require material constants like the radiation length, and for studies of EAS development in thunderclouds electric fields need to be modeled. To accomodate this, we make use of *mixin inheritance*, which is a template class that inherits from its template argument (see e.g. ref. [146]). This way, several interface classes, each one responsible for a single aspect, can be chained together to form the complete interface. Listing 5.1 shows a simplified example of how three different properties are combined to form a single abstract base class.

Listing 5.1: Mixin-based environment interface composition (simplified)

```
struct IMediumModel {
  virtual DensityType getMassDensity(Point const &) const = 0;
};

template <typename T> struct IRefractiveIndexModel : public T {
  virtual double getRefractiveIndex(Point const &) const = 0;
};

template <typename T> struct IMagneticFieldModel : public T {
  virtual MagneticFieldVector getMagneticField(Point const &) const = 0;
};

using MediumInterface =
    IMagneticFieldModel<IRefractiveIndexModel<IMediumModel>>;
```

Listing 5.2: Mixin-based composition of implementations (simplified)

```
template <typename T> struct FlatExponentialDensity : public T {
  // [...]

  virtual DensityType getMassDensity(Point const &P) const override {
    return rho0 * exp(axis.dot(P - Pref) / scaleHeight);
  }
};

template <typename T> struct HomogeneousDensity : public T {
  // [...]

  virtual DensityType getMassDensity(Point const &p) const override {
    return rho0;
  }
};

template <typename T> struct ExponentialRefractiveIndex : public T {
  // [...]

  virtual double getRefractiveIndex(Point const &p) const override {
    // [...] some implementation
  }
};

template <typename T> struct UniformMagneticField : public T {
  // [...]

  virtual MagneticFieldVector getMagneticField(Point const &p) const override
      {
    // some implementation [...]
  }
};

void printMassDensity(MediumInterface const& medium) {
  Point const p = make_some_point(); // obtain a Point
```

```
  // call implementation via dynamic dispatch
  DensityType rho = medium.getMassDensity(p);
  std::cout << "density at p = " << rho << std::endl;
}

int main() {
  ExponentialRefractiveIndex<
      FlatExponentialDensity<UniformMagneticField<MediumInterface>>>
      modelA;

  ExponentialRefractiveIndex<
      HomogeneousDensity<UniformMagneticField<MediumInterface>>>
      modelB;

  printMassDensity(modelA);
  printMassDensity(modelB);

  return 0;
}
```

The actual implementations of the interfaces, which in the end must inherit from their respective interface class, follow the same pattern, so that they can be freely combined and orthogonal aspects stay independent. Listing 5.2 shows example implementations of the interfaces of listing 5.1 and their usage via dynamic polymorphism. Two objects `modelA` and `modelB` are created which contain different implementations of the `getMassDensity()` interface. The function `printMassDensity()` handles both objects via their interface and is agnostic about the implementation, which is selected only at runtime. In a similar way, volume nodes are linked to models only via the interfaces.

The most fundamental medium properties are density and the fractions of the isotopes, which always need to be specified.

5.5 Processes

Processes are the entities that act on particles in various ways. They are grouped in six categories:

IteractionProcess This class of processes models interactions and related functionality. Typical examples are wrappers around hadronic interaction models. Two methods must be provided: `GrammageType getInteractionLength(Particle const&)` needs to return the (possibly infinite) interaction length of the modeled physical process for the particle currently being propagated in the current medium. The actual event generation has to be performed in the `doInteraction(SecondaryView&)` method, which typically adds secondary particles via the `SecondaryView`.

DecayProcess This class of processes models decays. Methods to be provided are: `TimeType getDecayTime(Particle const&)`, which returns the lab-frame decay time of the

particle currently being propagated. The decay event is generated in the `doDecay (SecondaryView&)` method, which typically fills the `SecondaryView` with the decay products.

BoundaryCrossingProcess This class of processes is relevant if any action shall be performed when a particle exits its current volume to enter another. In that case, the method `doBoundaryCrossing(Particle&, VolumeNode&, VolumeNode&)` is called. The last two parameters refer to the current and new volume nodes, respectively.

SecondariesProcess After an interaction or decay event has been performed, the newly generated secondaries can be processed further, e.g. filtered and/or modified. This happens in the `doSecondaries(StackView&)` method, in which the secondaries are typically iterated over.

ContinuousProcess In this class of processes aspects concerning the continuous movement of a particle along its trajectory are handled. Before the actual propagation takes place, its `LengthType getMaxStepLength(Particle const&, Trajectory const&)` method is called, which returns a maximum (possibly infinite) step-length. It serves as a hint to the propagation to limit the step-size to that value if necessary. This functionality is provided to ensure numerical accuracy. For instance, a process implementing energy losses may limit the step-size to make sure that the energy of the particle does not change too much, so that the decay time stays approximately valid. `ProcessReturn doContinuous(Particle&, Trajectory const&, bool)` is executed after the length of the trajectory has been determined. The particle properties may be modified, e.g. the energy reduced. The boolean input parameter indicates whether the previous step-length limitation had been caused by the `getMaxStepLength()` method by the same process.

The process can indicate whether the particle shall be regarded as absorbed via the return value.

StackProcess Primarily for statistical purposes it is beneficial to execute code periodically after each N cascade steps. This functionality is provided with this class of processes, which require a `doStack(Stack&)` method. An example usecase is the estimation of the remaining runtime of the simulation by checking how much of the initial energy is still stored in particles on the stack.

An arbitrary number of processes can be combined to constitute the *process sequence*.

5.6 Particle stack

The particle stack is the central object that manages the particles in memory. The current implementation uses a *structure of arrays*-like memory layout for the particles, meaning that each particle property is stored in a separate, contiguous array. This also means that no independent, compact ”particle object” that one could create anywhere exists. Instead,

a ”particle” is a mere reference to the actual data in the individual arrays. This reference object provides methods like `particle.getEnergy()`, `particle.setPID(Code)`, etc., so that the developer can be oblivious to the underlying memory layout.

The default particle properties stored on the stack are:

- particle ID (`Code`, integer)
- kinetic energy (`HEPEnergyType`)
- position (`Point`)
- direction vector (dimensionless, normalized `Vector`)
- time (`TimeType`)
- pointer to current volume
- nuclear isotope data (A, Z, integer)
- a boolean flag indicating whether a particle is deleted

To avoid the need to search the volume tree each time the current volume is needed, a pointer is stored. It is updated each time a boundary is crossed as described below. Nuclei are not fully specified by their particle ID, which only states that it is a nucleus (`Code::Nucleus`). Therefore, the A and Z values are stored in separate fields.

The *is-deleted* flag allows marking a particle at any position on the stack as ”removable”. Only whenever such a marked particle occurs and is read again on top of the stack it is really removed. This procedure is useful e.g. for thinning and similar filtering purposes in-place without the need for a (costly) reordering of the particles. It is also a cornerstone of the cascade history described in section 5.8.

5.7 Program flow

We are now in a position to discuss the how the individual building blocks gear into each other to process a full particle cascade. The basic principle is not different from the standard Monte Carlo cascade code: Particles from the stack are propagated and if they produce any secondaries these are pushed onto the stack. In CORSIKA 8 this main loop is implemented in the `Cascade` class, which has access to the stack, the environment and the process sequence. Furthermore, one may select a certain tracking algorithm, i.e. the procedure how the trajectory is determined. For propagation in magnetic fields this task is much more complex than without magnetic fields. A propagation step of a single particle consists of four parts: First, the trajectory is calculated, followed by the determination of the step-length to be taken. Then the particle is propagated by that length. Finally, the action corresponding to the chosen step-length is executed. Let us consider each of these substeps in more detail.

5.7.1 Trajectory determination

Without any Lorentz force the trajectory is a straight line (ray). The only actual calculation is that of its maximum length, which is determined by the intersection with volume boundaries. In the volume tree, only a subset of all existing volumes need to be considered: the current volume itself, its child volume and its excluded volumes. At the moment, the only implemented geometric primitive is a sphere. For the intersection of a straight line, which is parameterized with a real number ℓ such that $\ell = 0$ corresponds to the current position and $\ell > 0$ ($\ell < 0$) to points in front of (behind) the current position, with a sphere, a quadratic equation has to be solved, yielding up to two real solutions for ℓ. Some care must be taken to choose the correct one: Negative solutions are always excluded because they refer to the past. If both solutions are positive, the lower (greater) value has to be chosen when entering (exiting) a sphere. A complication arises when a particle was moved to the boundary in the previous step. Due to limited numerical precision the particle appears to be slightly before or after the boundary. When the intersection is recalculated in the current step and the particle appears to be still before the boundary, the same crossing can be proposed again, resulting in a stuck particle and an infinite loop. This situation is avoided by keeping a reference to the current node in memory and updating it during a boundary crossing. This way it is always clear if the considered volume is to be entered or exited.

The situation with magnetic fields is more involved. The required modifications have been studied in the bachelor thesis by Schmidt [147], using the leapfrog algorithm implemented in AIRES and described in ref. [148]. The update of the position $\boldsymbol{P}_0$ to $\boldsymbol{P}_1$ consists of two half steps, in between which the direction $\vec{u}$ is updated:

$$\boldsymbol{P}' = \boldsymbol{P}_0 + \frac{\ell}{2}\vec{u}_0, \tag{5.3}$$

$$\vec{u}_1 = \vec{u}_0 + \vec{u}_0 \times \ell \frac{q}{p}\vec{B}, \tag{5.4}$$

$$\boldsymbol{P}_1 = \boldsymbol{P}' + \frac{\ell}{2}\vec{u}_1. \tag{5.5}$$

If we consider the magnetic field $\vec{B}$ constant during the propagation (we set $\vec{B} = \vec{B}(\boldsymbol{P}_0)$) and do not normalize u_1, $P_1(\ell)$ is a parabola. The intersection with a sphere is described by a quartic equation, which is solved analytically, albeit with more computational effort. The right solution to select is the smallest positive value. Additionally, several safeguards are introduced to handle situations when the procedure turns out to be numerically fragile.

5.7.2 Step-length determination

After the trajectory has been proposed, it is determined how far the particle will be propagated. For this, a number of candidate step-lengths are taken into account, out of which the minimum is chosen:

- The maximum step-length to reach a volume boundary as described in the previous section.

- In case of magnetic fields, the angular deflection per step is limited to a pre-defined value, e.g. 0.1 rad.
- A candidate time of decay is sampled from the current (lab-frame) lifetime and converted to a length using the current velocity. The current lifetime is queried by summing the contributions (branching ratios) of all `DecayProcesses` in the process sequence:

$$\frac{1}{\tau_{\text{tot}}} = \sum_i \frac{1}{\tau^{(i)}}. \tag{5.6}$$

- A candidate interaction point is sampled: First, a grammage is sampled from the exponential distribution with the current interaction length λ_{int} as parameter, which is determined from the contributions of all (competing) `InteractionProcesses` in the process sequence:

$$\frac{1}{\lambda_{\text{int,tot}}} = \sum_i \frac{1}{\lambda_{\text{int}}^{(i)}}. \tag{5.7}$$

  For this, the individual `InteractionProcesses` themselves typically calculate the weighted sum of the cross-sections for each isotope of the medium in the current volume. The grammage is then converted to length, which is a function provided by the density model. Except for the trivial case of homogeneous density, the density models usually assume the trajectory to be straight.
- The minimum of the values obtained from the step-length limiting `ContinuousProcess`es, i.e. the return values of their `getMaxStepLength()` methods.

5.7.3 Propagation

After the actual step-length has been selected, the initially proposed trajectory is shortened to that length. Subsequently, continuous processes are applied to the particle along the trajectory, which may modify the particle properties. Also purely ”observing” continuous processes are possible. Examples are radio emission and recording the longitudinal profile.

Since energy losses can cause a particle to fall below the energy threshold, continuous processes state whether a particle is absorbed. In this case, it is deleted from the stack.

5.7.4 Final action

In case a particle survives the continuous processes, the action corresponding to the selected minimum step-length is performed:

- If the step-length was set by a step-length limitation of a continuous process, nothing happens. All steps described in sections 5.7.1 to 5.7.3 will be repeated.
- In case of an interaction, the interaction length is recalculated based on the particle properties after the propagation, denoted $\lambda'_{\text{int,tot}}$, which may be different from the value $\lambda_{\text{int,tot}}$ before the propagation that was also used for the sampling. One of the competing

`InteractionProcesses` needs to be selected to perform the interaction. This is done by randomly choosing the i-th process with probability $p_i = \max(\lambda_{\text{int,tot}}, \lambda'_{\text{int,tot}})/\lambda^{(i)}_{\text{int}}$. If $\lambda_{\text{int,tot}} < \lambda'_{\text{int,tot}}$, these probabilities do not sum up to one. The remaining fraction corresponds to a rejection of the interaction. As explained in section 4.4.1, the sampling in this case is still exact despite the change of $\lambda_{\text{int,tot}}$. If, on the other hand $\lambda_{\text{int,tot}} > \lambda'_{\text{int,tot}}$, the sampling is in principle not exact but for sufficiently small changes in energy the incurred error is negligible.

- For decays the procedure is analogous.
- In case of a volume boundary transition, the `doBoundaryCrossing()` methods of all `BoundaryCrossingProcesses` are called, in which the particle may be deleted. Additionally, the pointer to the current volume is updated to the new volume.

Interactions and decays create new secondaries, which are to be pushed onto the stack via the `SecondaryView` that offers an `addSecondary…()` method, which is to be called with the properties of the secondary. After it has been filled, it is passed subsequently to all `SecondariesProcesses`. These may iterate over the new secondaries and alter or even delete (more accurately, flag them as deleted) some of them again. Finally, the projectile particle, now no longer on top of the stack, is marked as deleted.

5.8 Cascade history

A special feature of CORSIKA 8 is the option to retain information about all predecessor generations. Such a feature was available in MOCCA and has been used by Hillas to study features of hadronic interactions with respect to their relevance for particle production in EAS [149]. His aim was to demonstrate the usefulness of the Hillas Splitting Algorithm despite its severely simplified nature. In contrast to MOCCA and somewhat surprisingly, none of the current generation of EAS simulation codes (AIRES, CONEX, CORSIKA) provide this feature to the required extent. CORSIKA [133] nevertheless allows accessing and storing the mother and grandmother particles of a ground-hitting particle. This tool has proven itself to be valuable in studying the production of muons in air showers [150]. Here, the muon's mother most often corresponds to a low-energy decaying meson, while the grandmother represents the projectile of the hadronic interaction in which that meson had been produced. It is nevertheless desirable to extend this genealogical information further since the only reason why only two preceding generations have been available is of purely technical nature: In the implementation in legacy CORSIKA, a record on the stack is extended to comprise not only the state of a single particle but also the states of its parent and grandparent.

In CORSIKA 8 we follow a different approach. When undergoing an interaction the projectile particle is not immediately removed from the stack but only maked as *deleted* just before the secondaries get pushed onto the stack. It is only absolutely removed as soon as the sub-cascade it initiates is completely processed, which is exactly the case when the deleted particle is the top particle of the stack for the second time.

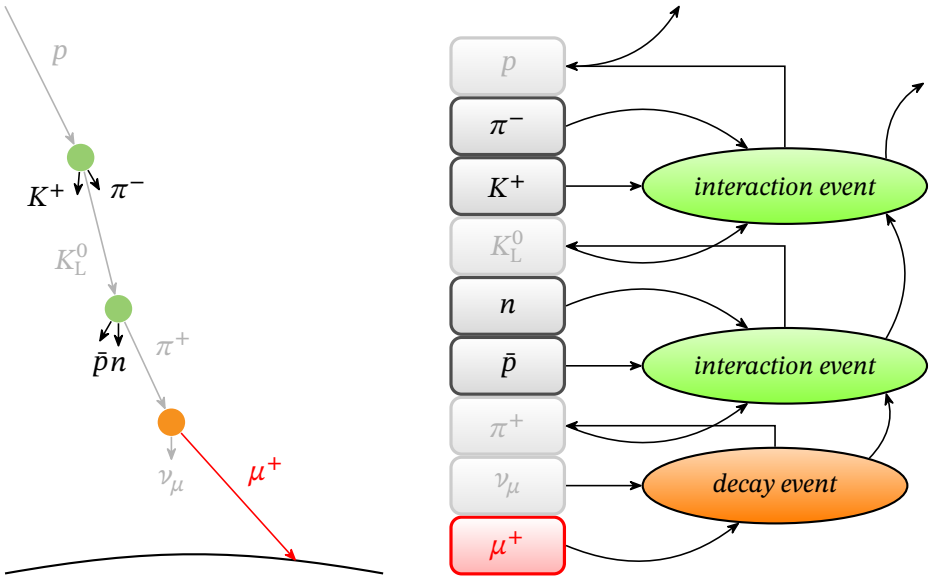


Figure 5.3: Snapshot of a stack with history information. Greyed-out particles represent deleted particles which still reside in memory. Red color indicates the particle currently being propagated.

In order to link particles with their parent generation, we introduce *event* objects and the particle object hold shared pointers to their event of creation, which in the case of the primary particle is just a null pointer. The event objects themselves keep a reference to the (deleted) projectile to establish the link to the parent particle. Furthermore each event maintains a copy of its secondaries that is filled at the time of their production and not modified afterwards. A particle on the stack is modified during propagation as its position and 4-momentum change. For certain studies however it is desirable to retain the particle state as it was produced. By keeping an additional, constant copy of the secondary in the event object the state at production is preserved and allows for relating this state to the state at reinteraction. It is foreseen that the event class can be subclassed so that specific types of events can store additional information, e.g. hadronic events might store data like impact parameter or number of wounded nucleons, that can be correlated with particle distributions at ground. An illustration of the design is given in fig. 5.3.

File size considerations

To estimate the impact of saving the complete lineage to disk we compare the particle output file sizes with and without inclusion of the lineage. For this experiment a simplified binary format is used in which a *plain particle record* consists only of the PDG code (4-byte integer) and the energy (8-byte double precision floating-point number). For the standard particle output, plain particle records are written one after the other without any additional header/-footer data. The total file size is given by the number of particles times the size of a plain

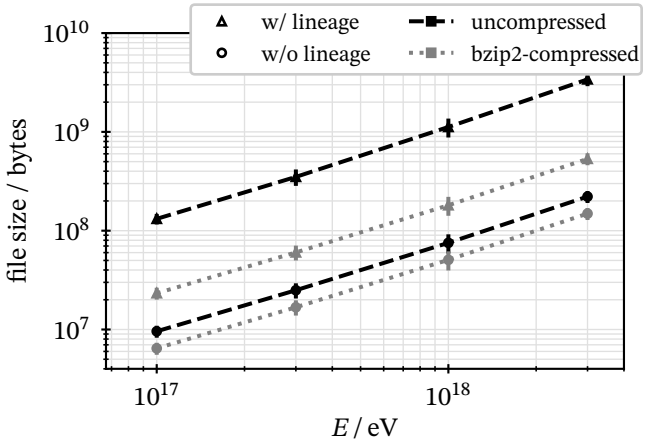


Figure 5.4: Average particle output file size

particle record (12 bytes).

By contrast, the enhanced data format that includes the lineage must preserve the jagged array structure: For a ground particle, a plain particle record is written for the particle itself, followed by plain particle records for each ancestor. This chain of ancestors is terminated by a trailing `0xFFFF FFFF`. Since this cannot be meaningfully interpreted as PDG code in this context[2], no ambiguity with another ancestor record arises. The file size increases by a factor given by the average number of ancestor generations plus the overhead of the trailing 4-byte sequence.

Clearly, the enhanced data format contains a high degree of redundancy since ground particles share parts of their lineages. The more ”ancient” any of the entries in a particular lineage is, the more likely it is to be found in the lineages of other ground particles as well. In particular, the primary particle is the final ancestor of every ground particle. It is therefore plausible that these redundancies lead to a substantial file size reduction when the file is compressed.

In fig. 5.4 the resulting file sizes are presented dependent on the primary energy, averaged over 20 runs. The error bars indicate the standard deviation of the file size distribution. The uncompressed enhanced data format requires about 15 times as much disk space as the standard data format. The number increases slightly with energy since the average number of generations grows logarithmically (see also section 7.1.2). Compressing the enhanced data format with bzip2 reduces the size by a factor of ~ 6. Finally we note that compression also reduces the standard-format file size by about 30 %, which can be attributed to the fact that only a few values out of the 4-byte range of the PDG code in fact occur.

[2]The PDG code -1, encoded as 4-byte integer `0xFFFF FFFF`, represents the $\bar{d}$ quark, which is not a possible projectile in CORSIKA 8.

5.9 Available modules

Although many important aspects are still being developed and optimized, CORSIKA 8 is already usable for a number of tasks, including some specialized tasks that have not been possible with any existing codes. In the following, a summary of currently available modules and features is given.

High-energy hadronic interactions are possible with SIBYLL 2.3d and QGSJet-II.04. Low energy hadronic interactions can be performed with UrQMD 1.3cr, for which the interface was adapted from CORSIKA 7, although it should not be considered usable for production yet. Additionally, a link to the HSA routines of AIRES is available. For both low energy models the tabulated cross-sections from the UrQMD distribution in CORSIKA 7 are used, which contain data for nucleon, pion and kaon projectiles on nitrogen, oxygen and argon targets.

Decays are handled by Pythia 8 or, alternatively, by SIBYLL. As SIBYLL produces excited states which are not (yet) covered by the PDG and whose decay cannot be treated by Pythia, it makes sense to combine both, so that common particles decay in Pythia and the other ones in SIBYLL.

For the handling of EM interactions and energy losses the PROPOSAL code [122, 151] has recently been integrated into CORSIKA 8 [152, 153]. Its validation is currently work in progress. Hybrid simulations are possible with a link to CONEX: Whenever an EM particle is produced as secondary, it is immediately fed into CONEX where it is filled into tables representing sources in the CE. After the Monte Carlo part of hadrons and muons has been completely processed, CONEX solves the CE. The result are mainly the EM longitudinal profiles but also of muons and hadrons stemming from photohadronic interactions. Moreover, a module for energy losses of muons and other heavy charged particles is available that implements Bethe-Bloch losses with Sternheimer correction, parameterized for air and other materials.

A module for radio emission that incorporates both the endpoint and ZHS formalisms is under development and first results are presented in ref. [154]. Two independent modules for Cherenkov radiation are actively being developed and tested, one targeting GPUs [155], while the other aims to maximize performance on CPUs by vectorization [156].

For visualization purposes a *track writer* module allows saving the coordinates of trajectory starting and end points together with the particle ID.

With the *observation plane* module particles that cross a user-defined plane (which can have arbitrary orientation) are written to an output file. The observation plane may be opaque or transparent, meaning that particles are deleted or not when crossing the plane. Multiple observation planes can be present at the same time, in which case all but the last one are transparent. The observation plane is implemented as a `ContinuousProcess` due to its feature to limit the step-length. An obseration level that respects the curvature of the Earth is currently not implemented although it would be straight-forward to do so as a `BoundaryCrossingProcess`.

With the flexibility of CORSIKA it is often only a small effort to implement wrapper processes that add generic functionality to existing processes. Three of these have been

developed for this thesis:

SwitchProcess Conditionally selects between two alternative processes. The condition is user-definable and depends on the current particle. The main application is to switch between low and high energy hadronic interaction model based on the particle energy.

InteractionCounter Wraps an `InteractionProcess` and creates a histogram of the number of generated interactions binned in projectile energy and species.

InteractionLengthModifier A wrapper around an `InteractionProcess` whose interaction length is modified by applying a user-definable function.

5.10 Output format

Various kinds of data are produced by CORSIKA 8, ranging from small data sets like longitudinal profiles to potentially huge amounts of particle data from the observation plane. At the same time, the typical use cases are equally diverse: In the UHECR community, usually one shower per run is generated with particle file sizes of $\gtrsim 100\,\mathrm{MB}$. On the other hand, simulations for gamma ray observatories are most often performed with a high number of showers, each containing only a small number of particles.

For CORSIKA 8 several alternatives (HDF5 [157, 158], ROOT [159], ExDir [160] and others) have been considered and discussed. The final solution is based on the *Apache Parquet* [161, 162] format, which is a binary, columnar storage format. These files containing the raw data are supplemented by human readable YAML files for meta-data. Each module that creates output writes an independent file, one for each shower of a run, all organized in a hierarchial directory structure. A Python module is shipped together with CORSIKA 8 to read and analyze the generated datasets for convenience. For many other programming languages, readers for parquet files are available as well.

For histograms, CORSIKA 8 integrates the *Boost.Histogram* [163] library, which is extensively used for the work presented in this thesis. Boost.histogram facilitates the creation and handling of histograms with a great degree of flexibility and comfort: Histograms may have an arbitrary number of dimensions (”axes”). Axes themselves can consist of classical real-value intervals with fixed or variable bin widths, dense or sparse integer values and others. Boost.Histogram itself does not include a data format to save the histograms to disk, though. For this thesis, a simple data format based on NumPy [164] ”.npz” files has been implemented. This is complemented by a Python module that allows to read these files, returning an `Histogram` object of the Python bindings of Boost.Histogram.

5.11 Default settings

While CORSIKA 8 is a framework whose parts one can use for many purposes, some examples are included in the codebase that replicate complete air shower simulations as far as possible. Some default conventions used in these examples are:

- The atmospheric model is that of Linsley, consisting of five layers. They are implemented as concentric spheres (”shells”) with the origin of the root coordinate system as center, each having either a spherical exponential (the inner four layers) or constant (the outermost layer) density model. The radii are $R_\mathrm{E} + h_i$, with $R_\mathrm{E} = 6371\,\mathrm{km}$ denoting the Earth radius and h_i the heights of the boundaries above sea level (a.s.l.).
- The nuclear composition of the atmosphere is set to 78.47 % nitrogen and 21.53 % oxygen.
- The shower core is fixed to be at $(0, 0, R_\mathrm{E} + h_\mathrm{obs})$ in the root CS, where h_obs is the height of the observation level.
- The primary particle is injected at the top of the atmosphere (height h_5 of the outmost layer). The position on the surface is chosen such that the momentum vector $(-p \sin\vartheta, 0, -p\cos\vartheta)$ (p is the magnitude) points towards the shower core.
- Time is measured from the injection, i.e. the primary particle starts with time $t = 0$.

6 Comparisons of hadronic cascade with other codes

In this chapter we study the performance of CORSIKA 8 in comparison to other up-to-date air shower simulation codes. We focus on the hadronic and muonic components of UHECR showers. This exercise does not only serve to validate CORSIKA 8 but also as a cross-validation of all the codes. Furthermore, we will point out systematic differences that are particlarly relevant in the light of the muon deficit that require further investigation.

Besides CORSIKA 8, we consider CORSIKA 7.7410 [165] (with the `CURVED` flag), CONEX (as part of CORSIKA), and AIRES 19.04.06. Additionally, MCEq 1.2.2 in "air shower configuration", i.e. with a $\delta(E - E_0)$-shaped primary spectrum, is taken into account where possible. Electromagnetic interactions are disabled in CORSIKA 7 and CONEX via the `ELMFLG` steering option. In AIRES the cutoff energies for electromagnetic particles are set to 99.9 % of the primary energy, effectively disabling the whole EM cascade. The geomagnetic field is not considered.

We consider proton-induced vertical showers at 10^{18} eV. The model transition energy of the Monte Carlo codes is set to 63.1 GeV, which is the lowest value laying on a bin edge of a logarithmically scaled energy axis with five bins per decade that all high-energy models can handle safely. The observation level is set to 1400 m a.s.l., correspondging to 875 g cm^{-2} in Linsley's atmospheric model. We study two cases: First, we use QGSJet-II.04 and consider only the high-energy part, i.e. the energy cut is set to 63.1 GeV as well. Second, we use SIBYLL 2.3d for high energies supplemented by a low-energy interaction model for energies down to 1 GeV. In CORSIKA 7 and CONEX, UrQMD 1.3 is used. In AIRES the only available option is the HSA. MCEq uses UrQMD 3.4. With CORSIKA 8 we consider both UrQMD 1.3 and the HSA implementation from AIRES, linked to CORSIKA 8. 500 showers are simulated in each configuration except CORSIKA 7, where 5000 showers are simulated in order to serve as baseline with small statistical fluctuations. All energy values refer to kinetic energy.

Preliminary – and by now already partially outdated – results have been presented in refs. [166, 167]. The results shown here represent only the latest status after a continued effort to identify, understand and reduce differences. A relevant outcome of this is the identification of the *hyperon bug* outlined in section 6.4.

6.1 Interaction spectrum

A useful observable is the *interaction spectrum*, i.e. the number of (hadronic) interactions by projectile energy and species. It is related to the corresponding energy spectra by

$$\frac{\mathrm{d}N_{\mathrm{int}}}{\mathrm{d}E} = \frac{1}{\lambda_{\mathrm{int}}(E)} \int \mathrm{d}X \frac{\mathrm{d}N}{\mathrm{d}E}(E, X), \tag{6.1}$$

as can be inferred from the CE eq. (3.1). The interaction spectrum of a given particle species is mainly influenced by the multiplicity of this species as secondaries in hadronic interactions, as well as its critical energy. We will return the the physics interpretation in chapter 7.

The interaction spectrum allows studying rather rare and short-lived particles like hyperons, which is hardly possible with reasonable statistics when considering only the energy spectra at a certain shower depth.

The spectra are obtained from the Monte Carlo codes by filling corresponding histograms during the run. In CORSIKA 7 and CONEX this is done via the COAST interface, which provide a hook for custom code to be called just before an interaction with access to the projectile properties. In CORSIKA 8, the `InteractionCounter` process is used for the purpose. In MCEq the interaction spectra are obtained by integrating the energy spectra according to eq. (6.1) numerically on the X grid that MCEq sets up by itself.

The results for p, n, π^+, K^+ and Λ are presented in figs. 6.1 to 6.6 for QGSJet-II.04. Agreement between all codes on a percent-level is found for n, π^+ and K^0_L. In the K^+ spectra, however, only the CORSIKA 7 and CONEX lines match, while MCEq and CORSIKA 8 show differences up to $\sim 10\,\%$. The bigger discrepancies for p and Λ are due to the fact that the latter are not allowed as projectiles in QGSJet so that an interaction with a proton projectile is generated instead. The counting with COAST happens after this particle ID conversion, leading to a mislabeling of Λ as p. In all species, CORSIKA 7 and CONEX exhibit the same behaviour. In MCEq particles are tracked only above their critical energy, which is why their spectra are cut off close to the peak of their distributions. The spectra of $\bar{p}$, $\bar{n}$, π^- and $\bar{\Lambda}$, which are not shown here explicitly, do not differ from those of their antiparticles.

Figures 6.7 to 6.12 show the interaction spectra for SIBYLL 2.3d. Here, the largest differences in the high-energy regime are found in kaons: CORSIKA 8 and CONEX yield the same spectra while CORSIKA 7 and MCEq predict almost the same shapes but different normalizations. Since in SIBYLL also the hyperons are valid projectiles, no ambiguities arise because of ID conversions.

In the low-energy regime, the situation is more complicated. We observe that despite the same model (UrQMD 1.3) CORSIKA 8 displays a higher interaction rate of nucleons than CORSIKA 7 and CONEX. With the HSA, the interaction rate is even higher than with UrQMD by up to 40 %. Discrepancies of similar magnitude are also found in the kaon sector. Here, however, the agreement of CORSIKA 8 with the other codes is better with HSA than with UrQMD. Strikingly, the K^+ spectra of CORSIKA 7 and CONEX diverge significantly below 7 GeV, even though the interface codes to UrQMD are identical. Low energy Λ particles in general interact only in CORSIKA 7 and CONEX.

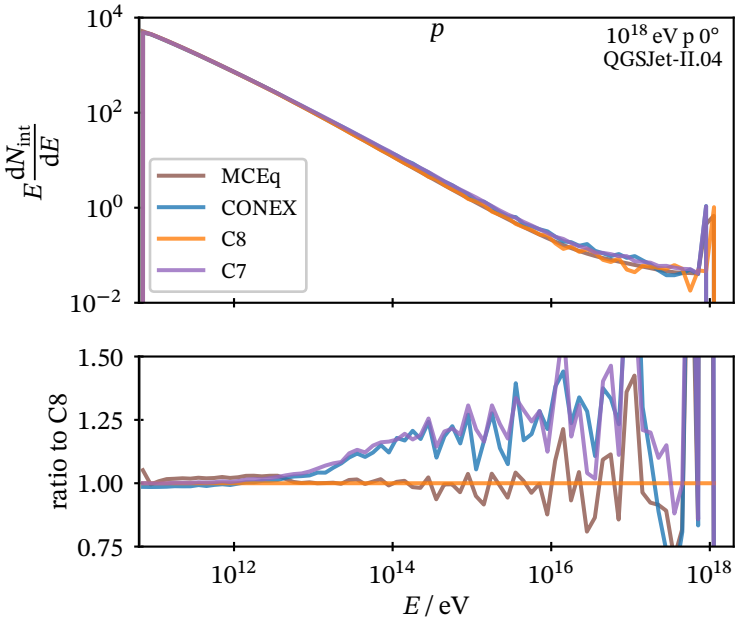


Figure 6.1: Interaction spectra of p with QGSJet-II.04

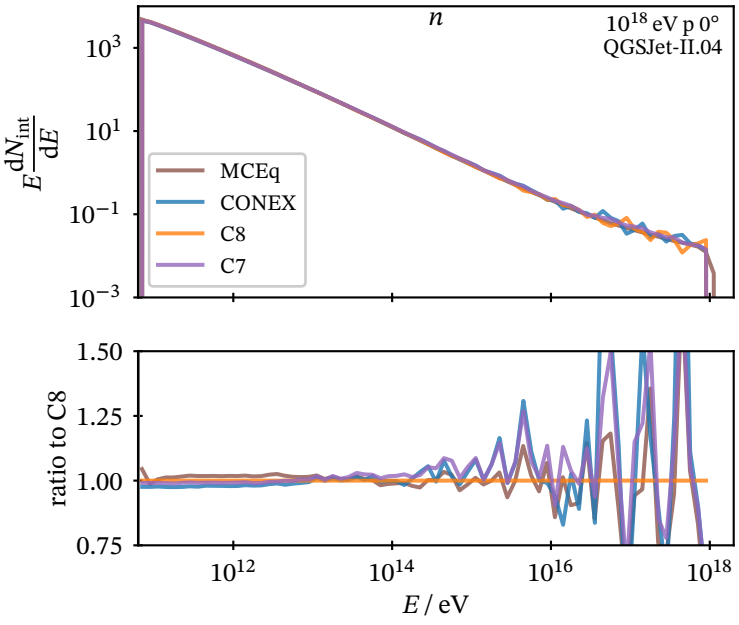


Figure 6.2: Interaction spectra of n with QGSJet-II.04

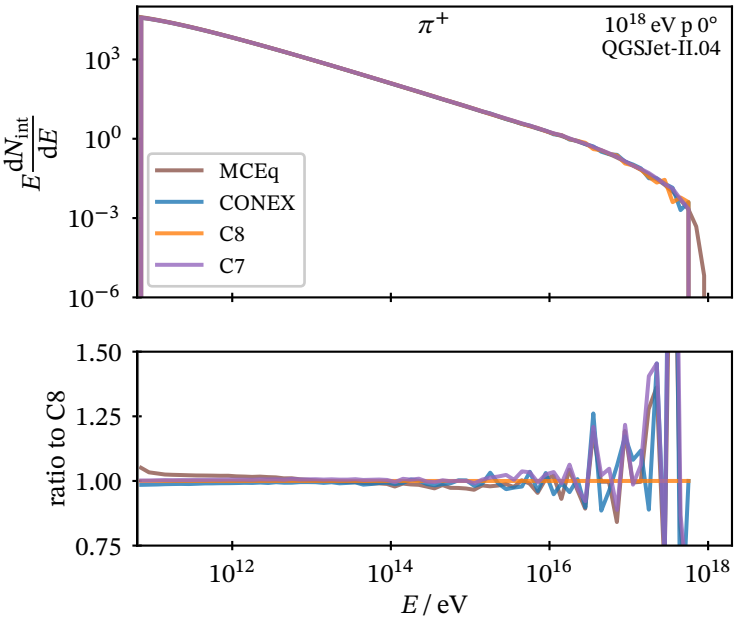


Figure 6.3: Interaction spectra of π^+ with QGSJet-II.04

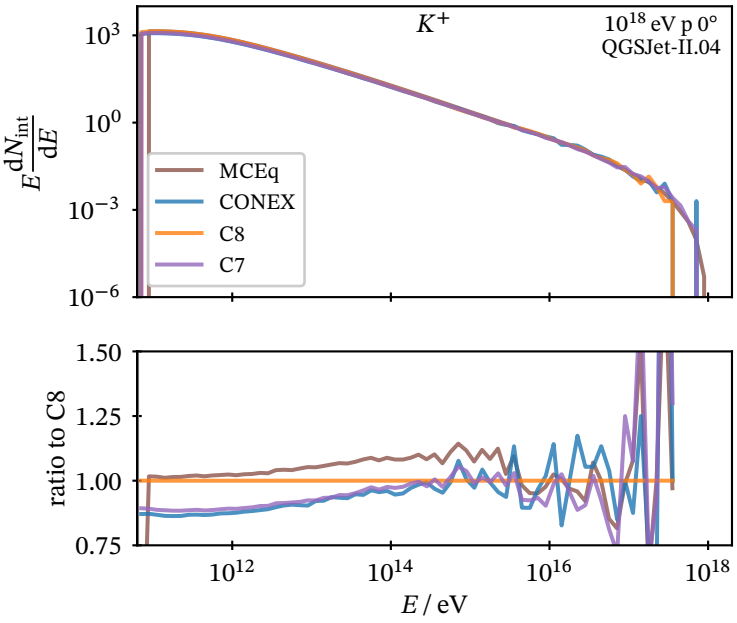


Figure 6.4: Interaction spectra of K^+ with QGSJet-II.04

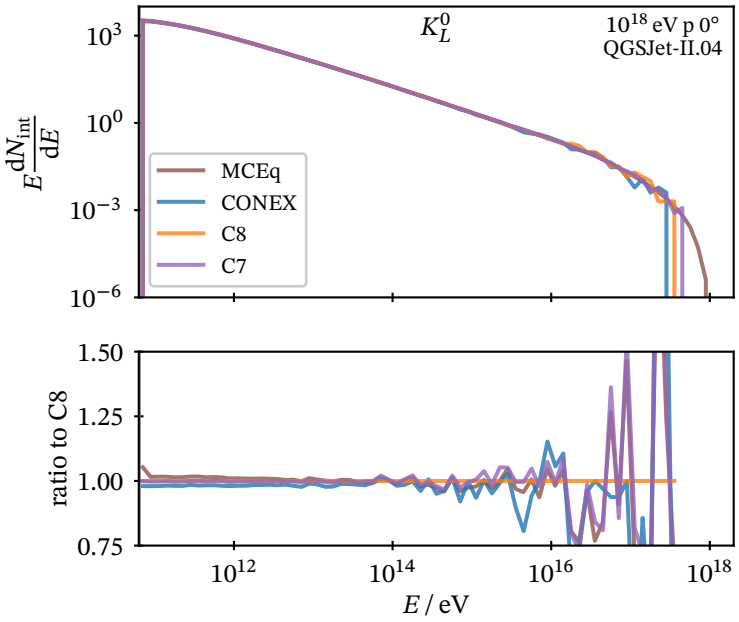


Figure 6.5: Interaction spectra of K_L^0 with QGSJet-II.04

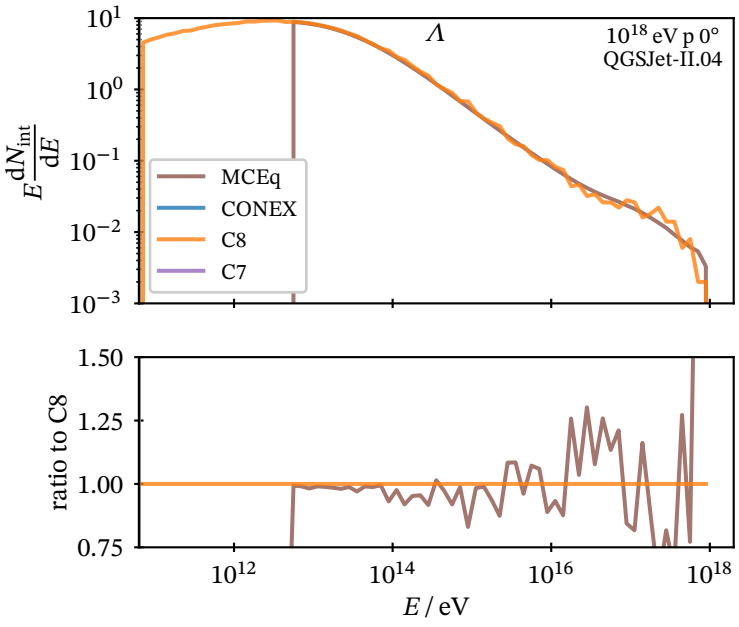


Figure 6.6: Interaction spectra of Λ with QGSJet-II.04

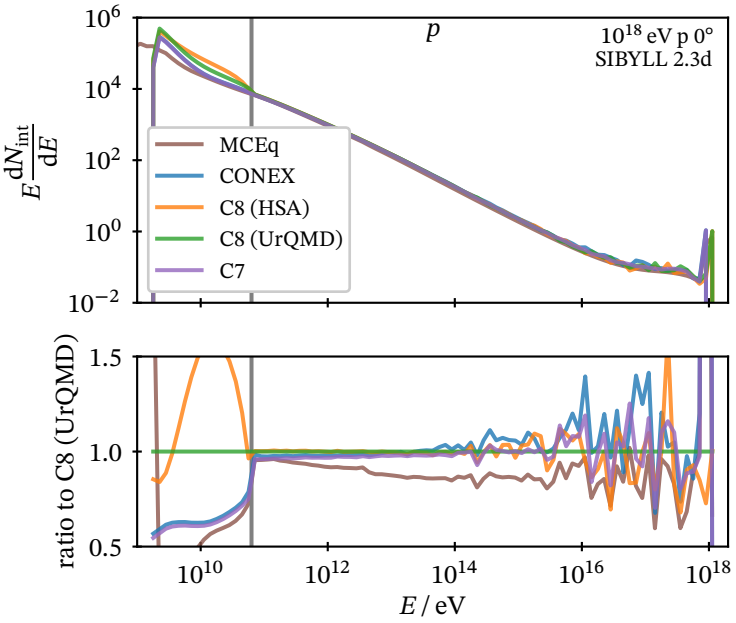


Figure 6.7: Interaction spectra of p with SIBYLL 2.3d

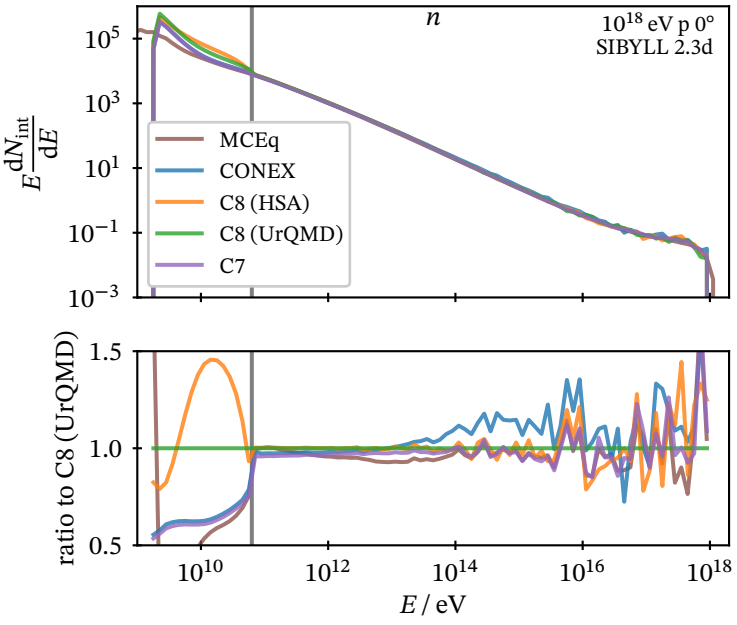


Figure 6.8: Interaction spectra of n with SIBYLL 2.3d

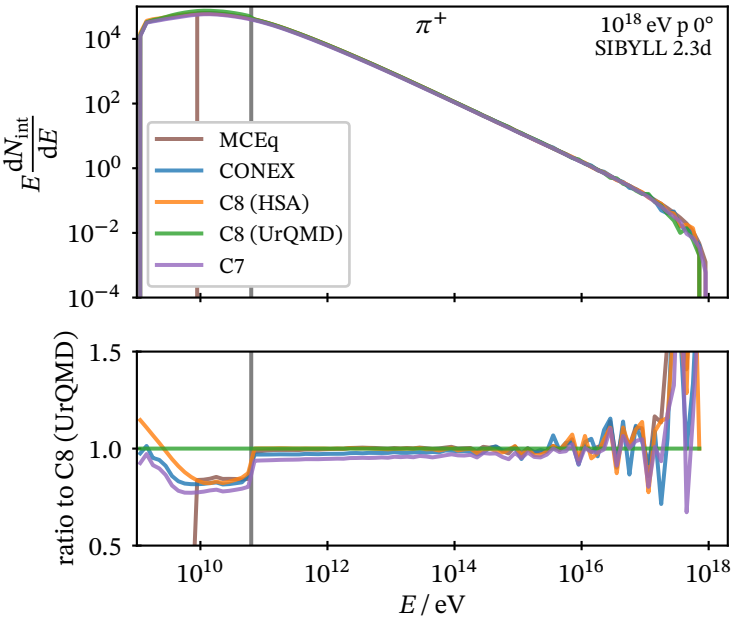


Figure 6.9: Interaction spectra of π^+ with SIBYLL 2.3d

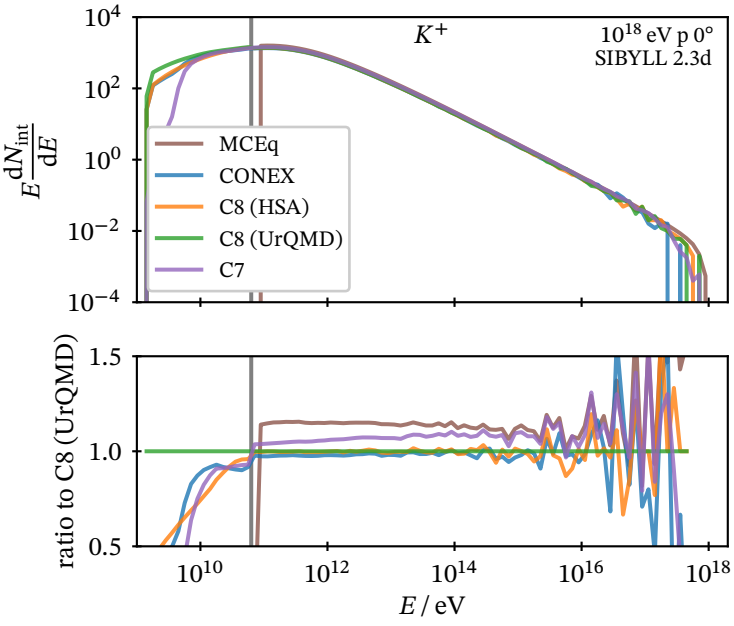


Figure 6.10: Interaction spectra of K^+ with SIBYLL 2.3d

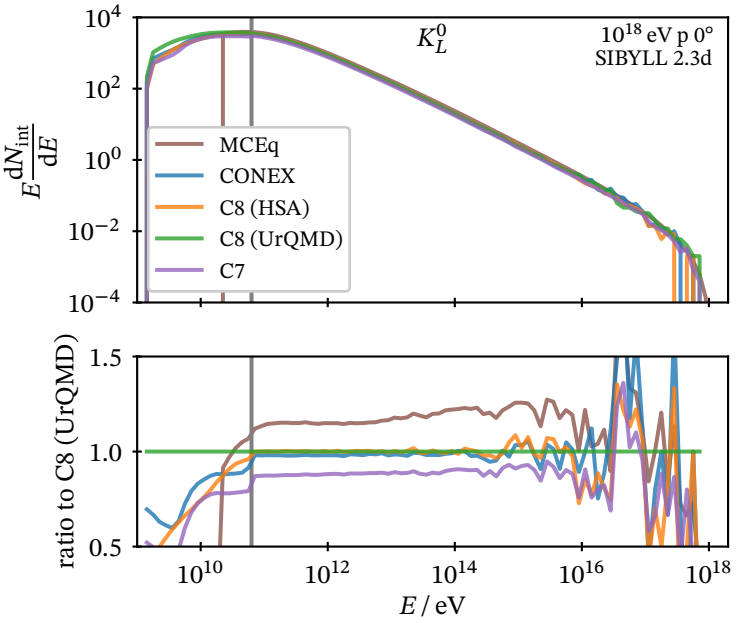


Figure 6.11: Interaction spectra of K_L^0 with SIBYLL 2.3d

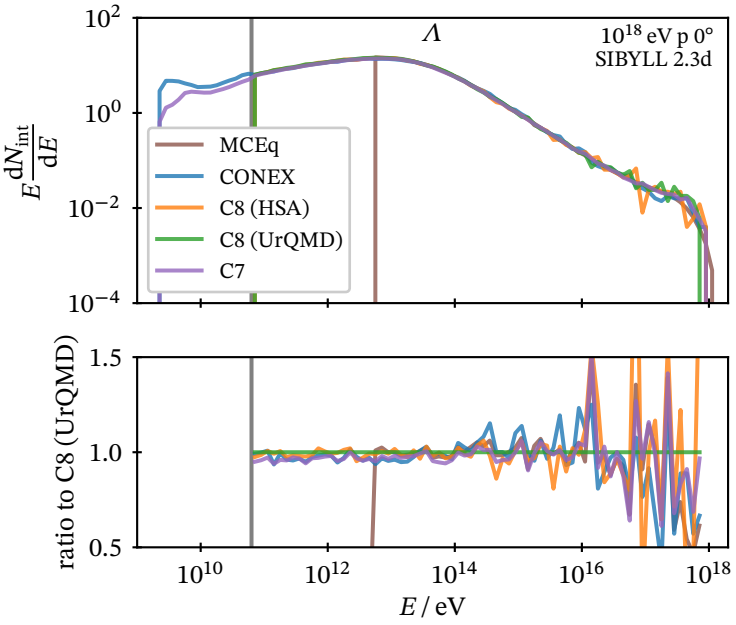


Figure 6.12: Interaction spectra of Λ with SIBYLL 2.3d

6.2 Muon energy spectra

The the muon energy spectra at ground for QGSJet-II.04 and SIBYLL 2.3d are shown in Figures 6.13 and 6.14, respectively. With QGSJet-II.04, CORSIKA 8, CORSIKA 7 and CONEX show virtually coinciding spectra below 10 TeV, the spectra of AIRES and MCEq both yield spectra which are offset by 5 %. Above that energy, CORSIKA 7 shows a softening not visible in the other spectra.

With SIBYLL 2.3d, the spread in the high-energy regime amounts to $\sim 15\,\%$ and no two codes yield coinciding spectra. In the low-energy regime, several points are noteworthy: Comparing CORSIKA 7 with AIRES, the difference does not widen up further despite the different low-energy models. On the other hand, the spectra calculated with HSA and UrQMD in CORSIKA 8 diverge quickly below the transition energy. Furthermore, MCEq features a sudden increase below 10 GeV, which coincides with the cutoff in the pion interaction spectrum and might be caused by the pions present at that energy suddenly forced to decay.

6.3 Lateral distribution

The lateral distribution of muons $\rho_\mu(r) = \mathrm{d}N_\mu/\mathrm{d}r/(2\pi r)$ as a function of radial distance r to the shower core, are shown in figs. 6.15 and 6.16 for QGSJet-II.04 and SIBYLL 2.3d, respectively. Considering only the high-energy muons above 63.1 GeV, the agreement is generally good up to around 1 km. With QGSJet-II.04 the differences are on a percent level, with SIBYLL 2.3d around 10 %. Above 1 km CORSIKA 7 predicts a higher density than the other codes.

Taking also the low-energy muons down to 1 GeV into account (fig. 6.16 left), the agreement stays on the same level in the < 1 km range. It is noteworthy, however, that above that distance the behaviour completely changes. Both AIRES and CORSIKA 8 with HSA yield a higher muon density than CORSIKA 7, while CORSIKA 8 with UrQMD shows a drop. At these distances, virtually all muons are of low energy, so that discrepancies observed also in the energy spectrum are particularly pronounced. Keeping in mind that the Pierre Auger Observatory is particularly sensitive to the lateral distribution at the km level, this underlines the importance of low-energy interactions.

This is further studied in fig. 6.17, where the two-dimensional distributions $\mathrm{d}^2N_\mu/\,\mathrm{d}E\,\mathrm{d}r$ of each code are compared to those of CORSIKA 7 by their ratio. Large deviations from one by more than 50 % are found especially in the upper left corners, i.e. at large distances and low energies, for all codes except CONEX.

6.4 The hyperon bug

During the course of this work an issue in CORSIKA 7.71 was found when inspecting the interaction histograms for hyperons, in particular those of Σ^0. The decay length of Σ^0, $c\tau = 2.2 \times 10^{-11}$ m, is a factor of 1000 times shorter than that of π^0, so that interactions are extremely unlikely. Figure 6.18 however shows that in v7.71 interactions of Σ^0 happen

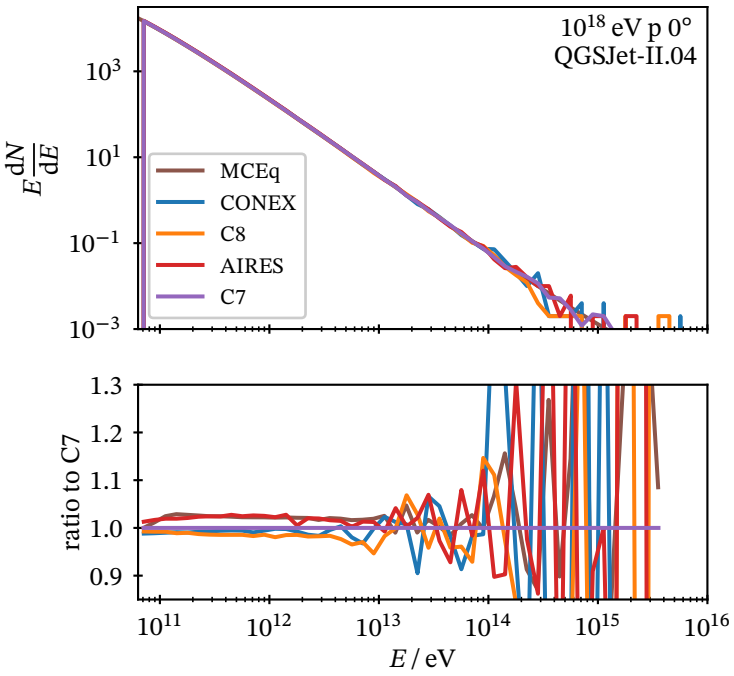


Figure 6.13: Muon energy spectra obtained with QGSJet-II.04

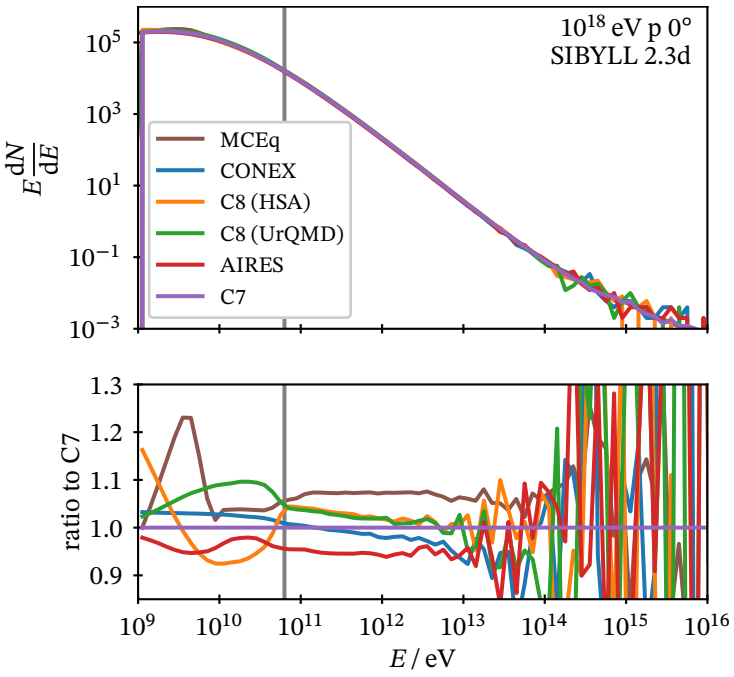


Figure 6.14: Muon energy spectra obtained with SIBYLL 2.3d

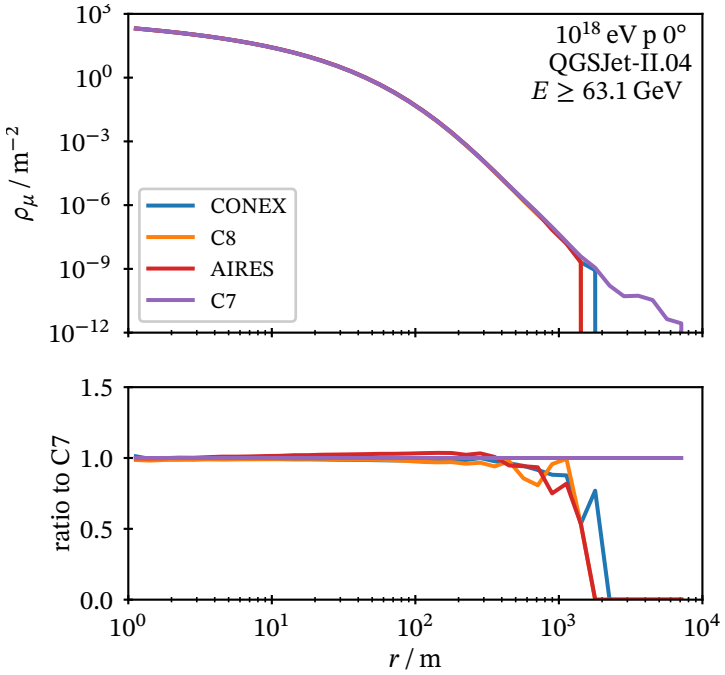


Figure 6.15: Muon lateral distribution obtained with QGSJet-II.04

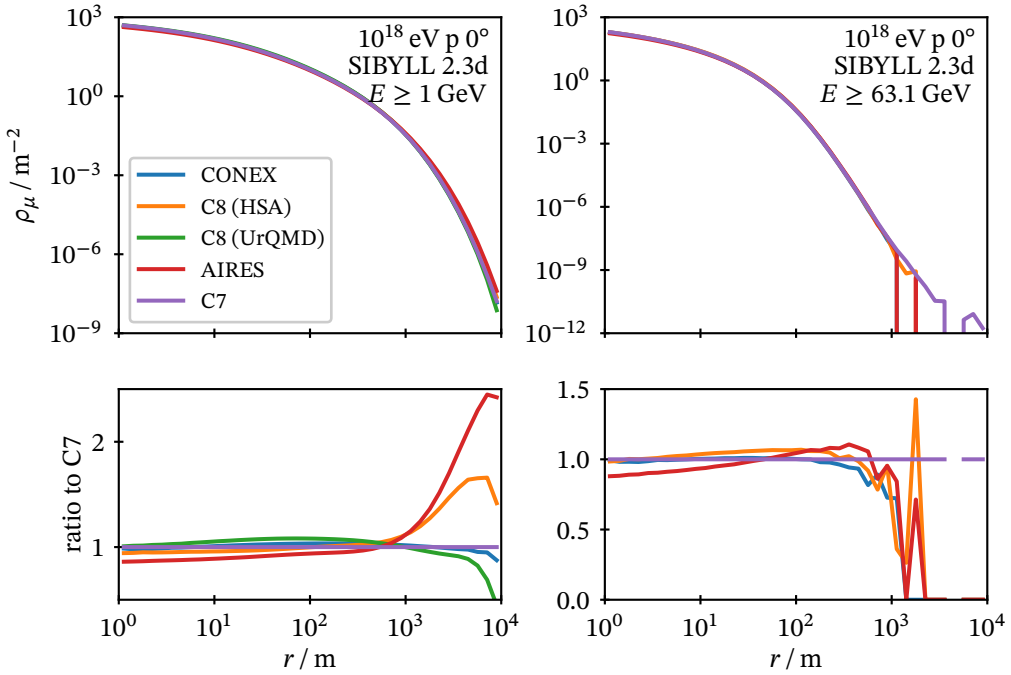
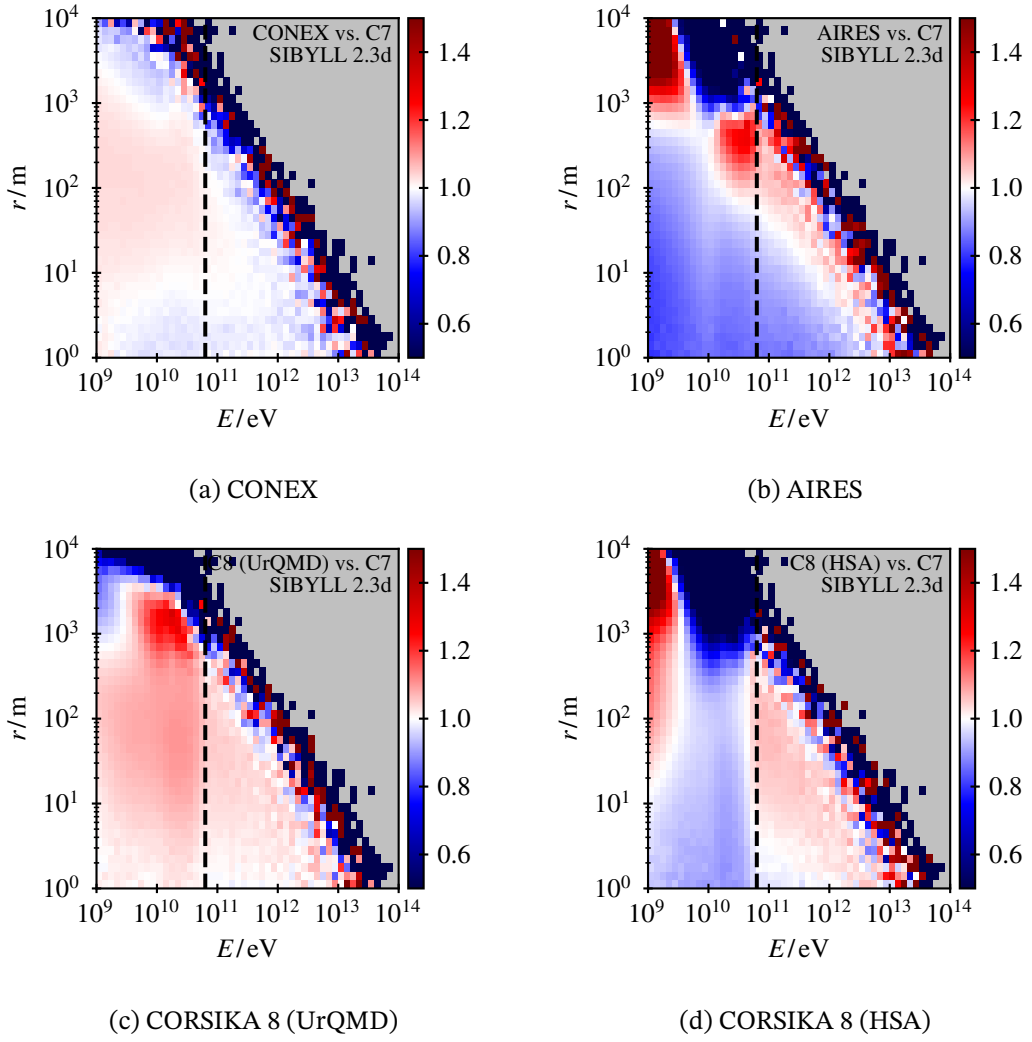


Figure 6.16: Muon lateral distribution obtained with SIBYLL 2.3d



(a) CONEX

(b) AIRES

(c) CORSIKA 8 (UrQMD)

(d) CORSIKA 8 (HSA)

Figure 6.17: Ratios of $\mathrm{d}^2 N_\mu / \mathrm{d}E\,\mathrm{d}r$ distributions for vertial proton showers at 10^{18}eV. The range of values is clipped to $[0.5, 1.5]$.

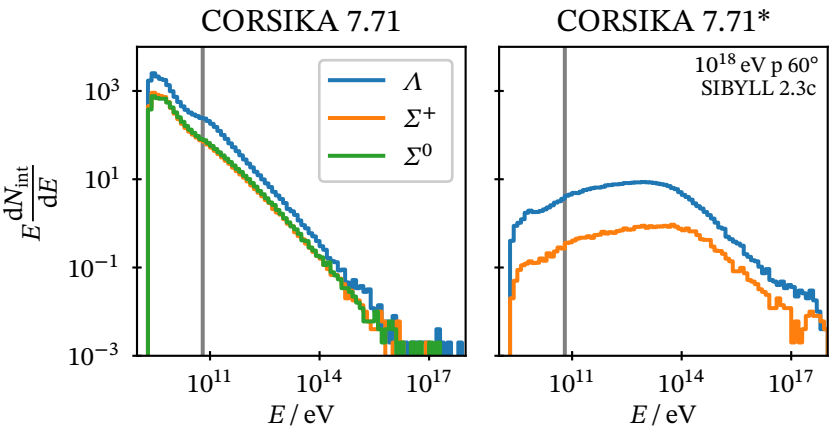


Figure 6.18: Interaction spectrum of hyperons in CORSIKA 7.71 before (left) and after (right) the bugfix

frequently, which eventually could be attributed to a bug that prevented hyperons from decaying when using EPOS-LHC or SIBYLL. The bugfix (commit `e7b96b02`, included in v7.74) re-enabled the decay, which completely prohibits interactions of Σ^0 and suppresses interactions of the longer lived hyperons below 10^{14} eV.

The effect on hadron and muon longitudinal profiles is shown in fig. 6.19. The maxima of the hadron profile differ by 6.3(34) g cm^{-2}. For the muon longitudinal profiles the shift is 9(4) g cm^{-2}. In both cases the maxima were obtained shower by shower by fitting a parabola in an interval ± 50 g cm^{-2} around the highest bin value. The height of the first interaction is fixed to 26 km (44 g cm^{-2}) to reduce statistical fluctuations.

Figure 6.20 displays the ratio of the muon energy spectrum. The bug causes a small softening up to 2 %.

Overall it can be concluded that the bug has only a small influence on shower observables (and therefore analyses carried out with affected versions) much smaller than any experimental uncertainties, yet it has improved the agreement with other codes.

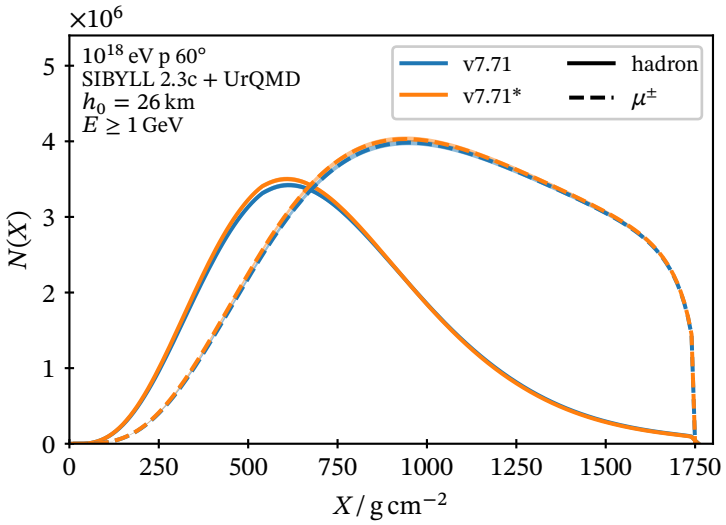


Figure 6.19: Longitudinal profiles of hadrons and muons before (v7.71) and after (v7.71*) the bugfix

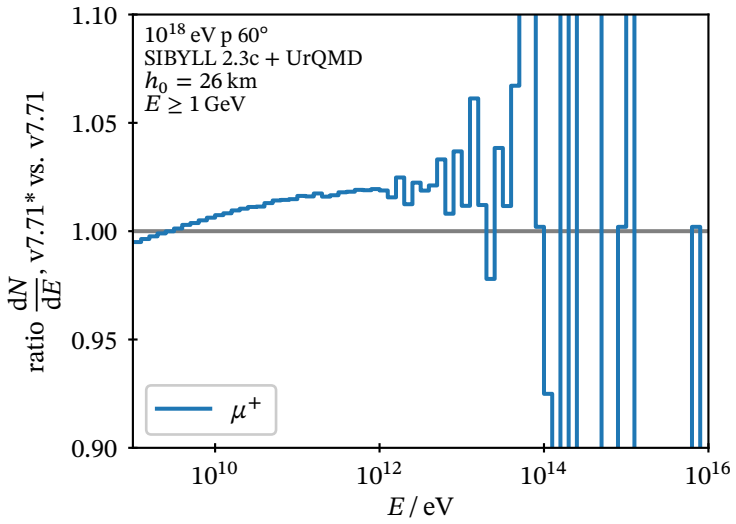


Figure 6.20: Ratio of muon energy spectra before (v7.71) vs. after (v7.71*) the bugfix

6.5 Conclusions

The overall picture shows that despite equal hadronic models, simulations with different codes do not result in equal predictions regarding typical muon observables. At high energies, possible causes include the handling of short-lived resonances, which is of relevance in particular for SIBYLL. These can be set do decay either inside SIBYLL during the event generation or be handed over to the shower simulation code as secondary. In the latter case, the ID conversion must be done carefully. Additionally, the implemented decay channels may not match.

Another issue can arise when a particle species produced as secondaries of an interaction model is long-lived but not a valid projectile for that model. In this case, an interaction event is generated with a different projectile species instead that resembles the original one as close as possible by conserving as many quantum numbers as possible, although some freedom exists. In QGSJet-II.04 this happens with $\Lambda/\bar{\Lambda}$, which are usually replaced by $n/\bar{n}$. In SIBYLL 2.3d such ambiguities do not appear since hyperon projectiles are allowed.

Regarding the situation at energies below the transition threshold, further studies are needed in particular to pinpoint the differences observed betwen CORSIKA 8 and CORSIKA 7/CONEX: The interaction rates of $\sim$ GeV nucleons and mesons with UrQMD in CORSIKA 8 are significantly higher than in CORSIKA 7 or CONEX, which has an imprint also on the muon energy spectrum. It is at present still unknown why UrQMD behaves so differently in CORSIKA 8.

At the same time, it is striking to see large differences between AIRES and CORSIKA 8 with HSA. Especially in the radial range relevant for Auger, large differences are observed in the muon lateral distribution, also compared to UrQMD.

In summary, we can conclude that regarding the high-energy interaction models CORSIKA 8 yields results that lie well within the spread between other established shower simulation codes. Until better understood, such differences should be treated as systematic uncertainty besides the choice of the interaction model itself. The comparative studies presented here have already led to bugfixes and modifications of several of the codes and a further convergence may be achieved in the future. Such studies are in particular called for in the light of the muon deficit.

A number of developments in low-energy models will have an impact if integrated into CORSIKA 8: Several updates of UrQMD have been released since version 1.3 was adapted for usage in air shower simulations. Furthermore, the SMASH model [168], which can to some degree be considered a successor of UrQMD, will provide more insight into microscopic effects of collisions. An interesting alternative will be the upcoming version of PYTHIA 8, which is able to treat also low-energy interactions [42]. Also a future version of EPOS will allow treating interactions down to the lowest energies.

7 Studies on hadronic interactions and muon production

In this chapter we will explore hadronic interactions of UHECR air showers through Monte Carlo simulation studies, with a focus on muon production in the light of the muon deficit. The first part of the chapter devoted to a study of the phase-space of hadronic interactions relevant to muon production. In the second and third parts we explore the implications on air shower phenomenology when applying different types of ad-hoc modifications to hadronic interactions where these are not well constrained by data from accelerator measurements.

7.1 Quantifying the importance of phase-space for muon production

Hadronic interaction models rely heavily on data from accelerator measurements for tuning and validation. The interactions happening in EAS however cover a huge phase space in terms of projectile species, $\sqrt{s}$, and kinematic distributions of their produced secondaries. In the light of the muon deficit it is desirable to quantify the importance or relevance of different phase-space regions: On the one hand, this helps to determine which features of hadronic interaction models require more attention and are good candidates for tweaking to enhance the muon production. On the other hand, such information can be compared with coverage by accelerator experiments to point out where additional measurements may be of use.

An early study has already been conducted by Hillas [149] in 1997 with the computational resources and tools (the EAS simulation code MOCCA) available. His goal at that time was to justify the use of his splitting algorithm against criticism of being to too oversimplified. A crucial ingredient of this study was the ability to record and inspect the lineage of particles reaching ground, which in turn allows relating a particle with its progeny. Here, we follow a similar approach, making use of the particle lineage in CORSIKA 8 described in section 5.8.

Another study by Meurer et al. [150] focussed on the phase-space of the *final* hadronic interaction, i.e. whose secondaries decay into mesons, of PeV showers.

A large fraction of this section has been published in ref. [169].

7.1.1 Interaction spectrum

Figure 7.1 shows the interaction spectrum of a 10^{19} eV proton shower of 60° zenith angle, grouped by several classes of hadrons. Going from high to low energies, we first observe a peak at the primary energy, which in the limit of infinitesimal bin widths would be a delta

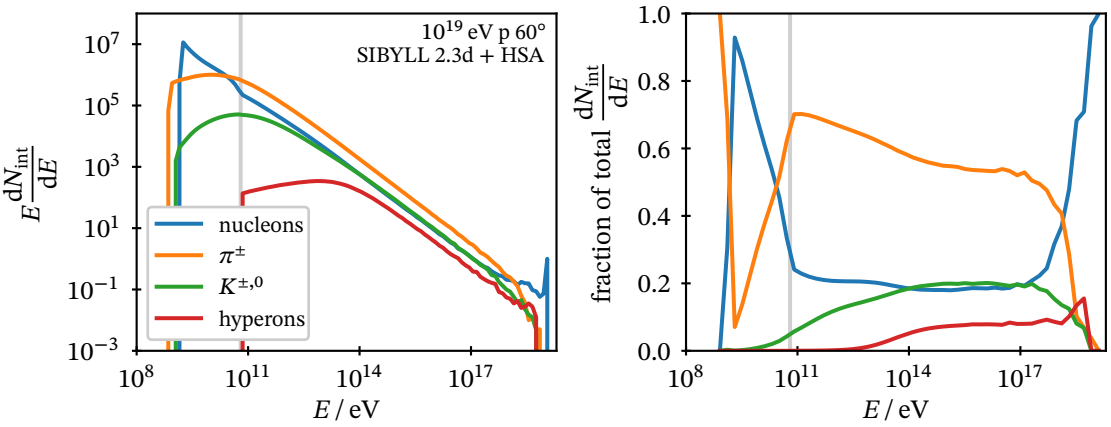


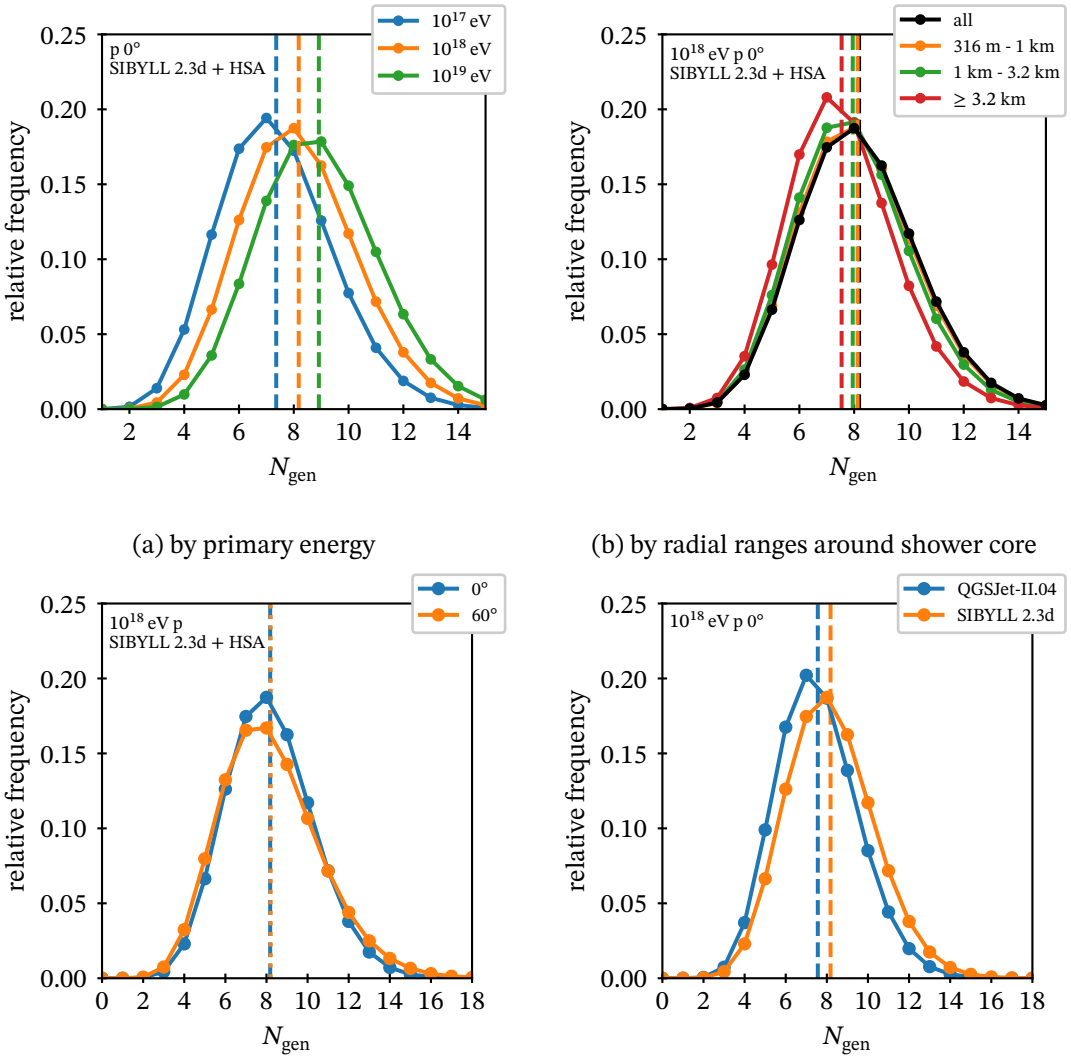
Figure 7.1: Number of hadronic interactions by energy and species. The grey vertical line indicates the transition between low- and high-energy interaction models.

function. For about one decade in energy below the primary energy, most interactions are those of nucleons, which can be attributed to the leading nucleons of the primary interaction. Below the crossing point, pion-air interactions are the dominating component, making up 50 % to 70 % of all hadronic interactions. Between 10^{14} eV to 10^{17} eV kaons and nucleons contribute equally to the total interaction spectrum with about 20 % each. Below 10^{14} eV, the kaon contribution decreases due to the K^0_S starting to predominantly decay instead of interacting. At that point also hyperons (mainly $\Lambda/\bar{\Lambda}$, the most long-lived hyperon), which are generally rare and rinteract only to a minor extent at all, fade away almost entirely. For a wide range in energy the total spectrum as well as the individual components follow a power-law. Performing a linear fit of the total $\log(E\mathrm{d}N_{\mathrm{int}}/\mathrm{d}E)$ vs. $\log E$ in the range 1 TeV to 0.1 EeV, we obtain an exponent of $-0.890(2)$. This power-law is slightly softer than the one reported in ref. [3, p. 318], where an exponent of -0.78 is stated (without derivation). We may also compare this to the Heitler–Matthews model. Here, we can derive

$$E\frac{\mathrm{d}N_{\mathrm{int}}}{\mathrm{d}E} = \sum_{k=0}^{N_{\mathrm{gen}}} \delta\left(E - \frac{E_0}{m^k}\right) \tag{7.1}$$

for a multiplicity m of long-lived hadrons.

The individual power-laws of unstable hadrons are broken when the corresponding species reaches its critical energy. Around 100 GeV the most long-lived and down to this energy most abundant $\pi^{\pm}$ also start to drop out. The switch from high- to low-energy interaction model causes a sudden change in the nucleon spectra, which is an artifact of the simplified treatment of the HSA. These low-energy nucleon interactions play only a minor role regarding muon production, however, as we will show later.



(a) by primary energy

(b) by radial ranges around shower core

(c) by zenith angle

(d) by hadronic interaction model

Figure 7.2: Number of muon ancestor generations. Vertical dashed lines indicate the mean value of the distribution of the same colour.

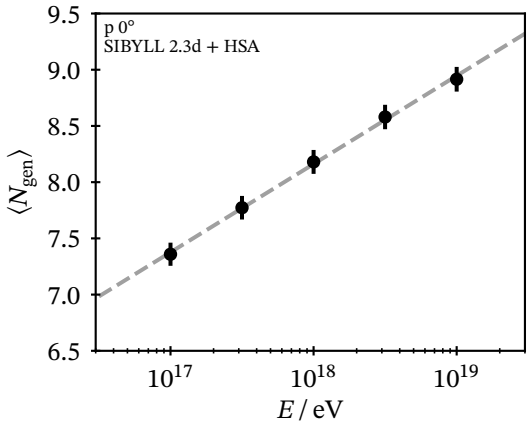
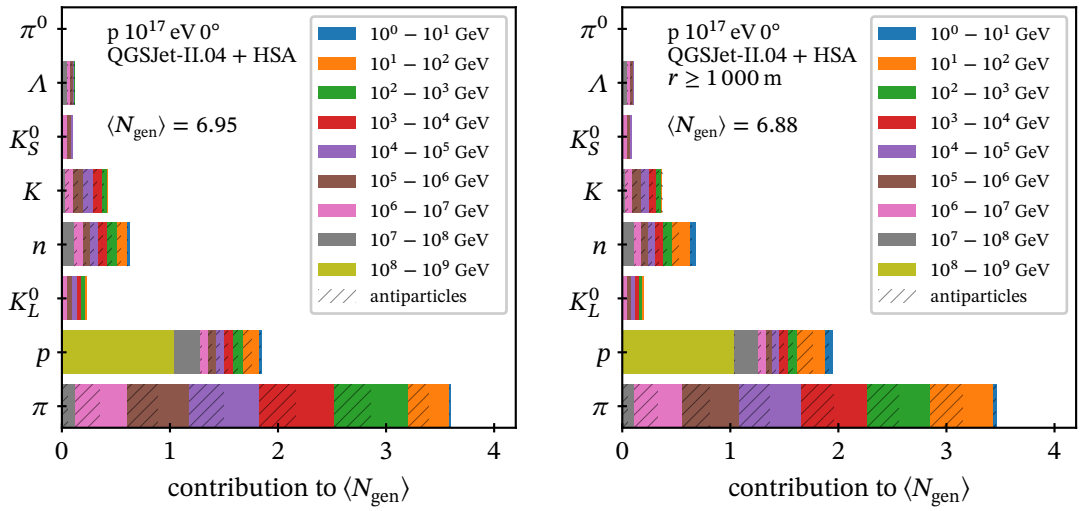


Figure 7.3: Mean number of muon ancestor generations as function of primary energy.

7.1.2 Number of generations

In fig. 7.2 we study the number of generations N_{gen} of ground-reaching muons, which is the total number of hadronic interactions that connect the primary particle with the muon in the shower. It is an important quantity since the number of muons grows exponentially with N_{gen} [52, 170] and small changes in hadronic interactions, e.g. the energy fraction transferred from the projectile onto further long-lived hadronic secondaries, are correspondingly amplified N_{gen} times by the multiplicative process [53, 171]. Making use of the lineage technique, N_{gen} is obtained by iterating over the muon ancestors and counting only interaction events (in contrast to decays). Figure 7.2a shows the distributions for different primary energies. The mean value grows logarithmically with the primary energy as expected from the Heitler–Matthews model [52]. A linear fit of $\langle N_{\text{gen}} \rangle$ vs. $\log_{10}(E)$, in which we include also data of $10^{17.5}$ eV and $10^{18.5}$ eV showers, yields an increase of N_{gen} of $s = 0.785(17)$ per decade of energy (see fig. 7.3). In the Heitler–Matthews model, s is related to the hadron multiplicty m via $s = 1/\log_{10}(m)$ (cf. eq. (2.11)), so that we can derive $m = 18.8(12)$. In fig. 7.2b we consider only muons within certain radial ranges r around the shower core. We observe that muons further away from the core tend to have slightly fewer generations than those close to the core. Figure 7.2c shows the dependency on the zenith angle. While the distribution widens slightly for inclined showers, the mean stays virtually unaffected. A greater change can be seen when comparing different hadronic interaction models (see fig. 7.2d). With SIBYLL 2.3d muons tend to have more ancestor generations than with QGSJet-II.04. The difference of their mean values also increases with energy.

It is instructive to quantify to which degree interactions in certain energy ranges and with certain projectiles contribute to the total N_{gen}. A priori we can only state the obvious: The first interaction, being the root of the shower, contributes exactly one generation. We build a histogram binned in projectile energy and species by iterating over the muon lineages and filling the histogram for each interaction according to its projectile energy and species.



(a) (b)

Figure 7.4: Distribution of muon ancestor projectile generations by energy and species (a) all muons (b) only muons with $r \geq 1000$ m.

Thereby each muon increases the total histogram count by its individual N_{gen}. Since muons share parts of their lineage, the corresponding interactions are counted multiple times – their *muon weight* is given by the number of muons stemming from that interaction. If we finally divide the bin counts by the total number of muons (possibly after applying a section criterion, e.g. on r), we end up with that bin's contribution to $\langle N_{\mathrm{gen}} \rangle$. The result of that procedure is shown in fig. 7.4, applied to 10^{17} eV showers. As expected, the bin containing solely the primary interaction (■-coloured) has a value of one. In the energy decade below the primary energy, the main contribution is due to nucleons with about twice as many protons as neutrons, conforming with fig. 7.1. In this energy range virtually no contribution of antinucleons is apparent. Charged pions contribute approximately half of the total $\langle N_{\mathrm{gen}} \rangle$. Each $\log E$ range between 10 PeV and 100 GeV carries comparable weight, slightly decreasing with energy. No distinction between positively and negatively charged pions can be observed. Below 100 GeV the importance of pion interactions decreases again as more and more pions do not reinteract. Comparing the distributions obtained when selecting only muons with at least 1 km lateral distance (fig. 7.4b) with those without any cut (fig. 7.4a), we find that for great lateral distance the importance of low-energy interactions increases. This can be understood considering that typically muons with higher energies stay close to the shower core. As the projectile energy of the last interaction of these muons needs to be higher than the final muon energy, the phase space that can contribute to these muons is necessarily cut off earlier. These results shed new light on findings by Drescher et al. [172] and Mariş et al. [173], where it is shown that the lateral distribution of muons at large distances, say $\gtrsim 1$ km, is greatly affected by the choice of the low-energy interaction model.

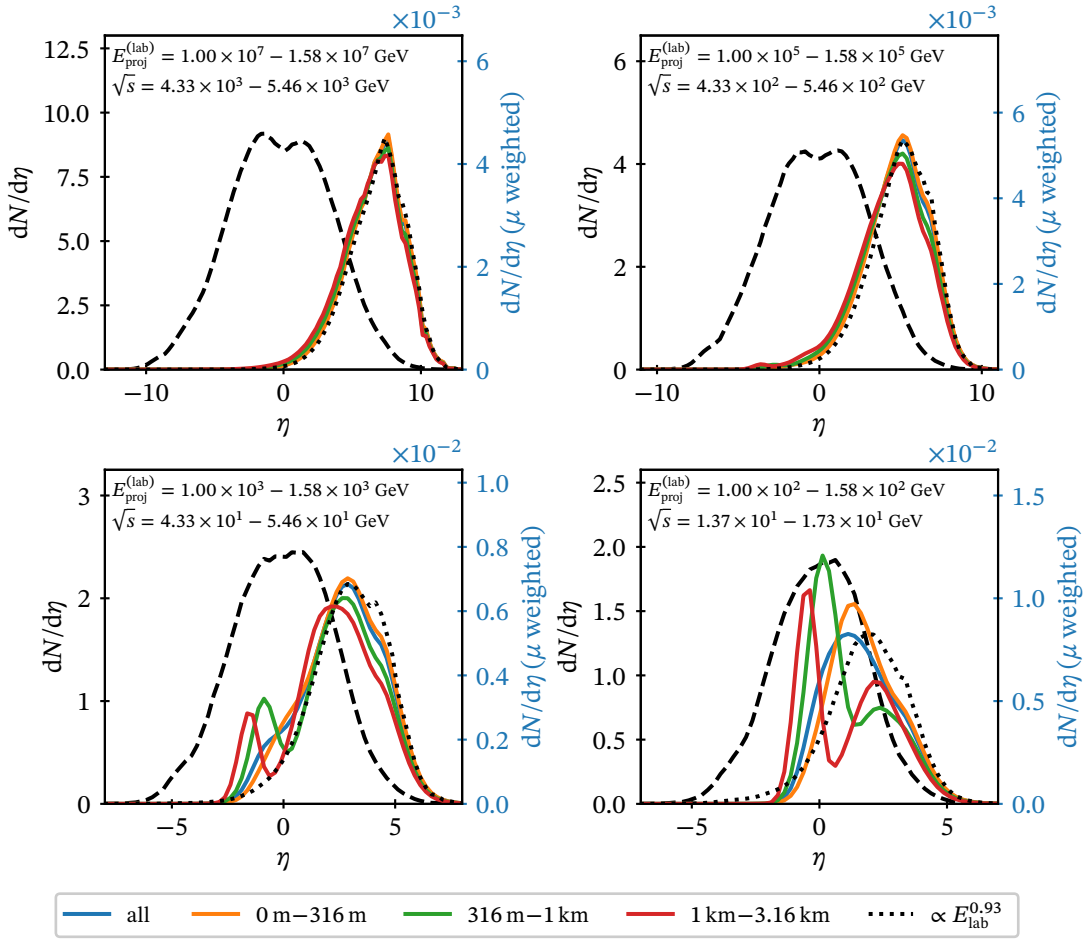


Figure 7.5: Pseudorapidity distributions of $\pi^{\pm}$ + Air $\rightarrow$ charged hadrons, 10^{19} eV vertical proton. The dashed line indicates generator-level distributions while the coloured and dotted lines shows muon-weighted distributions (in arbitrary units).

7.1.3 Pseudorapidity distributions

So far we have considered only the initial states of hadronic interactions, i.e. properties of the projectile. Also the kinematic distributions of the secondary particles are of interest since they can (ideally) be measured in collider experiments.

To quantify the importance of different regions of the phase-space of secondaries in hadronic interactions for muon production, one may weight specific phase-space element by the number of muons descending from particles produced in it. A simple prescription based on the Heitler–Matthews model is to weight by E_{lab}^{β} (with $\beta = 0.93$, E_{lab} is the energy of the secondary in the lab frame) [53, 170]. This, however, does not allow for any cut to be applied on the muons, e.g. on lateral distance. Having the full lineage available in our Monte Carlo simulations, we follow the approach of Hillas [149] and obtain the weight by counting the

muons as described in the previous section. In fig. 7.5, we show pseudorapidity distributions $\mathrm{d}N/\mathrm{d}\eta$ of $\pi^{\pm}$ + Air → charged hadrons (in centre-of-mass frame) for four different energies. The pure generator-level distributions (generated with SIBYLL 2.3d in CRMC [174] for a fixed projectile energy E_{p} and ^{14}N target) are plotted with dashed lines. The corresponding muon-weighted distributions are obtained from simulations of vertical 10^{19} eV proton showers in which we consider interactions within a range around E_{p}. The solid, coloured lines indicate the weighted distributions (in arbitrary units) after applying a lateral distance cut and normalized by the number of muons selected. Additionally, the black solid line shows the E_{lab}^{β}-based weighting for comparison. We find that at lab energies $\gtrsim 10^{14}$ eV the weighted distributions almost coincide irrespective of the muon lateral distance. Furthermore, the E_{lab}^{β}-based weight agrees very well with the distributions (up to an arbitrary scaling factor). These results quantitatively demonstrate the importance of the forward region of hadronic interactions for muon production. At lower energies, on the other hand, the muon lateral distance has an impact on the corresponding weight distributions. Besides the peak in the forward region, a second peak at mid-rapidity around $-2 \lesssim \eta \lesssim 0$ emerges when only muons with at least a few hundred meters distance are considered, which is not described by the E_{lab}^{β}-based weighting.

7.2 Modified characteristics of hadronic interaction models

In this section we consider two aspects of hadronic interaction models for which lab measurements exist only up to a few hundred GeV (lab), so that extrapolations up to the highest energies are necessary. We explore the impact on air showers of changes to the energy dependence, in particular with respect to muon production.

The common setup is as follows: CORSIKA 8 is used to treat the hadronic and muonic shower components with detailed Monte Carlo simulations. The electromagnetic component is treated in a hybrid fashion with CONEX. A limitation of this approach is that no feedback from the EM component back to the hadronic component is considered, so that muons from photonuclear interactions are missing, which contribute up to $\sim$ 10 % to the total muon number for proton showers, depending on the cutoff energy [175]. Since we are interested only in relative changes of the muon content, this is only of minor relevance here. Keilhauer's atmopsheric model is used because muon production depth distributions are one of the observables of interest in this section. The observation level is set to 1400 m. To reduce statistical fluctuations the height of the first interaction is fixed to a value corresponding to a slant depth of 43 g cm^{-2}, which is approximately one interaction length. The energy cut is set to 631 MeV.

To obtain distributions of X_{max} and $X_{\mathrm{max}}^{(\mu)}$, the following procedure is applied shower-by-shower on $\mathrm{d}E/\mathrm{d}X$ profiles (10 g cm^{-2} resolution) and the AMPD distribution (5 g cm^{-2} resolution): First, the maximum entry is searched and taken as initial guess. Then, a second-degree polynomial is fitted in the region ±2 (for X_{max}) or ±10 (for $X_{\mathrm{max}}^{(\mu)}$) bins around the initial guess. The maximum of this parabola is then taken as final value.

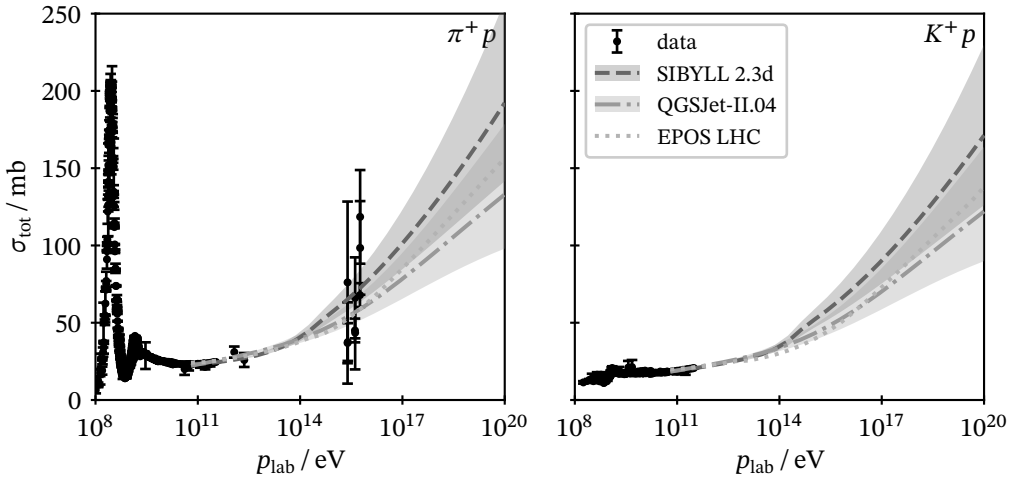


Figure 7.6: Total cross-sections of π^+/K^+ p. Data are from the PDG compilation [11]. The bands around the SIBYLL 2.3d and QGSJet-II.04 lines show the range obtained when varying $f_{19}^{(\pi/K)}$ in $[0.77, 1.30]$. The error bars of the PDG data designate systematic and statistical uncertainties added in quadrature.

7.2.1 Hadron-air cross-sections

While $p\,p$ cross-sections from accelerator measurements are available up to $E_{\mathrm{lab}} = 10^{17}$ eV, much less is known for meson projeciles. The highest-energy direct measurement of total cross-sections with pion beams on carbon was made at the SELEX experiment at $E_{\mathrm{lab}} \simeq 600$ GeV [176]. Above that energy, only data from indirect measurements are available, which make use of charge-exchange reactions in $p\,p$ collisions [177–179]. In this type of interactions a π^+ is exchanged and one of the protons is converted into a neutron. This virtual pion can be thought of as a projectile. These indirect measurements suffer from large uncertainties, however, and barely constrain the hadronic interaction models. The lack of constraining measurements is reflected in a rather large spread between model predictions (see fig. 7.6).

Data of inealstic cross-sections, especially for air-like targets like carbon, are more directly suited for air shower simulations. A compilation of such data are shown in fig. 7.7 together with model predictions. While the absolute spread between the models is smaller for proton than for carbon targets, the opposite is true when considering relative spreads. To a great degree the nuclear cross-section is set by the geometric size of the nucleus, so that differences in the hadronic physics modelling have a smaller impact.

The kink visible in the SIBYLL 2.3d lines at $10^{14.5}$ eV stems from a switch between different parameterizations.

Ulrich et al. [9] study, among other variables, the extrapolation uncertainty of hadron-air cross-sections on basic EAS observables by introducing an ad hoc modification factor to scale the model cross-sections in an energy-dependent way above $E_{\mathrm{lab}} = 1$ PeV. In this

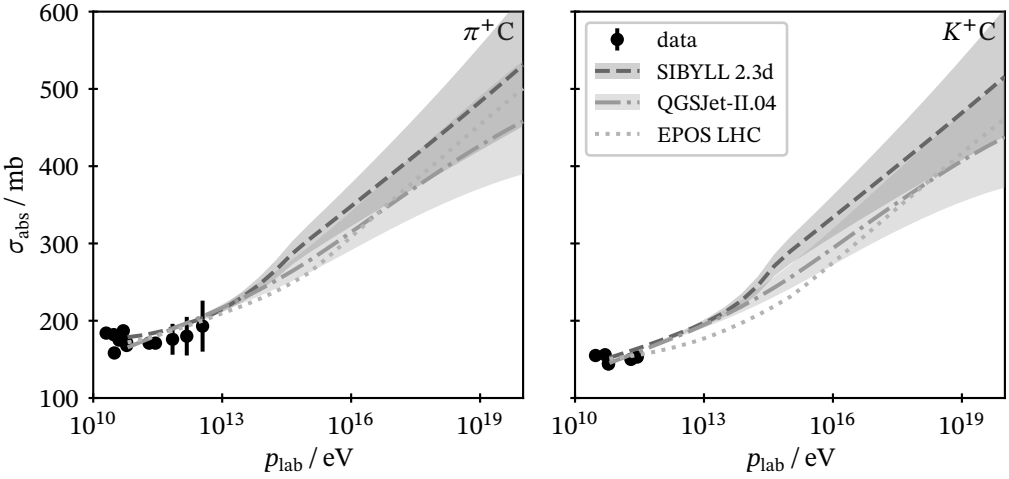


Figure 7.7: Absorption/inelastic cross-sections for π/K on C targets. Data are from Denisov et al. [180] (20 GeV to 50 GeV), Aduszkiewicz et al. [181] (31 GeV and 60 GeV), Carroll et al. [182] (60 GeV to 280 GeV) and Avakian et al. [183] (700 GeV to 3500 GeV). The bands around the SIBYLL 2.3d and QGSJet-II.04 lines show the range obtained when varying $f_{19}^{(\pi/K)}$ in $[0.87, 1.15]$.

section we consider an extension of that study that is different in several aspects. Since meson cross-sections are well constrained only up to several hundred GeV, the threshold is set to $E_{\mathrm{th}} = 1\,\mathrm{TeV}$. Additionally, the modification factors for pions, kaons and nucleons are independent. For consistency, also the nucleon cross-sections are modified starting at 1 TeV, although for these reliable measurements are available up to 10^{17} eV from LHC data. The functional form of the modification factor follows ref. [9]. For the modified inelastic cross-section $\sigma_{\mathrm{inel}}^{(s)}$ of projectiles of class $s \in \{\pi, K, N\}$ we define

$$\sigma_{\mathrm{inel}}^{(s)}(E_{\mathrm{lab}}) = \hat{\sigma}_{\mathrm{inel}}^{(s)}(E_{\mathrm{lab}}) \times \left(1 + \frac{f_{19}^{(s)} - 1}{\log \frac{E_{\mathrm{th}}}{10^{19}\,\mathrm{eV}}} \max\left(1, \log \frac{E_{\mathrm{lab}}}{E_{\mathrm{th}}}\right)\right), \tag{7.2}$$

where $\hat{\sigma}_{\mathrm{inel}}^{(s)}(E_{\mathrm{lab}})$ is the original cross-section of the hadronic interaction model. As visible in fig. 7.7 values of $f_{19}^{\pi/K}$ that deviate from 1 by 10 % to 15 % are sufficient to push the QGSJet-II.04 line almost onto the SIBYLL 2.3d line and vice-versa, which therefore poses a minimum natural range of the parameter. The implementation of this in CORSIKA 8 is based on a wrapper process that wraps the original interaction process and replaces the `getInteractionLength()` method with a generic function. The important part of the implementation is listed in listing 7.1.

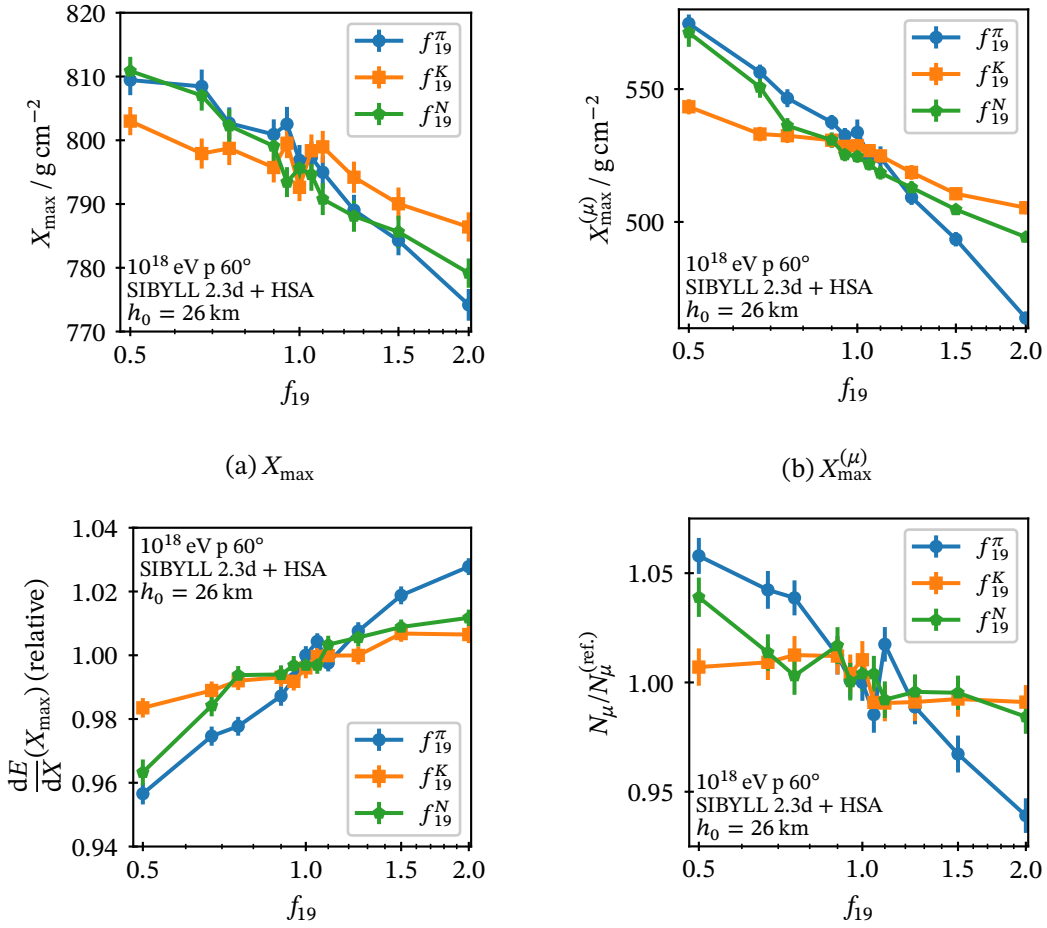
In figs. 7.8 and 7.9 the impact on X_{max}, $X_{\mathrm{max}}^{(\mu)}$, $\mathrm{d}E/\mathrm{d}X(X_{\mathrm{max}})$ and N_μ are shown for 60° proton showers at 10^{18} eV with SIBYLL 2.3d and QGSJetII-0.4, respectively. Each modification factors is varied independently with the two remaining ones staying unmodified. Varying the

Listing 7.1: Implementation of `InteractionLengthModifier`, abridged.

```
template <class TUnderlyingProcess>
class InteractionLengthModifier
    : public InteractionProcess<InteractionLengthModifier<TUnderlyingProcess
       >> {
public:
  template <typename TParticle>
  GrammageType getInteractionLength(TParticle const& particle) {
    GrammageType const original = process_.getInteractionLength(particle);
    Code const pid = particle.getPID();
    HEPEnergyType const energy = particle.getEnergy();
    return modifier_(original, pid, energy);
  }

  // [...]

  // signature of the modifying functor
  // arguments are original int. length , PID , energy
  using functor_signature = GrammageType(GrammageType, Code, HEPEnergyType);

private:
  TUnderlyingProcess& process_;
  std::function<functor_signature> const modifier_;
};
```

hadron cross-sections mainly affects the longitudinal shower development as the interaction length, which defines the characteristic scale, varies accordingly. Both $X_{\max}$ and $X_{\max}^{(\mu)}$ shift towards shallower depths for increasing cross-sections because the shower gets compressed as the interaction length decreases. For $X_{\max}$ the effect is less pronounced than for $X_{\max}^{(\mu)}$ because the bulk of the EM shower development quickly decouples from the hadronic part. Shallower depths also increase the probability of low-energy pions to decay instead of re-interacting so that the multiplicative cascade process stops earlier, producing fewer muons in total. Additionally, these muons have longer distances to travel to reach ground, during which a greater fraction decays, which reduces the observable number of muons even further. Keeping the results of the previous section in mind, it is not surprising that varying pion cross-sections typically has the largest, varying kaon cross-sections the smallest effect.

Compressing or stretching the longitudinal development also impacts the particle distributions at ground beause the shower has reached a different age compared to the unmodified one when reaching the ground level. The ratios of muon energy spectrum and lateral distribution are shown in fig. 7.10 for a mumber of values of f_{19}^{π}.



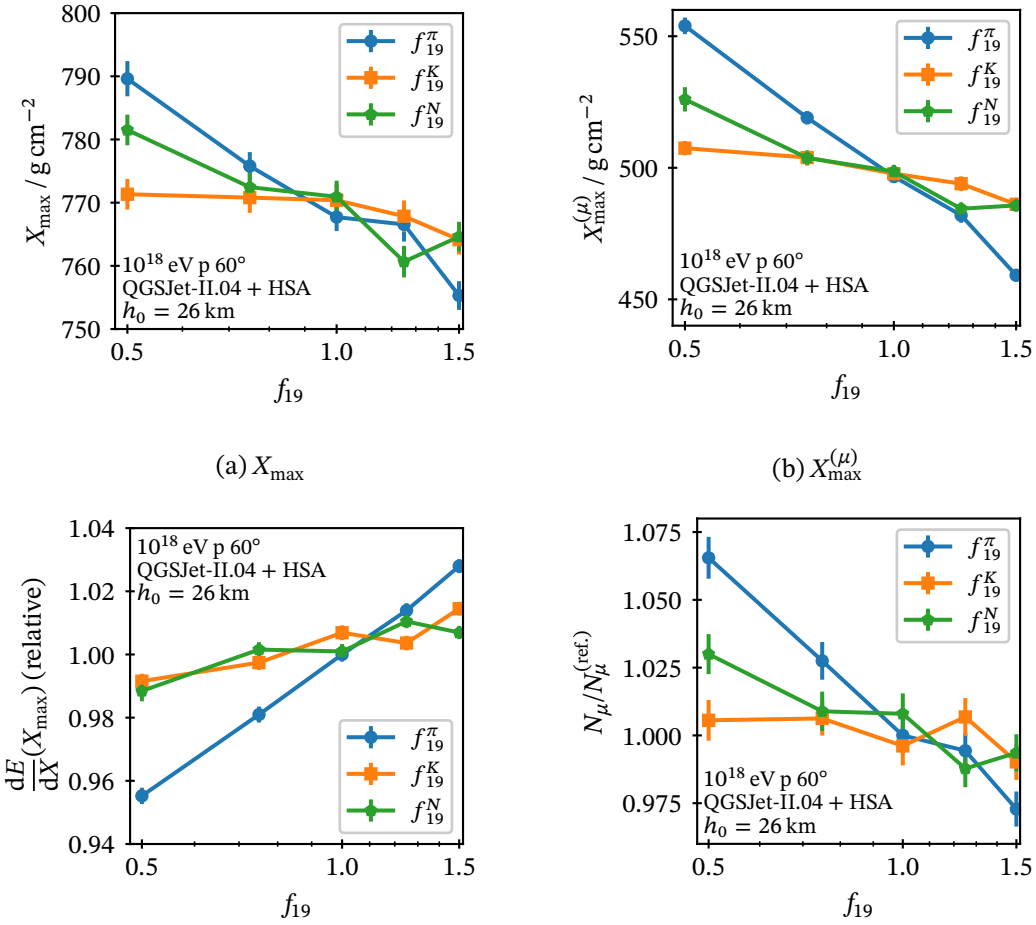
(a) $X_{\max}$

(b) $X_{\max}^{(\mu)}$

(c) relative change of $\mathrm{d}E/\,\mathrm{d}X(X_{\max})$

(d) relative change of N_μ

Figure 7.8: Change of observables with f_{19} for modified cross-sections, SIBYLL 2.3d, averaged over 500 showers



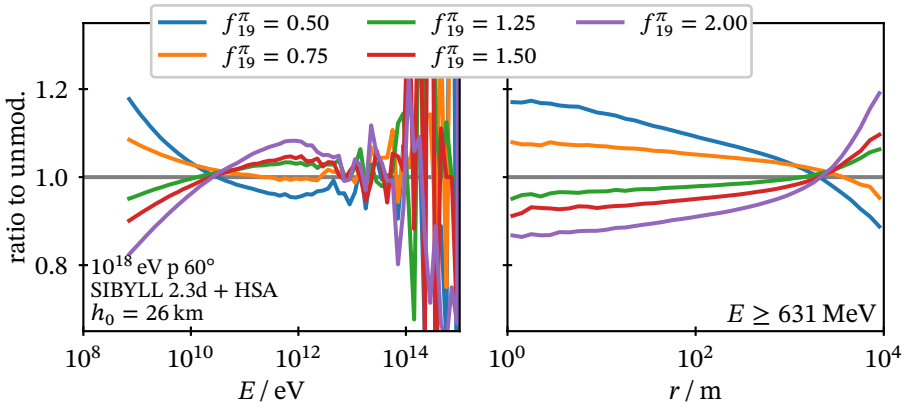
(a) $X_{\max}$

(b) $X_{\max}^{(\mu)}$

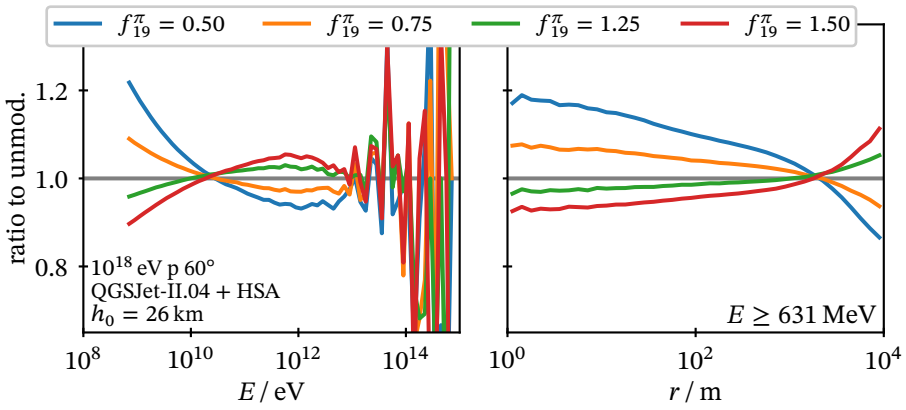
(c) relative change of $\mathrm{d}E/\,\mathrm{d}X(X_{\max})$

(d) relative change of N_μ

Figure 7.9: Change of observables with f_{19} for modified cross-sections, QGSJetII-0.4, averaged over 500 showers



(a) SIBYLL 2.3d



(b) QGSJetII-0.4

Figure 7.10: Ratios of muon energy spectrum and lateral distribution.

7.2.2 Pion charge exchange reactions and ρ^0 production

Enhanced production of leading ρ^0 mesons instead of leading π^0 in pion charge exchange reactions has been identified as possible way of increasing the muon number by retaining a greater fraction of energy in the hadronic cascade [186, 187]. ρ mesons are spin-1 excitations of pions and carry the same quark content. With masses of about 775 MeV they are heavier than pions and thus may decay into them. In particular, the virtually exclusive decay mode of ρ^0 is $\rho^0 \to \pi^+\pi^-$, which happens quasi-immediately with a lifetime of $\sim 10^{-24}$ s. Because of this, the production of a ρ^0 in place of a π^0 has consequences for the shower development. Producing a ρ^0 effectively increases the charged pion multiplicity and reduces the electromagnetic energy fraction, so that more energy is retained the hadronic cascade, leading to an increase in the number of muons. Depending on how early or late in the shower development such an enhancement of ρ^0 production takes place, the EM cascade is more or less affected since the amount of energy fed into it is greatest where the highest-energy π^0 are produced and is reduced in each subsequent generation.

Data of fixed-target $\pi^+ p$ collisions at $E_{\mathrm{lab}} = 250\,\mathrm{GeV}$ obtained at the EHS/NA22 experiment show that the x_{F} spectra cross at at $x_{\mathrm{F}} \simeq 0.5$, above which ρ^0 are more abundantly produced than π^0 [184, 185] (see fig. 7.11). More recently, this leading ρ^0 enhancement has been confirmed for carbon targets in the NA61/SHINE experiment at $E_{\mathrm{lab}} = 158\,\mathrm{GeV}$ [188]. A measurement of π^0 production in pion-nucleus interactions is still pending, however, leaving the total effect on the muon number unconstrained [37].

In hadronic interaction models, leading π^0 and ρ^0 occur as final state of the projectile remnant in charge exchange reactions of $\pi^\pm$ projectiles in which their charge is transferred to the target. Special attention to this enhancement has been paid in the update of SIBYLL 2.1 to 2.3 [37, 189], which has been tuned to reproduce the EHS/NA22 measurements. Due to the lack of measurements, little is known about the effect at higher energies [190]. In SIBYLL 2.3d no explicit energy dependence is built in.

Here we study the impact of extrapolation uncertainty to higher energies on EAS observables by applying an ad hoc prescription to exchange π^0 and ρ^0 in $\pi^\pm$ Air collisions. We generate standard SIBYLL 2.3d events and apply a probablisitic modification of the particle ID based on a function $p(E)$ depending on the projectile energy, again motivated the f_{19} approach of ref. [9]:

$$p(E) = \mathrm{clamp}\left(f_{19} \max\left(0, \log\frac{E}{E_{\mathrm{th}}}\right) \Big/ \log\frac{E_{\mathrm{th}}}{10^{19}\,\mathrm{eV}} \right), \tag{7.3}$$

where $\mathrm{clamp}(x)$ is defined as $\mathrm{clamp}(x) = \min(1, \max(-1, x))$ in order to limit the codomain to $[-1; 1]$. At energies below the threshold energy E_{th}, which is set to 100 GeV, $p(E)$ vanishes. Above the threshold $p(E)$ increases (decreases) logarithmically for positive (negative) values of f_{19}, which determines the slope. Some examples are illustrated in fig. 7.12. For positive values of $p(E)$ we replace $\pi^0 \to \rho^0$ with probability $p(E)$. For negative values we replace $\rho^0 \to \pi^0$ with probability $-p(E)$. We also introduce up to two additional selection criteria:

a) Only the leading particle is converted in contrast to all secondary π^0/ρ^0 independently of each other being subject to a random conversion.

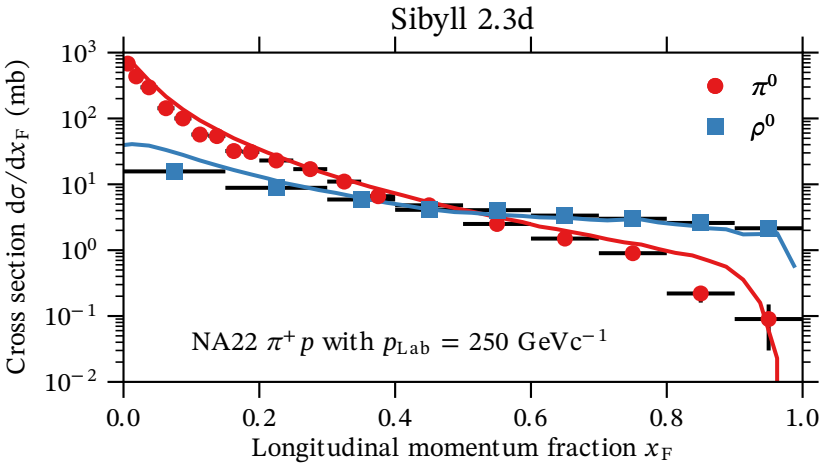


Figure 7.11: x_F spectra of $\pi^+ p \to \pi^0/\rho^0$ with SIBYLL 2.3d. Adapted from ref. [37]; data from ref. [184, 185]

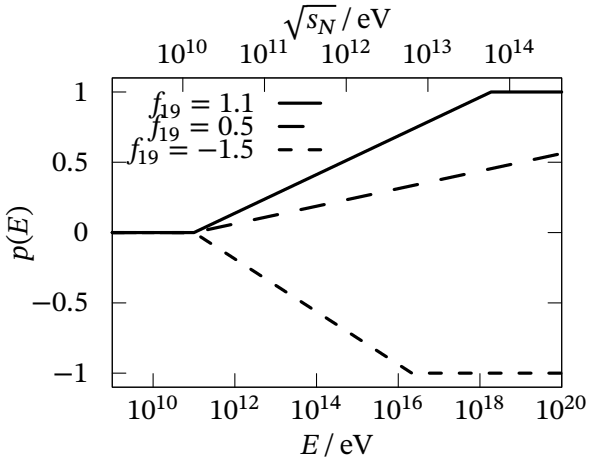


Figure 7.12: Conversion function $p(E)$.

b) Only secondaries having x_L values above a given threshold are considered for conversion. We choose x_L instead of x_F for simplicity. For the values considered here the difference is negligible.

These prescriptions are implemented in CORSIKA 8 as `SecondariesProcess`. Preliminary results of this study have been published as part of ref. [167].

We consider three cases. In case A, none of the selection criteria is applied. All secondary π^0/ρ^0 mesons may be converted, which means that not only the leading particle is affected and also multiple particles may be converted. This is the most extreme scenario, going beyond altering only pure charge-exchange reactions. In case B, we consider only the leading particle. In case C, a cut of $x_\mathrm{L} = 0.2$ is introduced in addition to the leading-only cut, which is a simplistic way of considering only particles stemming from the projectile remnant.

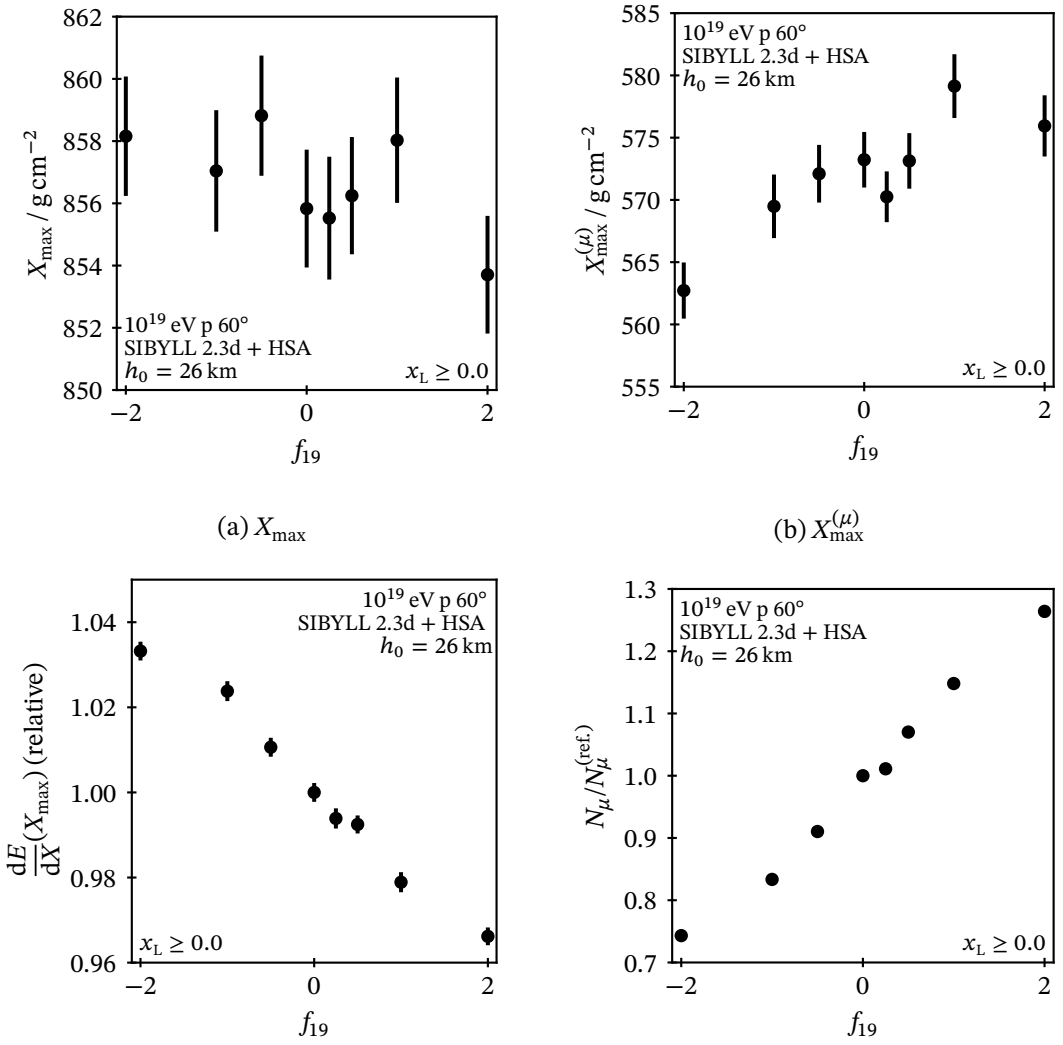
The effects on 10^{19} eV proton showers at 60° are summarized in fig. 7.13 for case A. The mean shower maximum X_max is hardly affected (fig. 7.13a). Neither the primary interaction nor the interactions of its leading nucleons are subject to the modificiation, so that the highest-energetic π^0 stay unchanged. This is consistent with the statement in ref. [37] that the introduction of the ρ^0 enhancement has negligible impact on X_max. The change of $X_\mathrm{max}^{(\mu)}$ is smaller than $\pm 10\,\mathrm{g\,cm^{-2}}$ within the parameter range considered (fig. 7.13b), which is very little compared to the difference between EPOS-LHC and QGSJet-II.04 of $\sim 50\,\mathrm{g\,cm^{-2}}$ shown in ref. [71]. All in all, the longitudinal development is affected only to a minor degree. On the other hand, the muon number is very sensitive to the exchange and displays a linear dependence on f_{19} (fig. 7.13d). The maximum change amounts to ±27 %. The electromagnetic shower size (fig. 7.13c), for which $\mathrm{d}E/\mathrm{d}X(X_\mathrm{max})$ serves as a proxy, behaves in an anticorrelated way as can be expeced from energy conservation. Here, the maximum change amounts to ±3.5 %. The relative changes of energy spectrum and lateral distribution are shown in fig. 7.14. We observe that muons at all energies and lateral distances are affected to almost equal degrees. Only in the energy spectra a slight slope is visible, which can be explained by the fact that muons at higher energies typically have fewer hadronic ancestor generations, so that the effect cannot accumulate as much as with low-energy muons.

The results for case B are shown in fig. 7.15. Since X_max and $X_\mathrm{max}^{(\mu)}$ hardly vary already in case A, we do not consider them here. The restriction to leading particles limits the achievable variation of N_μ (fig. 7.15a) and $\mathrm{d}E/\mathrm{d}X(X_\mathrm{max})$ (fig. 7.15b) significantly. Furthermore, the data in the positive f_{19} range change less than in the negative range. This suggests that only few leading π^0 remain that could be converted, while leading ρ^0 are more abundant.

This saturation effect is even more pronounced in case C (fig. 7.16). Here, the $\pi^0 \to \rho^0$ conversion can increase N_μ up to 5 %, while the $\rho^0 \to \pi^0$ conversion reduces N_μ by more than 10 %.

7.2.3 Conclusions

The two modifications studied in this section are orthogonal. Altered hadron-air cross-sections change mainly the longitudinal shower development. The MPD with $X_\mathrm{max}^{(\mu)}$ is even more sensitive than the EM shower maximum by a factor of about three. The EM component quickly decouples from the hadronic one, while for the MPD the mean free path between



(a) $X_{\max}$

(b) $X^{(\mu)}_{\max}$

(c) relative change of $\mathrm{d}E/\,\mathrm{d}X(X_{\max})$

(d) relative change of N_μ

Figure 7.13: Mean change of observables with f_{19} for modified ρ^0 production, case A, averaged over 500 showers

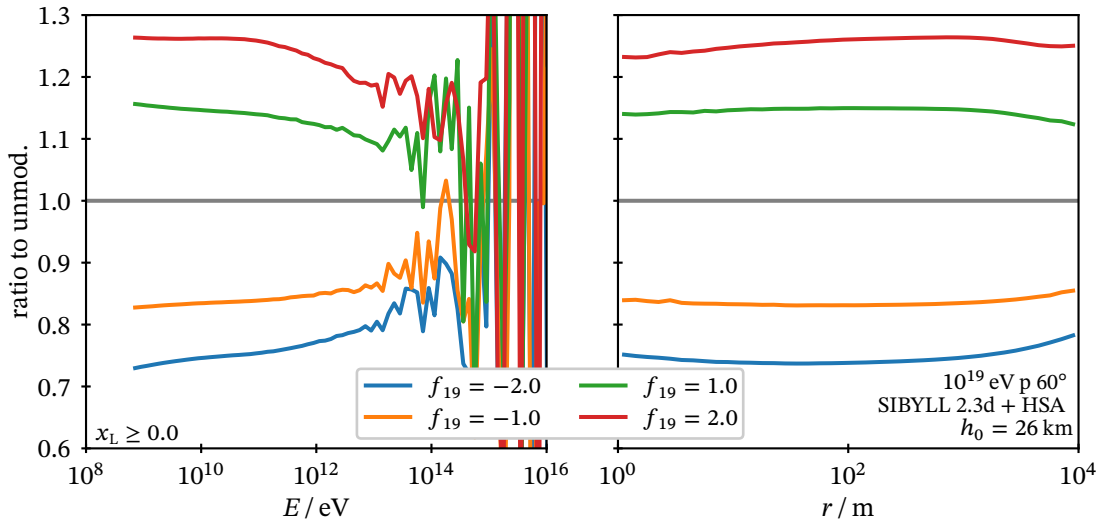
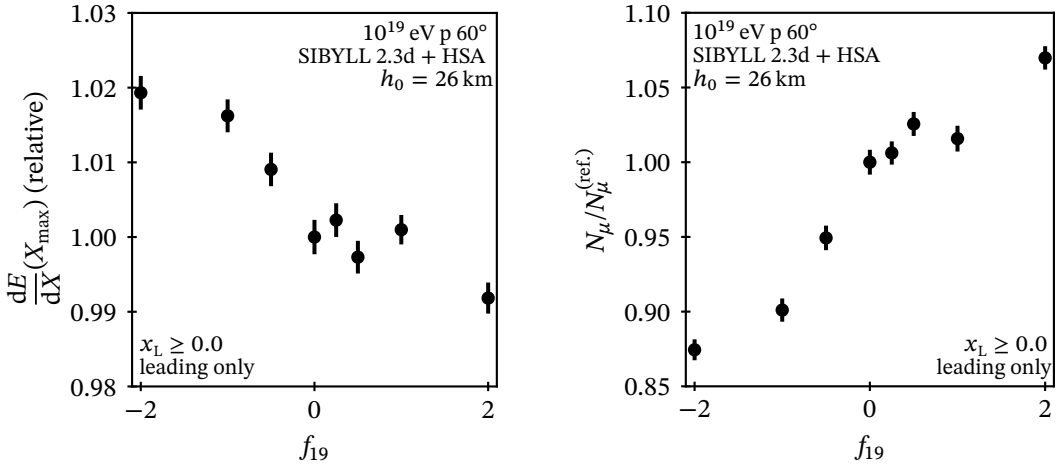


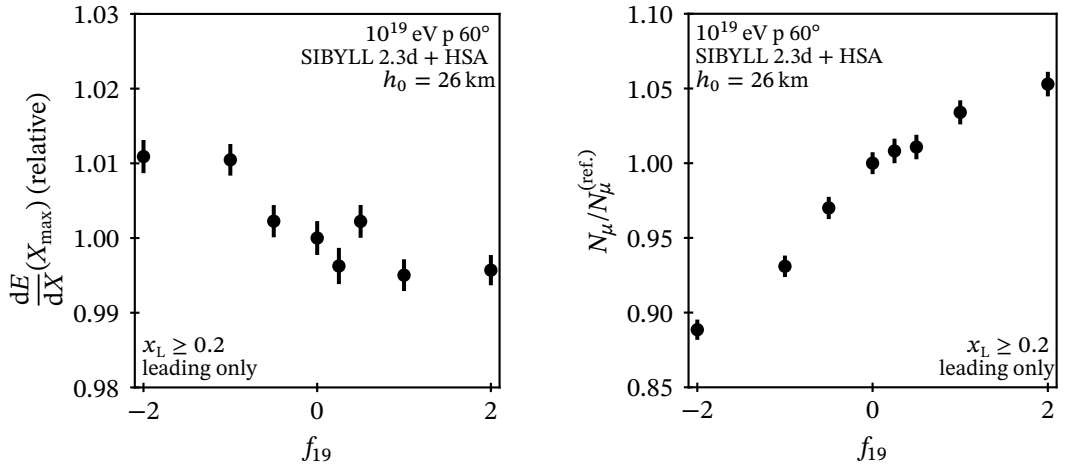
Figure 7.14: Ratios of muon energy spectrum and lateral distribution for modified ρ^0 production, case A, averaged over 500 showers



(a) Relative change of $dE/dX(X_{max})$

(b) Relative change of N_μ

Figure 7.15: Mean change of observables with f_{19} for modified ρ^0 production, case B, averaged over 500 showers



(a) Relative change of $\mathrm{d}E/\,\mathrm{d}X(X_{\max})$

(b) Relative change of N_μ

Figure 7.16: Mean change of observables with f_{19} for modified ρ^0 production, case C, averaged over 500 showers

collisions add up N_{gen} times. Especially the poorly constrained meson-air cross-sections together with some theoretical freedom in the relation between nucleon-air and meson-air cross-sections are an ingredient towards understanding the inconsistency of $X_{\max}$ and $X_{\max}^{(\mu)}$ seen in data of the Pierre Auger Collaboration [71]. The required changes, a reduction of model predictions of $X_{\max}^{(\mu)}$ towards data by increasing the cross-section, would decrese the muon content by a few percent, however.

Regarding enhanced ρ^0 production, the results show that SIBYLL 2.3d leaves little room for an even higher yield of forward ρ^0 from charge-exchange reactions. On the other hand, an unconstrained enhanced ρ^0 production (case A) is a very efficient way of increasing the muon content without affecting the longitudinal development. Although the concept shown here is unrealistic, it is an example demonstration of how a change of π^0 production (in other words the energy fraction going into EM particles) influences EAS observables. Collective effects and enhanced strangeness production, which have a similar effect, are entering the focus as they have lately been found (surprisingly) also in small systems (see ref. [53]). Dedicated measurements will improve the understanding and modelling of such systems and consequently also muon production in air showers.

8 Large multiplicity events

Up to now we have only considered exclusively strong interactions for EAS hadrons. The influence of weak processes in interactions is usually completely ignored in this context due to weak cross-sections being 10^6 times smaller than the strong cross-sections. In this chapter we consider a type of process where weak interactions might give rise to cross-sections comparable to usual hadronic interaction cross-sections. Via simulations with a simple toy model, we study the consequences for UHECR air showers if such an event happens in the first interaction.

The content of this chapter has been published in ref. [191].

8.1 Context

Scattering processes with large multiplicity of final state bosons at very high energies have been studied for instance for the standard model [192]. The amplitude of such processes, which have final states with $n \gg 1$ Higgs bosons, grows factorially with n due to the large number of contributing Feynman diagrams [192, 193]. Provided that this growth of the amplitude reflects reality (rather than an incomplete calculation) and under the assumption that perturbative unitarity is ensured [194], the cross-section must be bounded from above, favouring a specific large process energy and thus number of Higgs bosons produced. Figure 8.1 the exclusive cross-section of such processes in pp collisions as a function of the multiplicity for different centre-of-mass energies as estimated by Degrande et al. [195]. Taken at face value, the energy fraction at $\sqrt{s} \geq 50\,\mathrm{TeV}$ converted into Higgs bosons can reach $\mathcal{O}(30\,\%)$.

8.2 Model and implementation

We include the large-multiplicity process phenomenologically into our CR simulations with a simplified toy model.

We model the primary event in the following way: A fixed fraction f of the centre-of-mass-energy per nucleon $\sqrt{s_{\mathrm{NN}}} = \sqrt{2 m_{\mathrm{N}} E_0}$ is used to create Higgs particles, while the underlying event is generated with QGSJet-II.04 with the remaining energy $(1-f)\times E_0$. Strictly speaking, this procedure is only an approximation since a fraction of the actual phase space stays unavailable for hadron production in QGSJetII in our implementation. This could be alleviated by generating the event with the whole energy and replacing a fraction of the secondaries by Higgs particles afterwards.

A second parameter ε is introduced to model the kinetic energy of the Higgs particles relative to their mass. To ensure momentum conservation we always generate an even number

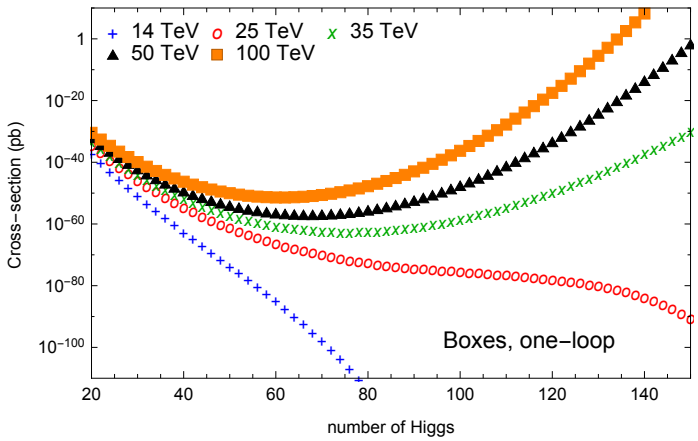


Figure 8.1: Cross-section for multi-Higgs production in pp interactions; taken from ref. [195]

n_h of Higgs particles in $n_\mathrm{pairs} = n_\mathrm{h}/2$ pairs with their momentum vectors back-to-back. The number of pairs is given by

$$n_\mathrm{pairs} = \left\lfloor \frac{f\sqrt{s_\mathrm{NN}}}{2m_\mathrm{h}(1+\varepsilon)} \right\rfloor, \tag{8.1}$$

where we consider $\varepsilon = 0.1,\ 1,\ 2$ for concreteness. The available total kinetic energy

$$T = f\sqrt{s_\mathrm{NN}} - n_\mathrm{h} m_\mathrm{h} \tag{8.2}$$

is split among the pairs by distributing the individual energy portions per pair $T^{(i)}_\mathrm{pair}$ uniformly on the $(n_\mathrm{pairs}-1)$-simplex defined by the conditions

$$\sum_{i=1}^{n_\mathrm{pairs}} T^{(i)}_\mathrm{pair} = T \quad \text{and} \quad 0 \le T^{(i)}_\mathrm{pair} \le T. \tag{8.3}$$

We distribute the momentum vectors isotropically in the proton-nucleon centre-of-mass system. In the subsequent course of the simulation we let the Higgs particles decay (using Pythia v8.235) into long-lived particles. EM particles occuring during the simulation are fed into CONEX to generate EM longitudinal profiles. Secondary hadronic interactions are treated with QGSJetII-04 above and with UrQMD below 60 GeV.

8.3 Results

We simulate proton-induced, vertical showers of the primary energy $E_0 = 10^{19}$ eV. Since the proton-nucleus cross-section, which determines the distribution of the height of the primary interaction h_0, of a large-multiplicity process is unknown, we set a fixed $h_0 = 20$ km above sea level.

In fig. 8.2 the first two moments of the X_max distribution, $\langle X_\mathrm{max}\rangle$ and $\sigma(X_\mathrm{max})$, are shown as a function of f and ε. $\langle X_\mathrm{max}\rangle$ decreases almost linearly with f with a slope only mildly depending on ε. Also the fluctuations $\sigma(X_\mathrm{max})$ are reduced with increasing f, almost independent of

ε. Furthermore, the values obtained for proton and iron showers without new physics are indicated in fig. 8.2 as dashed lines. Consequently, disregarding the artificially fixed h_0, for some values of f and ϵ the large-multiplicity showers could be incorrectly classified as due to heavier primaries if interpreted using only standard hadronic interactions. It appears that values of $f \gtrsim 0.4$ are disfavoured by the fact that such showers seem heavier than iron.

The energy spectrum of muons at ground is depicted in fig. 8.3. The most striking feature is the increase of muons with energies above $\gtrsim 10\,\mathrm{TeV}$. These are prompt muons stemming from the decay of the Higgs bosons. Their maximum energy is shown to be dependent on ε. Moreover, we observe an increase up to 50 % depending on the model parameters throughout the whole energy spectrum with only a slight change in the overall slope.

We notice that the here presented muon spectra are qualitatively comparable to those obtained by Brooijmans et al. [196] with sphalerons as first interaction, suggesting a somewhat general signature of electroweak physics within the first interaction in EAS.

8.4 Conclusions

The results presented here show that large multiplicity Higgs events in the first interaction of proton-induced air showers generally increase the number of muons and lead to shallower depths of X_{max}. Both make such showers appear heavier when interpreted only with standard hadronic interactions. A clear signature is the increase of very high energy muons ($E \gtrsim 10\,\mathrm{TeV}$). In the PeV range their number is increased by a factor of up to 1000. It is, however, experimentally virtually impossible to measure the flux of such high energy muons in single air showers.

Nevertheless, one needs to take the simplifications made here into account. The fixed height of first interaction affects mainly the X_{max} distribution (and therefore its mean and its variance), which in reality depends on the cross-section of the process. Moreover, we have not considered large multiplicity events in secondary interactions. These would increase the potential effects to a certain degree, provided that the primary energy is sufficiently high so that also secondaries are above the threshold where large multiplicity events become dominant. Following fig. 8.1, this threshold should be around 1 EeV to 10 EeV. Incorporating these secondary interactions in the simulations would require modelling the energy dependence of the corresponding hadron-air cross-sections.

Large multiplicity events from weak interactions remain an interesting topic for future research, since these may not be in conflict with X_{max} and N_μ measurements and even explain some features of those data.

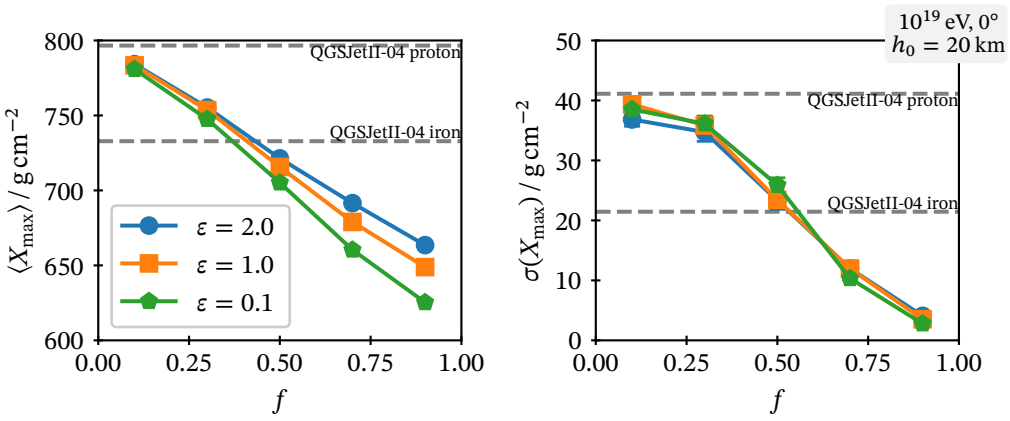


Figure 8.2: Dependence of the two moments of the $X_{\max}$ distribution on f and ε with fixed h_0

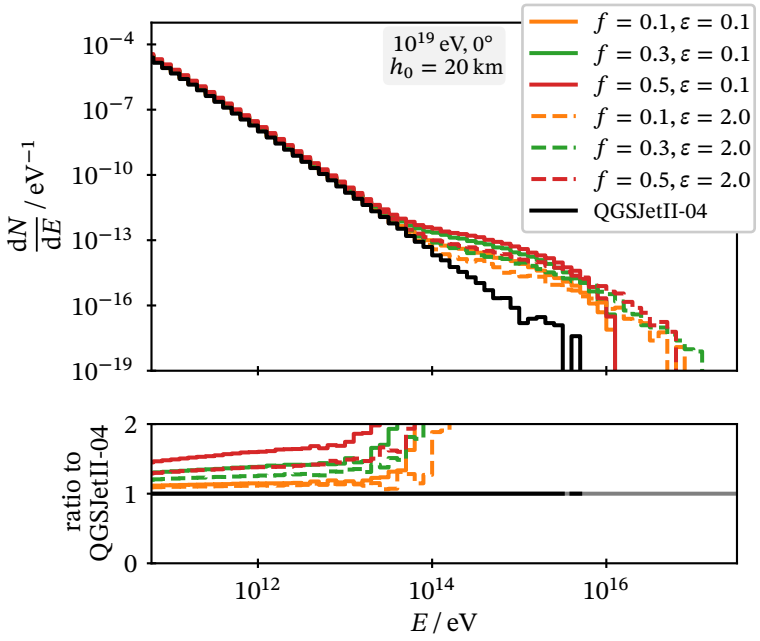


Figure 8.3: Secondary muon energy spectrum at sea level (1036 g cm^{-2})

9 Summary

Accurate simulations of extensive air showers are one of the foundations of modern astroparticle physics. The CORSIKA 8 framework is a next-generation effort to develop a simulation framework to accommodate the community's requirements to face challenges in current and upcoming large-scale experiments.

One of the main pillars of this thesis is the development of the foundations of CORSIKA 8. The overall architecture of CORSIKA 8 offers a great deal of modularity and flexibility, which allow performing simulation studies that have previously been difficult or impossible. Major contributions to the project include:

- Design and implementation of the geometry system, which makes working with geometric objects (points, vectors) convenient. Potential transformations of coordinate systems happen transparently and expressions are close to symbolic notation.
- Design and implementation of the environment system. The environment system allows defining setups in which showers propagate through multiple media, e.g. from air into ice. This is achieved by creating volumes, each of which can have different physical properties assigned, and assembling them in the volume tree.
- Several classes of user-defined processes can be defined, ranging from decay and interaction processes over continuous processes, useful for modelling energy losses or the emission of Cherenkov radiation, to filtering and modifying secondary particles.
- An algorithm to propagate particles is devised and implemented that takes all these conditions into account.
- The complete lineage of particles is kept and can be inspected. Corresponding data structures and algorithms are implemented to keep any overhead negligible.

With the groundwork described here, a collaborative effort has been initiated. Thanks to the growing community of developers, CORSIKA 8 has reached a level of maturity that all essential aspects of air showers are covered. Electromagnetic, hadronic and muonic components of air showers can be simulated, the latter two are the focus of this work.

In the second part of the thesis detailed comparisons between CORSIKA 8 and several other well-known and up-to-date shower simulation codes are performed. The main results are:

- Considering only high-energy hadronic interactions with QGSJet-II.04, agreement is found on a percent level regarding muon energy spectra and lateral distribution up to 1 km. With SIBYLL 2.3d, the predictions agree only on a level of 10 %.

- With both models, CORSIKA 7 predicts significantly more muons at distances greater than 1 km.
- Large differences between CORSIKA 8 and CORSIKA 7 or CONEX are observed when using UrQMD 1.3 as low-energy interaction model down to 1 GeV, having an impact of up to 10 % on the muon energy spectrum.
- Low-energy interactions have a significant impact on the lateral distribution at distances of a few km. The switch from UrQMD to the Hillas Splitting Algorithm increases the muon density at this distance by more than 50 %. Using AIRES with the same implementation of this algorithm increases the density further by another 50 %.

The results show that even among established codes systematic differences remain which are not fully understood. These detailed comparisons have already brought some issues to light that have been remedied in the versions shown here. Further studies are necessary to reach further convergence and identify relevant aspects of muon production.

Using the lineage technique, the phase space of hadronic interactions is quantified with respect to the relevance for muon production and the results compared to predictions of the Heitler-Matthews model. The most important findings are:

- The number of ancestor generations of muons is shown to increase logarithmically with energy in accordance with the Heitler-Matthews model.
- Muons at large lateral distance tend to have fewer ancestor generations than those close to the core. On average about half of the ancestors are pions. The interactions leading to these laterally displaced muons happen more often at low energies than it is the case for muons close to the core.
- Regarding the kinematic distribution of secondaries in hadronic interactions, it is demonstrated explicitly that in interactions at energies $\sqrt{s} \gtrsim 500\,\mathrm{GeV}$ particles emitted in the forward region are most relevant. With decreasing energy, also the central region becomes relevant, in particular for muons at large distances.

Additionally, two modifications of hadronic interactions are studied with respect to their effect on air shower observables, which is greatly facilitated with CORSIKA 8 owing to its flexibility. For the first time, also the effects on the muon production depth are considered:

- On the one hand, a modified extrapolation of hadron-air cross-section affects mainly the longitudinal development, leading to a stretching (decreased cross-sections) or compression (increased cross-sections). The change of $X_{\mathrm{max}}^{(\mu)}$ is shown to be three times larger than for X_{max} because the electromagnetic component quickly decouples from the hadronic one. A novelty is that the cross-sections of pions, kaons and nucleons are modified independently.
- On the other hand, the exchange of π^0 with ρ^0 is investigated. When considering only leading particles, ρ^0 hardly be enhanced further since SIBYLL 2.3d already operates

close to the maximum, so that the muon production can be increased only by a few percent more. The longitudinal development remains essentially unchanged.

Finally, large multiplicity boson production of Higgs bosons in the first interaction of ultra-high energy air showers is studied as exotic scenario with a simple toy model in CORSIKA 8 with the results that

- X_{max} decreases with the amount of Higgs particles produced.
- the number of muons increases at all energies at the same time. The largest impact, however, is the production of prompt muons, which increases the number of muons in the PeV range by a factor of 1000.

The work presented demonstrates some of the advanced capabilities of CORSIKA 8. In conclusion, many simulation studies that previously were prohibitively laborious or impossible to implement have turned into low-hanging fruits. New impulses are given for further directions of research towards solving the muon mystery in extensive air showers.

Bibliography

[1] R. Alves Batista et al., "Open Questions in Cosmic-Ray Research at Ultrahigh Energies", Front. Astron. Space Sci. **6**, 23 (2019) 10.3389/fspas.2019.00023, arXiv:1903.06714 [astro-ph.HE].

[2] L. A. Anchordoqui, "Ultra-High-Energy Cosmic Rays", Phys. Rept. **801**, 1–93 (2019) 10.1016/j.physrep.2019.01.002, arXiv:1807.09645 [astro-ph.HE].

[3] T. K. Gaisser, R. Engel and E. Resconi, *Cosmic Rays and Particle Physics*, 2nd ed. (Cambridge University Press, 2016), ISBN: 9780521016469, 10.1017/CBO9781139192194.

[4] K.-H. Kampert and A. A. Watson, "Extensive Air Showers and Ultra High-Energy Cosmic Rays: A Historical Review", Eur. Phys. J. H **37**, 359–412 (2012) 10.1140/epjh/e2012-30013-x, arXiv:1207.4827 [physics.hist-ph].

[5] D. Heck, J. Knapp, J. N. Capdevielle, G. Schatz and T. Thouw, *CORSIKA: A Monte Carlo code to simulate extensive air showers*, tech. rep. FZKA-6019 (Forschungszentrum Karlsruhe, 1998), 10.5445/IR/270043064.

[6] R. Engel, D. Heck, T. Huege, T. Pierog, M. Reininghaus, F. Riehn, R. Ulrich, M. Unger and D. Veberič, "Towards a Next Generation of CORSIKA: A Framework for the Simulation of Particle Cascades in Astroparticle Physics", Comput. Softw. Big Sci. **3**, 2 (2019) 10.1007/s41781-018-0013-0, arXiv:1808.08226 [astro-ph.IM].

[7] R. Engel, D. Heck and T. Pierog, "Extensive air showers and hadronic interactions at high energy", Ann. Rev. Nucl. Part. Sci. **61**, 467–489 (2011) 10.1146/annurev.nucl.012809.104544.

[8] K.-H. Kampert and M. Unger, "Measurements of the Cosmic Ray Composition with Air Shower Experiments", Astropart. Phys. **35**, 660–678 (2012) 10.1016/j.astropartphys.2012.02.004, arXiv:1201.0018 [astro-ph.HE].

[9] R. Ulrich, R. Engel and M. Unger, "Hadronic Multiparticle Production at Ultra-High Energies and Extensive Air Showers", Phys. Rev. D **83**, 054026 (2011) 10.1103/PhysRevD.83.054026, arXiv:1010.4310 [hep-ph].

[10] A. Aab et al. (Pierre Auger Collaboration), "Direct measurement of the muonic content of extensive air showers between $\mathbf{2 \times 10^{17}}$ and $\mathbf{2 \times 10^{18}}$ eV at the Pierre Auger Observatory", Eur. Phys. J. C **80**, 751 (2020) 10.1140/epjc/s10052-020-8055-y.

[11] P. A. Zyla et al. (Particle Data Group), "Review of Particle Physics", PTEP **2020**, 083C01 (2020) 10.1093/ptep/ptaa104.

[12] R. Alfaro et al. (HAWC), "All-particle cosmic ray energy spectrum measured by the HAWC experiment from 10 to 500 TeV", Phys. Rev. D **96**, 122001 (2017) 10.1103/PhysRevD.96.122001, arXiv:1710.00890 [astro-ph.HE].

[13] Y. A. Fomin, G. B. Khristiansen, G. B. Kulikov, V. G. Pogorely, V. I. Solovjeva, V. P. Sulakov and A. V. Trubitsyn, "Energy Spectrum of Cosmic Rays at Energies of 5×10^{15} eV to 5×10^{17} eV", in 22nd Int. Cosmic Ray Conf. Vol. 2 (1991), p. 85.

[14] M. Amenomori et al. (TIBET III), "The All-particle spectrum of primary cosmic rays in the wide energy range from 10^{14} eV to 10^{17} eV observed with the Tibet-III air-shower array", Astrophys. J. **678**, 1165–1179 (2008) 10.1086/529514, arXiv:0801.1803 [hep-ex].

[15] M. A. K. Glasmacher et al., "The cosmic ray energy spectrum between 10^{14} eV and 10^{16} eV", Astropart. Phys. **10**, 291–302 (1999) 10.1016/S0927-6505(98)00070-X.

[16] T. Antoni et al. (KASCADE), "KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems", Astropart. Phys. **24**, 1–25 (2005) 10.1016/j.astropartphys.2005.04.001, arXiv:astro-ph/0505413.

[17] W. D. Apel et al. (KASCADE Grande), "Kneelike structure in the spectrum of the heavy component of cosmic rays observed with KASCADE-Grande", Phys. Rev. Lett. **107**, 171104 (2011) 10.1103/PhysRevLett.107.171104, arXiv:1107.5885 [astro-ph.HE].

[18] K. Andeen and M. Plum (IceCube), "Cosmic Ray Spectrum and Composition from PeV to EeV from the IceCube Neutrino Observatory", PoS **ICRC2019**, 172 (2021) 10.22323/1.358.0172, arXiv:1908.08139 [astro-ph.HE].

[19] R. U. Abbasi et al. (HiRes), "First observation of the Greisen-Zatsepin-Kuzmin suppression", Phys. Rev. Lett. **100**, 101101 (2008) 10.1103/PhysRevLett.100.101101, arXiv:astro-ph/0703099.

[20] V. Verzi (Pierre Auger), "Measurement of the energy spectrum of ultra-high energy cosmic rays using the Pierre Auger Observatory", PoS **ICRC2019**, 450 (2020) 10.22323/1.358.0450.

[21] D. Ivanov (Telescope Array), "Energy Spectrum Measured by the Telescope Array", PoS **ICRC2019**, 298 (2020) 10.22323/1.358.0298.

[22] K. Greisen, "End to the cosmic ray spectrum?", Phys. Rev. Lett. **16**, 748–750 (1966) 10.1103/PhysRevLett.16.748.

[23] G. T. Zatsepin and V. A. Kuzmin, "Upper limit of the spectrum of cosmic rays", JETP Lett. **4**, 78–80 (1966).

[24] J. Bellido (Pierre Auger), "Depth of maximum of air-shower profiles at the Pierre Auger Observatory: Measurements above $10^{17.2}$ eV and Composition Implications", PoS **ICRC2017**, 506 (2018) 10.22323/1.301.0506.

[25] A. Abada et al. (FCC), "FCC-hh: The Hadron Collider: Future Circular Collider Conceptual Design Report Volume 3", Eur. Phys. J. ST **228**, 755–1107 (2019) 10.1140/epjst/e2019-900087-0.

[26] H. Bethe and W. Heitler, "On the Stopping of fast particles and on the creation of positive electrons", Proc. Roy. Soc. Lond. A **146**, 83–112 (1934) 10.1098/rspa.1934.0140.

[27] W. Heitler, *The quantum theory of radiation*, 3rd ed. (Oxford Clarendon Press, 1954).

[28] J. R. Hörandel, "Cosmic rays from the knee to the second knee: 10^4 to 10^{18} eV", Mod. Phys. Lett. A **22**, 1533–1552 (2007) 10.1142/S0217732307024139, arXiv:astro-ph/0611387 [astro-ph].

[29] S. Klein, "Suppression of Bremsstrahlung and pair production due to environmental factors", Rev. Mod. Phys. **71**, 1501–1538 (1999) 10.1103/RevModPhys.71.1501, arXiv:hep-ph/9802442.

[30] L. D. Landau and I. Pomeranchuk, "Limits of applicability of the theory of bremsstrahlung electrons and pair production at high-energies", Dokl. Akad. Nauk Ser. Fiz. **92**, 535–536 (1953).

[31] A. B. Migdal, "Bremsstrahlung and pair production in condensed media at high-energies", Phys. Rev. **103**, 1811–1820 (1956) 10.1103/PhysRev.103.1811.

[32] M. Ter-Mikaelian, Zh. Eksp. Teor. Fiz. **25**, 289 (1953).

[33] S. Meighen-Berger, A. Fedynitch and M. Huber, "EmCa – Electromagnetic-Cascades Simulation Package", (2019), arXiv:1907.06924 [astro-ph.HE].

[34] A. N. Cillis, H. Fanchiotti, C. A. Garcia Canal and S. J. Sciutto, "Influence of the LPM effect and dielectric suppression on particle air showers", Phys. Rev. D **59**, 113012 (1999) 10.1103/PhysRevD.59.113012, arXiv:astro-ph/9809334.

[35] R. Alkofer and J. Greensite, "Quark Confinement: The Hard Problem of Hadron Physics", J. Phys. G **34**, S3 (2007) 10.1088/0954-3899/34/7/S02, arXiv:hep-ph/0610365.

[36] T. Pierog, "Hadronic Interactions and Air Showers: Where Do We Stand?", EPJ Web Conf. **208**, 02002 (2019) 10.1051/epjconf/201920802002.

[37] F. Riehn, R. Engel, A. Fedynitch, T. K. Gaisser and T. Stanev, "Hadronic interaction model Sibyll 2.3d and extensive air showers", Phys. Rev. D **102**, 063002 (2020) 10.1103/PhysRevD.102.063002, arXiv:1912.03300 [hep-ph].

[38] E.-J. Ahn, R. Engel, T. K. Gaisser, P. Lipari and T. Stanev, "Cosmic ray interaction event generator SIBYLL 2.1", Phys. Rev. D **80**, 094003 (2009) 10.1103/PhysRevD.80.094003, arXiv:0906.4113 [hep-ph].

[39] S. Ostapchenko, "Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: I. QGSJET-II model", Phys. Rev. D **83**, 014018 (2011) 10.1103/PhysRevD.83.014018, arXiv:1010.1869 [hep-ph].

[40] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko and K. Werner, "EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider", Phys. Rev. C **92**, 034906 (2015) 10.1103/PhysRevC.92.034906, arXiv:1306.0121 [hep-ph].

[41] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen and P. Z. Skands, "An introduction to PYTHIA 8.2", Comput. Phys. Commun. **191**, 159–177 (2015) 10.1016/j.cpc.2015.01.024, arXiv:1410.3012 [hep-ph].

[42] T. Sjöstrand and M. Utheim, "Hadron Interactions for Arbitrary Energies and Species, with Applications to Cosmic rays", (2021), arXiv:2108.03481 [hep-ph].

[43] D. d'Enterria, T. Pierog and G. Sun, "Impact of QCD jets and heavy-quark production in cosmic-ray proton atmospheric showers up to 10^{20} eV", Astrophys. J. **874**, 152 (2019) 10.3847/1538-4357/ab01e2, arXiv:1809.06406 [astro-ph.HE].

[44] A. Ferrari, P. R. Sala, A. Fasso and J. Ranft, *FLUKA: A multi-particle transport code (Program version 2005)*, tech. rep. CERN-2005-010, SLAC-R-773, INFN-TC-05-11, CERN-2005-10 (2005), 10.2172/877507.

[45] G. Battistoni et al., "Overview of the FLUKA code", Annals Nucl. Energy **82**, 10–18 (2015) 10.1016/j.anucene.2014.11.007.

[46] M. Bleicher et al., "Relativistic hadron hadron collisions in the ultrarelativistic quantum molecular dynamics model", J. Phys. G **25**, 1859–1896 (1999) 10.1088/0954-3899/25/9/308, arXiv:hep-ph/9909407 [hep-ph].

[47] S. A. Bass et al., "Microscopic models for ultrarelativistic heavy ion collisions", Prog. Part. Nucl. Phys. **41**, 255–369 (1998) 10.1016/S0146-6410(98)00058-1, arXiv:nucl-th/9803035 [nucl-th].

[48] A. M. Hillas, "Two interesting techniques for Monte-Carlo simulation of very high energy hadron cascades", in Proc. 17th Int. Cosmic Ray Conf. Vol. 8 (1981), p. 193.

[49] J. C. Collins, "Light cone variables, rapidity and all that", (1997), arXiv:hep-ph/9705393.

[50] R. Sahoo, *Relativistic Kinematics*, 2016, arXiv:1604.02651 [nucl-ex].

[51] R. P. Feynman, "Very high-energy collisions of hadrons", Phys. Rev. Lett. **23**, 1415–1417 (1969) 10.1103/PhysRevLett.23.1415.

[52] J. Matthews, "A Heitler model of extensive air showers", Astropart. Phys. **22**, 387–397 (2005) 10.1016/j.astropartphys.2004.09.003.

[53] J. Albrecht et al., "The Muon Puzzle in cosmic-ray induced air showers and its connection to the Large Hadron Collider", (2021), arXiv:2105.06148 [astro-ph.HE].

[54] J. Aichelin et al., "Future physics opportunities for high-density QCD at the LHC with heavy-ion and proton beams", CERN Yellow Rep. Monogr. **7**, edited by Z. Citron, A. Dainese, M. Mangano, A. B. Meyer, A. Nisati, G. Salam and M. A. Vesterinen, 1159–1410 (2019) 10.23731/CYRM-2019-007.1159, arXiv:1812.06772 [hep-ph].

[55] H. P. Dembinski, R. Ulrich and T. Pierog, "Future Proton-Oxygen Beam Collisions at the LHC for Air Shower Physics", PoS **ICRC2019**, 235 (2020) 10.22323/1.358.0235.

[56] L. Cazon, R. A. Vazquez and E. Zas, "Depth development of extensive air showers from muon time distributions", Astropart. Phys. **23**, 393–409 (2005) 10.1016/j.astropartphys.2005.01.009, arXiv:astro-ph/0412338.

[57] S. Andringa, L. Cazon, R. Conceição and M. Pimenta, "The muonic longitudinal shower profiles at production", Astropart. Phys. **35**, 821–827 (2012) 10.1016/j.astropartphys.2012.03.010, arXiv:1111.1424 [hep-ph].

[58] T. Abu-Zayyad et al. (HiRes, MIA), "Evidence for Changing of Cosmic Ray Composition between 10^{17} eV and 10^{18} eV from Multicomponent Measurements", Phys. Rev. Lett. **84**, 4276–4279 (2000) 10.1103/PhysRevLett.84.4276, arXiv:astro-ph/9911144.

[59] A. Aab et al. (Pierre Auger), "Muons in Air Showers at the Pierre Auger Observatory: Mean Number in Highly Inclined Events", Phys. Rev. D **91**, 032003 (2015) 10.1103/PhysRevD.91.032003, arXiv:1408.1421 [astro-ph.HE].

[60] H. P. Dembinski et al. (EAS-MSU, IceCube, KASCADE-Grande, NEVOD-DECOR, Pierre Auger, SUGAR, Telescope Array, Yakutsk EAS Array), "Report on Tests and Measurements of Hadronic Interaction Properties with Air Showers", EPJ Web Conf. **210**, 02004 (2019) 10.1051/epjconf/201921002004, arXiv:1902.08124 [astro-ph.HE].

[61] L. Cazon (EAS-MSU, IceCube, KASCADE Grande, NEVOD-DECOR, Pierre Auger, SUGAR, Telescope Array, Yakutsk EAS Array), "Working Group Report on the Combined Analysis of Muon Density Measurements from Eight Air Shower Experiments", PoS **ICRC2019**, 214 (2020) 10.22323/1.358.0214, arXiv:2001.07508 [astro-ph.HE].

[62] D. Soldin (EAS-MSU, IceCube, KASCADE-Grande, NEVOD-DECOR, Pierre Auger, SUGAR, Telescope Array, Yakutsk EAS Array), "Update on the Combined Analysis of Muon Measurements from Nine Air Shower Experiments", PoS **ICRC2021**, 349 (2021) 10.22323/1.395.0349, arXiv:2108.08341 [astro-ph.HE].

[63] A. G. Bogdanov, D. M. Gromushkin, R. P. Kokoulin, G. Mannocchi, A. A. Petrukhin, O. Saavedra, G. Trinchero, D. V. Chernov, V. V. Shutenko and I. I. Yashin, "Investigation of the properties of the flux and interaction of ultrahigh-energy cosmic rays by the method of local-muon-density spectra", Phys. Atom. Nucl. **73**, 1852–1869 (2010) 10.1134/S1063778810110074.

[64] A. G. Bogdanov, R. P. Kokoulin, G. Mannocchi, A. A. Petrukhin, O. Saavedra, V. V. Shutenko, G. Trinchero and I. I. Yashin, "Investigation of very high energy cosmic rays by means of inclined muon bundles", Astropart. Phys. **98**, 13–20 (2018) 10.1016/j.astropartphys.2018.01.003.

[65] A. Aab et al. (Pierre Auger), "Testing Hadronic Interactions at Ultrahigh Energies with Air Showers Measured by the Pierre Auger Observatory", Phys. Rev. Lett. **117**, 192001 (2016) 10.1103/PhysRevLett.117.192001, arXiv:1610.08509 [hep-ex].

[66] J. A. Bellido, R. W. Clay, N. N. Kalmykov, I. S. Karpikov, G. I. Rubtsov, S. V. Troitsky and J. Ulrichs, "Muon content of extensive air showers: comparison of the energy spectra obtained by the Sydney University Giant Air-shower Recorder and by the Pierre Auger Observatory", Phys. Rev. D **98**, 023014 (2018) 10.1103/PhysRevD.98.023014, arXiv:1803.08662 [astro-ph.HE].

[67] A. Aab et al. (Pierre Auger), "Measurement of the Fluctuations in the Number of Muons in Extensive Air Showers with the Pierre Auger Observatory", Phys. Rev. Lett. **126**, 152002 (2021) 10.1103/PhysRevLett.126.152002, arXiv:2102.07797 [hep-ex].

[68] R. Abbasi et al. (IceCube), "Density of GeV Muons Measured with IceTop", PoS **ICRC2021**, 342 (2021) 10.22323/1.395.0342, arXiv:2107.09583 [astro-ph.HE].

[69] F. Gesualdi, A. D. Supanitsky and A. Etchegoyen, "Muon deficit in air shower simulations estimated from AGASA muon measurements", Phys. Rev. D **101**, 083025 (2020) 10.1103/PhysRevD.101.083025, arXiv:2003.03385 [astro-ph.HE].

[70] F. Gesualdi, H. Dembinski, K. Shinozaki, D. A. Supanitsky, T. Pierog, L. Cazon, D. Soldin and R. Conceição, "On the muon scale of air showers and its application to the AGASA data", PoS **ICRC2021**, 473 (2021) 10.22323/1.395.0473, arXiv:2108.04824 [astro-ph.HE].

[71] A. Aab et al. (Pierre Auger), "Muons in Air Showers at the Pierre Auger Observatory: Measurement of Atmospheric Production Depth", Phys. Rev. D **90**, 012012 (2014) 10.1103/PhysRevD.90.012012, arXiv:1407.5919 [hep-ex].

[72] T. Pierog and K. Werner, "Muon Production in Extended Air Shower Simulations", Phys. Rev. Lett. **101**, 171101 (2008) 10.1103/PhysRevLett.101.171101, arXiv:astro-ph/0611311.

[73] S. Baur, H. Dembinski, M. Perlin, T. Pierog, R. Ulrich and K. Werner, "Core-corona effect in hadron collisions and muon production in air showers", (2019), arXiv:1902.09265 [hep-ph].

[74] I. Allekotte et al. (Pierre Auger), "The Surface Detector System of the Pierre Auger Observatory", Nucl. Instrum. Meth. A **586**, 409–420 (2008) 10.1016/j.nima.2007.12.016, arXiv:0712.2832 [astro-ph].

[75] A. Aab et al. (Pierre Auger), "Prototype muon detectors for the AMIGA component of the Pierre Auger Observatory", JINST **11**, P02012 (2016) 10.1088/1748-0221/11/02/P02012, arXiv:1605.01625 [physics.ins-det].

[76] J. Abraham et al. (Pierre Auger), "The Fluorescence Detector of the Pierre Auger Observatory", Nucl. Instrum. Meth. A **620**, 227–251 (2010) 10.1016/j.nima.2010.04.023, arXiv:0907.4282 [astro-ph.IM].

[77] T. Huege, "Radio detection of cosmic ray air showers in the digital era", Phys. Rept. **620**, 1–52 (2016) 10.1016/j.physrep.2016.02.001, arXiv:1601.07426 [astro-ph.IM].

[78] T. Bergmann, R. Engel, D. Heck, N. N. Kalmykov, S. Ostapchenko, T. Pierog, T. Thouw and K. Werner, "One-dimensional Hybrid Approach to Extensive Air Shower Simulation", Astropart. Phys. **26**, 420–432 (2007) 10.1016/j.astropartphys.2006.08.005, arXiv:astro-ph/0606564 [astro-ph].

[79] R. E. Caflisch, "Monte Carlo and quasi-Monte Carlo methods", Acta Numer. **7**, 1–49 (1998) 10.1017/S0962492900002804.

[80] N. Metropolis and S. Ulam, "The Monte Carlo Method", J. Am. Stat. Assoc. **44**, 335–341 (1949) 10.1080/01621459.1949.10483310.

[81] E. Horowitz, "The Art of the Algorithmist", in Proceedings of the 1976 annual conference, ACM '76 (1976), pp. 442–444, ISBN: 9781450374897, 10.1145/800191.805636.

[82] K. Miura, "EGS4V: Vectorization of the Monte Carlo cascade shower simulation code EGS4", Comput. Phys. Commun. **45**, 127–136 (1987) 10.1016/0010-4655(87)90147-0.

[83] H.-J. Drescher and G. R. Farrar, "Air shower simulations in a hybrid approach using cascade equations", Phys. Rev. D **67**, 116001 (2003) 10.1103/PhysRevD.67.116001, arXiv:astro-ph/0212018 [astro-ph].

[84] D. Heck and J. Knapp, *Upgrade of the Monte Carlo code CORSIKA to simulate extensive air showers with energies* $> 10^{20}$ *eV*, tech. rep. FZKA-6097 (Forschungszentrum Karlsruhe, 1998), 10.5445/IR/270043705.

[85] G. Amadio et al., "GeantV: Results from the Prototype of Concurrent Vector Particle Transport Simulation in HEP", Comput. Softw. Big Sci. **5**, 3 (2021) 10.1007/s41781-020-00048-6, arXiv:2005.00949 [physics.comp-ph].

[86] M. Kobal (Pierre Auger), "A thinning method using weight limitation for air-shower simulations", Astropart. Phys. **15**, 259–273 (2001) 10.1016/S0927-6505(00)00158-4.

[87] R. R. Wilson, "Monte Carlo Study of Shower Production", Phys. Rev. **86**, 261–269 (1952) 10.1103/PhysRev.86.261.

[88] S. J. Sciutto, *AIRES: A System for air shower simulations. User's guide and reference manual. Version 2.2.0*, 1999, 10.13140/RG.2.2.12566.40002, arXiv:astro-ph/9911331 [astro-ph].

[89] S. J. Sciutto, *AIRES: A minimum document*, tech. rep. GAP-97-029 (1997), http://www-f9.ijs.si/auger/private/GAP_Notes/GAP_97_029.ai.

[90] M. Tueros and S. Sciutto, "TIERRAS: A package to simulate high energy cosmic ray showers underground, underwater and under-ice", Comput. Phys. Commun. **181**, 380–392 (2010) 10.1016/j.cpc.2009.09.022.

[91] J. Alvarez-Muñiz, W. R. Carvalho Jr. and E. Zas, "Monte Carlo simulations of radio pulses in atmospheric showers using ZHAireS", Astropart. Phys. **35**, 325–341 (2012) 10.1016/j.astropartphys.2011.10.005, arXiv:1107.1189 [astro-ph.HE].

[92] K. Kasahara, "Introduction to Cosmos and some Relevance to Ultra High Energy Cosmic Ray Air Showers", in Proc. 24th Int Cosmic Ray Conf. (1995), p. 399.

[93] K. Kasahara, *The Users Manual of Cosmos*, 2002, http://cosmos.icrr.u-tokyo.ac.jp/cosmosHome/man.pdf.

[94] T. Sako, T. Fuji, K. Kasahara, H. Menjo, N. Sakaki, N. Sakurai, A. Taketa and Y. Tameda, "COSMOS X as a general purpose air shower simulation tool", PoS **ICRC2021**, 431 (2021) 10.22323/1.395.0431.

[95] P. K. F. Grieder, "ASICO, A new system of programs to simulate, analyze and correlate air showers and all of their observables within the same unique frame of completeness, to investigate ultrahigh-energy interactions and astrophysical aspects", in Proc. 16th Int. Cosmic Ray Conf. (1979), pp. 161–166.

[96] W. R. Nelson, H. Hirayama and D. W. O. Rogers, *The EGS4 Code System*, tech. rep. SLAC-265 (Stanford Linear Accelerator Center, 1985).

[97] I. Kawrakow, E. Mainegra-Hing, D. Rogers, F. Tessier and B. Walters, *The EGSnrc code system: Monte Carlo simulation of electron and photon transport*, tech. rep. PIRS-701 (National Research Council Canada, 2018).

[98] H. Hirayama, Y. Namito, A. F. Bielajew, S. J. Wilderman and W. R. Nelson, *The EGS5 code system*, tech. rep. SLAC-R-730 (Stanford Linear Accelerator Center, 2005).

[99] H. J. Gils, D. Heck, J. Oehlschlaeger, G. Schatz, T. Thouw and A. Merkel, "A multi-transputer system for parallel Monte Carlo simulations of extensive air showers", Comput. Phys. Commun. **56**, 105–113 (1989) 10.1016/0010-4655(89)90011-8.

[100] J. N. Capdevielle, P. Gabriel, H. J. Gils, P. Grieder, D. Heck, J. Knapp, H. J. Mayer, J. Oehlschläger, H. Rebel, G. Schatz and T. Thouw, *The Karlsruhe Extensive Air Shower Simulation Code CORSIKA*, tech. rep. KfK-4998 (Kernforschungszentrum Karlsruhe, 1992), 10.5445/IR/270033168.

[101] T. Huege, M. Ludwig and C. W. James, "Simulating radio emission from air showers with CoREAS", AIP Conf. Proc. **1535**, 128 (2013) 10.1063/1.4807534, arXiv:1301.2132 [astro-ph.HE].

[102] K. Bernlöhr, "Impact of atmospheric parameters on the atmospheric Cherenkov technique", Astropart. Phys. **12**, 255–268 (2000) 10.1016/S0927-6505(99)00093-6, arXiv:astro-ph/9908093.

[103] T. Pierog, R. Engel, D. Heck and R. Ulrich, "3D Hybrid Air Shower Simulation in CORSIKA", in 32nd Int. Cosmic Ray Conf. Vol. 2 (2011), p. 222, 10.7529/ICRC2011/V02/1170.

[104] A. Fedynitch, R. Engel, T. K. Gaisser, F. Riehn and T. Stanev, "Calculation of conventional and prompt lepton fluxes at very high energy", EPJ Web Conf. **99**, 08001 (2015) 10.1051/epjconf/20159908001, arXiv:1503.00544 [hep-ph].

[105] T. Kozynets, A. Fedynitch and D. J. Koskinen, "A Numerical Approach to Angular Distributions in Hadronic Cascades", PoS **ICRC2021**, 1209 (2021) 10.22323/1.395.1209.

[106] G. Lente and K. Ősz, "Barometric formulas: various derivations and comparisons to environmentally relevant observations", ChemTexts **6**, 13 (2020) 10.1007/s40828-020-0111-6.

[107] J. Cruz Moreno and S. Sciutto, "Characterization of the atmospheric depth profile using the ground-level temperature: The case of Malargüe, Argentina", Eur. Phys. J. Plus **128**, 104 (2013) 10.1140/epjp/i2013-13104-3.

[108] National Oceanic and Atmospheric Administration, National Aeronautics and Space Administration and United States Air Force, *U.S. Standard Atmosphere, 1976*, NOAA - SIT 76-1562 (1976).

[109] B. Keilhauer, J. Blümer, R. Engel, H. O. Klages and M. Risse, "Impact of varying atmospheric profiles on extensive air shower observation: Atmospheric density and primary mass reconstruction", Astropart. Phys. **22**, 249–261 (2004) 10.1016/j.astropartphys.2004.08.004, arXiv:astro-ph/0405048.

[110] C. Forbes, M. Evans, N. Hastings and B. Peacock, *Statistical distributions*, 4th ed. (Wiley, 2011), ISBN: 9780470390634, 10.1002/9780470627242.

[111] S. Agostinelli et al. (GEANT4), "GEANT4: A Simulation toolkit", Nucl. Instrum. Meth. A **506**, 250–303 (2003) 10.1016/S0168-9002(03)01368-8.

[112] F. Schröder and D. Heck (HEGRA), "Observations at large zenith angles", AIP Conf. Proc. **515**, 343 (2000) 10.1063/1.1291391.

[113] I. Gradshteyn and I. Ryzhik, *Table of integrals, series, and products*, trans. by A. Jeffrey and D. Zwillinger, 7th ed. (Academic Press, 2007), ISBN: 9780123736376, 10.1016/C2009-0-22516-5.

[114] V. Chernatkin, "Simulation des gerbes dues aux rayons cosmiques à très hautes énergies", PhD thesis (Université de Nantes, 2005), HDL: 10068/774839.

[115] M. Stadelmaier, R. Engel, D. Schmidt and D. Veberič, *A Useful Model of the Atmosphere*, tech. rep. GAP-2021-10 (2021).

[116] J. C. Glaser, "Absolute energy calibration of the Pierre Auger observatory using radio emission of extensive air showers", PhD thesis (RWTH Aachen, 2017), 10.18154/RWTH-2017-02960.

[117] D. Heck, *The CURVED version of the air shower simulation program CORSIKA*, tech. rep. FZKA-6954 (2004), 10.5445/IR/270057207.

[118] K.-H. Kampert, J. Kulbartz, L. Maccione, N. Nierstenhoefer, P. Schiffer, G. Sigl and A. R. van Vliet, "CRPropa 2.0 – a Public Framework for Propagating High Energy Nuclei, Secondary Gamma Rays and Neutrinos", Astropart. Phys. **42**, 41–51 (2013) 10.1016/j.astropartphys.2012.12.001, arXiv:1206.3132 [astro-ph.IM].

[119] T. Stanev, R. Engel, A. Mücke, R. J. Protheroe and J. P. Rachen, "Propagation of ultrahigh-energy protons in the nearby universe", Phys. Rev. D **62**, 093005 (2000) 10.1103/PhysRevD.62.093005, arXiv:astro-ph/0003484.

[120] I. Kawrakow, "Accurate condensed history Monte Carlo simulation of electron transport. I. EGSnrc, the new EGS4 version", Med. Phys. **27**, 485–498 (2000) 10.1118/1.598917.

[121] D. Chirkin and W. Rhode, "Muon Monte Carlo: A High-precision tool for muon propagation through matter", (2004), arXiv:hep-ph/0407075.

[122] J. H. Koehne, K. Frantzen, M. Schmitz, T. Fuchs, W. Rhode, D. Chirkin and J. Becker Tjus, "PROPOSAL: A tool for propagation of charged leptons", Comput. Phys. Commun. **184**, 2070–2090 (2013) 10.1016/j.cpc.2013.04.001.

[123] K. Bernlöhr, *Atmosphere implementation and approximations*, talk given at the CORSIKA Cosmic Ray Simulation Workshop, 2019, https://indico.scc.kit.edu/event/529/contributions/5507/.

[124] J. Dormand and P. Prince, "A family of embedded Runge-Kutta formulae", J. Comput. Appl. Math. **6**, 10.1016/0771-050x(80)90013-3 (1980) 10.1016/0771-050x(80)90013-3.

[125] P. Bogacki and L. Shampine, "A 3(2) pair of Runge–Kutta formulas", Appl. Math Lett. **2**, 321–325 (1989) 10.1016/0893-9659(89)90079-7.

[126] P. Virtanen et al., "SciPy 1.0 – Fundamental Algorithms for Scientific Computing in Python", Nature Meth. **17**, 261 (2020) 10.1038/s41592-019-0686-2, arXiv:1907.10121 [cs.MS].

[127] H. P. Dembinski, L. Nellen, M. Reininghaus and R. Ulrich (CORSIKA 8), "Technical Foundations of CORSIKA 8: New Concepts for Scientific Computing", PoS **ICRC2019**, 236 (2020) 10.22323/1.358.0236.

[128] A. A. Alves, A. Poctarev and R. Ulrich, "Counter-based pseudorandom number generators for CORSIKA 8 - A multi-thread friendly approach", EPJ Web Conf. **251**, 03039 (2021) 10.1051/epjconf/202125103039.

[129] M. Reininghaus and R. Ulrich, "CORSIKA 8 – Towards a modern framework for the simulation of extensive air showers", EPJ Web Conf. **210**, 02011 (2019) 10.1051/epjconf/201921002011, arXiv:1902.02822 [astro-ph.IM].

[130] A. A. Alves, M. Reininghaus, A. Schmidt, R. Prechelt and R. Ulrich (CORSIKA), "CORSIKA 8 - A novel high-performance computing tool for particle cascade Monte Carlo simulations", EPJ Web Conf. **251**, 03038 (2021) 10.1051/epjconf/202125103038.

[131] J. Albrecht et al. (HEP Software Foundation), "A Roadmap for HEP Software and Computing R&D for the 2020s", Comput. Softw. Big Sci. **3**, 7 (2019) 10.1007/s41781-018-0018-8, arXiv:1712.06982 [physics.comp-ph].

[132] D. Heck, *Air shower simulation with CORSIKA at arbitrary direction of incidence*, tech. rep. FZKA-7254 (Forschungszentrum Karlsruhe, 2006), 10.5445/IR/270066237.

[133] D. Heck and R. Engel, *The EHISTORY Option of the Air-Shower Simulation Program CORSIKA*, tech. rep. FZKA-7495 (Forschungszentrum Karlsruhe, 2009), 10.5445/IR/270078292.

[134] L. Arrabito, K. Bernlöhr, J. Bregeon, M. Carrère, A. Khattabi, P. Langlois, D. Parello and G. Revy, "Optimizing Cherenkov Photons Generation and Propagation in CORSIKA for CTA Monte–Carlo Simulations", Comput. Softw. Big Sci. **4**, 9 (2020) 10.1007/s41781-020-00042-y, arXiv:2006.14927 [astro-ph.IM].

[135] ISO/IEC JTC1/SC22/WG5, *Information Technology — Programming languages — Fortran — Units of measure for numerical quantities*, tech. rep. N2113 (2016).

[136] Mars Climate Orbiter Mishap Investigation Board, *Phase I report*, 1999.

[137] N. Gehani, "Units of measure as a data attribute", Comput. Lang. **2**, 93–111 (1977) 10.1016/0096-0551(77)90010-8.

[138] S. McKeever, O. Bennich-Björkman and O.-A. Salah, "Unit of measurement libraries, their popularity and suitability", Softw. Pract. Exper. **51**, 711–734 (2021) 10.1002/spe.2926.

[139] Z. D. Umrigar, "Fully Static Dimensional Analysis with C++", SIGPLAN Not. **29**, 135–139 (1994) 10.1145/185009.185036.

[140] M. Moene, *PhysUnits C++11*, https://github.com/martinmoene/PhysUnits-CT-Cpp11.

[141] Á. P. Raposo, "The algebraic structure of quantity calculus", Measur. Sci. Rev. **18**, 147–157 (2018) 10.1515/msr-2017-0021, arXiv:1611.01502 [math.GM].

[142] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*, Student Economy Edition (Westview Press, 2016), ISBN: 9780813350196.

[143] S. Argirò, S. L. C. Barroso, J. Gonzalez, L. Nellen, T. C. Paul, T. A. Porter, L. Prado Jr., M. Roth, R. Ulrich and D. Veberič, "The Offline Software Framework of the Pierre Auger Observatory", Nucl. Instrum. Meth. A **580**, 1485–1496 (2007) 10.1016/j.nima.2007.07.010, arXiv:0707.1652 [astro-ph].

[144] G. Guennebaud, B. Jacob et al., *Eigen*, version 3, https://eigen.tuxfamily.org.

[145] I. R. Shafarevich and A. O. Remizov, *Linear algebra and geometry* (Springer, 2013), pp. 289–317, ISBN: 9783642309946, 10.1007/978-3-642-30994-6.

[146] D. Nesteruk, *Design Patterns in Modern C++: Reusable Approaches for Object-Oriented Software Design* (Apress, 2018), ISBN: 9781484236031, 10.1007/978-1-4842-3603-1.

[147] A. Schmidt, "Implementierung von Magnetfeldern in CORSIKA 8", B.Sc. thesis (KIT, 2020), https://gitlab.ikp.kit.edu/AirShowerPhysics/corsika/-/wikis/CORSIKA-Theses.

[148] A. Cillis and S. Sciutto, "Geomagnetic field and air shower simulations", (1997), arXiv:astro-ph/9712345.

[149] A. M. Hillas, "Shower simulation: lessons from MOCCA", Nucl. Phys. B Proc. Suppl. **52**, 29–42 (1997) 10.1016/S0920-5632(96)00847-X.

[150] C. Meurer, J. Blümer, R. Engel, A. Haungs and M. Roth, "Muon production in extensive air showers and its relation to hadronic interactions", Czech. J. Phys. **56**, A211 (2006) 10.1007/s10582-006-0156-9, arXiv:astro-ph/0512536 [astro-ph].

[151] M. Dunsch, J. Soedingrekso, A. Sandrock, M. Meier, T. Menne and W. Rhode, "Recent Improvements for the Lepton Propagator PROPOSAL", Comput. Phys. Commun. **242**, 132–144 (2019) 10.1016/j.cpc.2019.03.021, arXiv:1809.07740 [hep-ph].

[152] J.-M. Alameddine, J. Soedingrekso, A. Sandrock, M. Sackel and W. Rhode, "PROPOSAL: A library to propagate leptons and high energy photons", J. Phys. Conf. Ser. **1690**, 012021 (2020) 10.1088/1742-6596/1690/1/012021.

[153] J.-M. Alameddine, J. Alvarez-Muñiz, J. Ammerman-Yebra, L. Bollmann, W. Rhode, M. Sackel, A. Sandrock, J. Soedingrekso and E. Zas (CORSIKA 8), "Electromagnetic Shower Simulation for CORSIKA 8", PoS **ICRC2021**, 428 (2021) 10.22323/1.395.0428.

[154] N. Karastathis, R. Prechelt, T. Huege and J. Ammerman-Yebra (CORSIKA 8), "Simulations of radio emission from air showers with CORSIKA 8", PoS **ICRC2021**, 427 (2021) 10.22323/1.395.0427.

[155] D. Baack and W. Rhode (CORSIKA 8), "GPU Accelerated optical light propagation in CORSIKA8", PoS **ICRC2021**, 705 (2021) 10.22323/1.395.0705.

[156] M. Carrère, L. Arrabito, J. Bregeon, D. Parello, P. Langlois and G. Vasileiadis, "A C++ Cherenkov photons simulation in CORSIKA 8", EPJ Web Conf. **251**, 03011 (2021) 10.1051/epjconf/202125103011.

[157] The HDF Group, *Hierarchical Data Format, version 5*, (1997-2021) https://www.hdfgroup.org/HDF5/.

[158] M. Folk, G. Heber, Q. Koziol, E. Pourmal and D. Robinson, "An Overview of the HDF5 Technology Suite and Its Applications", in Proceedings of the EDBT/ICDT 2011 Workshop on Array Databases (2011), pp. 36–47, 10.1145/1966895.1966900.

[159] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework", Nucl. Instrum. Meth. A **389**, 81–86 (1997) 10.1016/S0168-9002(97)00048-X.

[160] S.-A. Dragly, M. Hobbi Mobarhan, M. E. Lepperød, S. Tennøe, M. Fyhn, T. Hafting and A. Malthe-Sørenssen, "Experimental Directory Structure (Exdir): an alternative to HDF5 without introducing a new file format", Front. Neuroinform. **12**, 16 (2018) 10.3389/fninf.2018.00016.

[161] Apache Software Foundation, *Apache parquet*, https://parquet.apache.org/.

[162] D. Vohra, "Apache Parquet", in *Practical Hadoop Ecosystem: A Definitive Guide to Hadoop-Related Frameworks and Tools* (Apress, 2016), pp. 325–335, 10.1007/978-1-4842-2199-0_8.

[163] H. P. Dembinski, J. Pivarski and H. Schreiner, "Recent developments in histogram libraries", EPJ Web Conf. **245**, 05014 (2020) 10.1051/epjconf/202024505014.

[164] C. R. Harris et al., "Array programming with NumPy", Nature **585**, 357–362 (2020) 10.1038/s41586-020-2649-2, arXiv:2006.10256 [cs.MS].

[165] T. Pierog, D. Heck and R. Ulrich, *CORSIKA*, version 7.7410, 2021, 10.5281/zenodo.5246070.

[166] D. Melo, M. Reininghaus, F. Riehn and R. Ulrich (CORSIKA 8), "First results of the CORSIKA 8 air shower simulation framework", PoS **ICRC2019**, 399 (2019) 10.22323/1.358.0399.

[167] R. Ulrich, F. Riehn, M. Reininghaus, A. Fedynitch and T. Pierog, "Hadron cascades in CORSIKA 8", PoS **ICRC2021**, 474 (2021) 10.22323/1.395.0474.

[168] J. Weil et al., "Particle production and equilibrium properties within a new hadron transport approach for heavy-ion collisions", Phys. Rev. C **94**, 054905 (2016) 10.1103/PhysRevC.94.054905, arXiv:1606.06642 [nucl-th].

[169] M. Reininghaus, R. Ulrich and T. Pierog, "Air shower genealogy for muon production", PoS **ICRC2021**, 463 (2021) 10.22323/1.395.0463, arXiv:2108.03266 [astro-ph.HE].

[170] L. Cazon, R. Conceição and F. Riehn, "Probing the energy spectrum of hadrons in proton air interactions at ultrahigh energies through the fluctuations of the muon content of extensive air showers", Phys. Lett. B **784**, 68–76 (2018) 10.1016/j.physletb.2018.07.026, arXiv:1803.05699 [hep-ph].

[171] L. Cazon, R. Conceição, M. A. Martins and F. Riehn, "Constraining the energy spectrum of neutral pions in ultra-high-energy proton-air interactions", Phys. Rev. D **103**, 022001 (2021) 10.1103/PhysRevD.103.022001, arXiv:2006.11303 [astro-ph.HE].

[172] H.-J. Drescher, M. Bleicher, S. Soff and H. Stoecker, "Model dependence of lateral distribution functions of high energy cosmic ray air showers", Astropart. Phys. **21**, 87–94 (2004) 10.1016/j.astropartphys.2003.10.007, arXiv:astro-ph/0307453 [astro-ph].

[173] I. C. Mariş, R. Engel, X. Garrido, A. Haungs, M. Roth, R. Ulrich and M. Unger, "Influence of Low Energy Hadronic Interactions on Air-shower Simulations", Nucl. Phys. B Proc. Suppl. **196**, 86–89 (2009) 10.1016/j.nuclphysbps.2009.09.013, arXiv:0907.0409 [astro-ph.CO].

[174] R. Ulrich, T. Pierog and C. Baus, *Cosmic Ray Monte Carlo Package (CRMC)*, version 1.8.0, 2021, 10.5281/zenodo.4558706.

[175] S. Müller, R. Engel, T. Pierog and M. Roth, "Impact of muon detection thresholds on the separability of primary cosmic rays", Astropart. Phys. **97**, 174–185 (2018) 10.1016/j.astropartphys.2017.11.005.

[176] U. Dersch et al. (SELEX), "Total cross-section measurements with π^-, Σ^- and protons on nuclei and nucleons around 600 GeV/c", Nucl. Phys. B **579**, 277–312 (2000) 10.1016/S0550-3213(00)00204-2, arXiv:hep-ex/9910052.

[177] V. A. Petrov, R. A. Ryutin and A. E. Sobol, "LHC as $\pi\, p$ and $\pi\,\pi$ collider", Eur. Phys. J. C **65**, 637–647 (2010) 10.1140/epjc/s10052-009-1202-0, arXiv:0906.5309 [hep-ph].

[178] R. A. Ryutin, "Total pion–proton cross section from the new LHCf data on leading neutrons spectra", Eur. Phys. J. C **77**, 114 (2017) 10.1140/epjc/s10052-017-4690-3, arXiv:1612.03418 [hep-ph].

[179] V. A. Khoze, A. D. Martin and M. G. Ryskin, "Total $\pi^+ p$ cross section extracted from the leading neutron spectra at the LHC", Phys. Rev. D **96**, 034018 (2017) 10.1103/PhysRevD.96.034018, arXiv:1705.03685 [hep-ph].

[180] S. P. Denisov, S. V. Donskov, Y. P. Gorin, R. N. Krasnokutsky, A. I. Petrukhin, Y. D. Prokoshkin and D. A. Stoyanova, "Absorption cross-sections for pions, kaons, protons and anti-protons on complex nuclei in the 6 to 60 GeV/c momentum range", Nucl. Phys. B **61**, 62–76 (1973) 10.1016/0550-3213(73)90351-9.

[181] A. Aduszkiewicz et al. (NA61/SHINE), "Measurements of total production cross sections for π^++C, π^++Al, K^++C, and K^++Al at 60 GeV/c and π^++C and π^++Al at 31 GeV/c", Phys. Rev. D **98**, 052001 (2018) 10.1103/PhysRevD.98.052001, arXiv:1805.04546 [hep-ex].

[182] A. S. Carroll et al., "Absorption Cross-Sections of $\pi^\pm$, $K^\pm$, p and $\bar{p}$ on Nuclei Between 60 GeV/c and 280 GeV/c", Phys. Lett. B **80**, 319–322 (1979) 10.1016/0370-2693(79)90226-0.

[183] V. V. Avakian et al., "Determination of the cross section of the pion and nucleon interaction with iron nuclei in the energy range 0.5–5.0 TeV", Nucl. Phys. B **259**, 163–169 (1985) 10.1016/0550-3213(85)90304-9.

[184] N. M. Agababyan et al. (EHS/NA22), "Inclusive production of vector mesons in $\pi^+ p$ interactions at 250 GeV/c", Z. Phys. C **46**, 387–395 (1990) 10.1007/BF01621026.

[185] M. R. Atayan et al. (EHS/NA22), "π^0 and η meson production in $\pi^+ p$ and $K^+ p$ collisions at 250 GeV/c", Z. Phys. C **54**, 247–254 (1992) 10.1007/BF01566653.

[186] S. Ostapchenko, "QGSJET-II: physics, recent improvements, and results for air showers", EPJ Web Conf. **52**, 02001 (2013) 10.1051/epjconf/20135202001.

[187] H.-J. Drescher, "Remnant Break-up and Muon Production in Cosmic Ray Air Showers", Phys. Rev. D **77**, 056003 (2008) 10.1103/PhysRevD.77.056003, arXiv:0712.1517 [hep-ph].

[188] A. Aduszkiewicz et al. (NA61/SHINE), "Measurement of meson resonance production in π^-+ C interactions at SPS energies", Eur. Phys. J. C **77**, 626 (2017) 10.1140/epjc/s10052-017-5184-z, arXiv:1705.08206 [nucl-ex].

[189] F. Riehn, "Hadronic multiparticle production with Sibyll", PhD thesis (KIT, 2015), 10.5445/IR/1000052699.

[190] F. Riehn, R. Engel, A. Fedynitch, T. K. Gaisser and T. Stanev, "The hadronic interaction model Sibyll 2.3c and muon production in extensive air-showers", EPJ Web Conf. **208**, 11002 (2019) 10.1051/epjconf/201920811002.

[191] O. Fischer, M. Reininghaus and R. Ulrich, "Avenues to new-physics searches in cosmic ray air showers", PoS **ICHEP2020**, 602 (2021) 10.22323/1.390.0602, arXiv:2012.14293 [hep-ph].

[192] H. Goldberg, "Breakdown of perturbation theory at tree level in theories with scalars", Phys. Lett. B **246**, 445–450 (1990) 10.1016/0370-2693(90)90628-J.

[193] J. M. Cornwall, "On the High-energy Behavior of Weakly Coupled Gauge Theories", Phys. Lett. B **243**, 271–278 (1990) 10.1016/0370-2693(90)90850-6.

[194] V. V. Khoze and M. Spannowsky, "Higgsplosion: Solving the Hierarchy Problem via rapid decays of heavy states into multiple Higgs bosons", Nucl. Phys. B **926**, 95–111 (2018) 10.1016/j.nuclphysb.2017.11.002, arXiv:1704.03447 [hep-ph].

[195] C. Degrande, V. V. Khoze and O. Mattelaer, "Multi-Higgs production in gluon fusion at 100 TeV", Phys. Rev. D **94**, 085031 (2016) 10.1103/PhysRevD.94.085031, arXiv:1605.06372 [hep-ph].

[196] G. Brooijmans, P. Schichtel and M. Spannowsky, "Cosmic ray air showers from sphalerons", Phys. Lett. B **761**, 213–218 (2016) 10.1016/j.physletb.2016.08.030, arXiv:1602.00647 [hep-ph].

Acknowledgements

My endeavour to work on, write and finally finish this thesis could not have been successful without the support of numerous colleagues and friends, many of whom have become both.

I am indebted to Ralph Engel for giving me the opportunity to join the cosmic ray business at IAP, entrusting me with the CORSIKA 8 project, making it possible to join the double doctoral degree program and acting as referee of this thesis.

My gratitude goes to Diego Melo for hosting me during my (unfortunately only) stay at ITeDA and for being the second referee. Without your help I would probably not have looked into AIRES.

I am particularly grateful for Ralf Ulrich committing to be my day-to-day advisor. You always have an open ear and solve any upcoming problem whatsoever quickly. Your engagement in the project has been and is a key to its success. Many thanks also for proofreading this thesis and suggesting numerous improvements.

I thank Tanguy Pierog for teaching me many things about hadronic interactions and providing valuable guidance. Without your insight into CORSIKA and CONEX, I would have been totally lost, and I'm happy that I could share the office with you (even if it was only virtual for quite some time...).

I also thank Felix Riehn, Antonio Augusto Alves Jr. and André Schmidt for their various contributions to the CORSIKA 8 projects that also had a significant impact on this thesis.

I gratefully acknowledge the collaboration with Oliver Fischer. Thank you for all the efforts on the project on large multiplicity events and countless suggestions to use CORSIKA 8 for new and exotic physics. I hope we can follow up on them some time soon.

Kudos to (in order of appearance) Nicolás González, Alexander Streich, David Schmidt, Ariel Bridgeman, Ana Botti, Daniela Mockler, William Painter, Álvaro Taboada, Matías Perlín, Isabel Goos, Ana Laura Müller, Max "Klausi" Stadelmaier, Steffen Hahn, Marina Scornavacche, Felix Schlüter and Nikolaos Karastathis for all the good times at the institutes and all sorts of activities in various parts of the world (I count at least eight countries on three continents...).

Special thanks go to Johannes Hulsman for being an awesome office mate in Bodenhaltung as well as in the batcave, and to Flavia Gesualdi for always being cheerful, kind and helping me with so many things. I always enjoyed the daily lunch breaks and meriendas with you.

Many thanks to Sabine Bucher, Marie-Christine Mundt and Doris Wochele for all kinds of administrative support.

I would also like express my gratitude towards my buddy Alexander Himrich for all the good times we have had for more than half of my life.

My deepest appreciation goes to Angelika, Hannah and Sigurd Reininghaus as well as Kathrin Bismark for their continued support and encouragement to find my way and for standing by my side all the time.

Computational resources of the BinAC computing cluster, supported by the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen, the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant no. INST 37/935-1 FUGG, were of great avail for this work. Financial support was provided by Landesgraduiertenförderung Baden-Württemberg (LGF), the Karlsruhe School of Elementary Particle and Astroparticle Physics: Science and Technology (KSETA), and the Helmholtz International Research School for Astroparticle Physics and Enabling Technologies (HIRSAP).