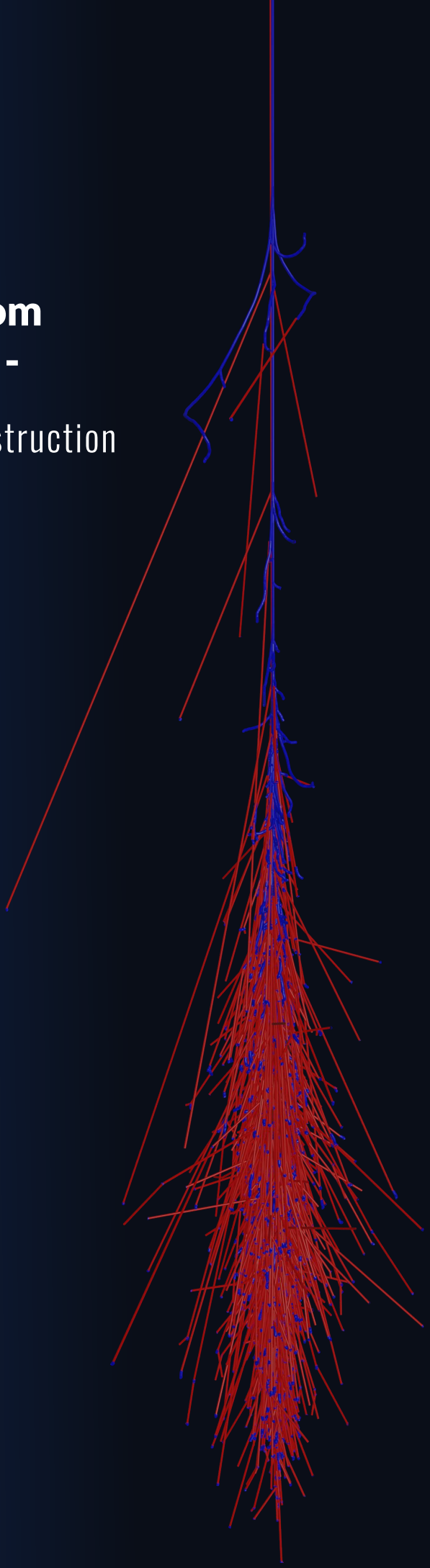
Simulating radio emission from air showers with CORSIKA 8 -

Relevance for energy and mass reconstruction



Nikolaos Karastathis

Karlsruher Institut für Technologie

Simulating radio emission from air showers with CORSIKA 8 - Relevance for energy and mass reconstruction

Zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.)

von der KIT-Fakultät für Physik des Karlsruher Instituts für Technologie (KIT) genehmigte DISSERTATION

von Nikolaos Karastathis (M.Sc.)
aus Patra, Griechenland

Tag der mündlichen Prüfung: **03.05.2024**
Referent: **Prof. Dr. Ralph Engel**
(Karlsruhe Institut für Technologie)
Korreferent: **Prof. Dr. Tim Huege**
(Vrije Universiteit Brussel)

Abstract

The radio detection technique has undergone an impressive renaissance over the last two decades establishing itself as a reliable method to reconstruct basic properties of Cosmic Rays and of the Extensive Air Showers they initiate when impinging Earth's atmosphere. To achieve this, simulation tools are of crucial importance. While existing simulation codes that are used for many years are in place, the benefit of developing new simulation software that takes advantage of modern technologies and is flexible enough to accommodate the growing needs of modern experiments becomes evident. The predictions of simulations need to be very accurate and not be burdened with uncertainties. The radio emission from Extensive Air Showers is a process well understood which is derived from first principles. Two of the most common formalisms that describe it, namely the "Endpoints" formalism and ZHS are employed by the well established air shower simulation codes, CORSIKA 7 and ZHAireS respectively. Both of them are calculated under different assumptions and the level of their agreement acts as a good indication on how well the radio emission from Extensive Air Showers is understood. Trustworthy radio simulations and a very good understanding of the radio emission calculations are the key ingredients of successful energy and mass reconstruction of cosmic rays.

The work of this thesis can be divided into two parts. The first part is focused on the development of a modern and flexible radio module as an integral part of the next-generation Monte Carlo simulation code, CORSIKA 8. Its main goal is to provide a fully functional calculation of the radio emission from air showers, utilizing both the "Endpoints" and ZHS formalisms. At the same time, it incorporates a novel design that allows the inclusion of complex physical scenarios that one cannot simulate with the existing software. Once I have presented the code design, the features and the capabilities of the radio module, it is going to be validated with existing software. The "Endpoints" and ZHS formalisms are going to be directly compared for air showers and their level of agreement is going to be quantified. In addition, I will inspect the impact of simulation details on energy and mass reconstruction. Once I have established that the radio emission calculation is indeed well understood and the simulations are trustworthy I will proceed to the second part of this thesis.

A key quantity for Extensive Air showers is the shower maximum, or $X_{\max}$, which reveals information regarding the mass composition of the primary particle. Well established $X_{\max}$ reconstruction techniques utilize the information carried in the signal pulse in the form of its fluence, discarding information with respect to the pulse shape. The primary objective of this part of the thesis is to develop a reconstruction scheme that does not rely on fluence, instead relies solely on the information contained in the pulse shape. I will present and benchmark such a reconstruction scheme in terms of its accuracy and point out potential opportunities.

Zusammenfassung

Die Radiodetektionstechnik hat in den letzten zwei Jahrzehnten eine beeindruckende Renaissance erlebt und sich als zuverlässige Methode zur Rekonstruktion grundlegender Eigenschaften der kosmischen Strahlung und der ausgedehnten Luftschauer, die sie beim Auftreffen auf die Erdatmosphäre auslösen, etabliert. Um dies zu erreichen, sind Simulationsmethoden von entscheidender Bedeutung. Während bestehende Simulationssoftwares, die seit vielen Jahren verwendet werden, vorhanden sind, wird der Nutzen der Entwicklung einer neuen Simulationssoftware, die moderne Technologien nutzt und flexibel genug ist, um den wachsenden Anforderungen moderner Experimente gerecht zu werden, offensichtlich. Die Vorhersagen von Simulationen müssen sehr genau sein und dürfen keine Unsicherheiten erzeugen. Die Radioemission ausgedehnter Luftschauer ist ein gut verstandener Prozess, der auf Grundprinzipien beruht. Zwei der gebräuchlichsten Formalismen, die es beschreiben, nämlich der „Endpoint“-Formalismus und ZHS, werden von den etablierten Luftschauer-Simulationssoftwares CORSIKA 7 bzw. ZHAireS verwendet. Beide werden unter unterschiedlichen Annahmen berechnet und der Grad ihrer Übereinstimmung ist ein guter Hinweis darauf, wie gut die Radioemission von ausgedehnten Luftschauern verstanden ist. Zuverlässige Radiosimulationen und ein sehr gutes Verständnis der Radioemissionsberechnungen sind die Schlüsselfaktoren für eine erfolgreiche Energie- und Massenrekonstruktion der kosmischen Strahlung.

Die Arbeit dieser Thesis lässt sich in zwei Teile gliedern. Der erste Teil konzentriert sich auf die Entwicklung eines modernen und flexiblen Radiomoduls als integraler Bestandteil der Monte-Carlo- Simulationssoftware der nächsten Generation, CORSIKA 8. Ihr Hauptziel ist die Bereitstellung einer voll funktionsfähigen Berechnung der Radioemission von Luftschauern, unter Verwendung sowohl des „Endpoint“- als auch des ZHS-Formalismus. Gleichzeitig verfügt es über ein neuartiges Design, das die Einbeziehung komplexer physikalischer Szenarien ermöglicht, die mit der vorhandenen Software nicht simuliert werden können. Sobald ich das Code-Design, die Features und die Fähigkeiten des Radiomoduls vorgestellt habe, wird es mit vorhandener Software validiert. Die „Endpoint“- und ZHS-Formalismen werden für Luftschauer direkt verglichen und der Grad ihrer Übereinstimmung quantifiziert. Darüber hinaus werde ich den Einfluss von Simulationsdetails auf die Energie- und Massenrekonstruktion untersuchen. Sobald ich festgestellt habe, dass die Berechnung der Radioemission tatsächlich gut verstanden ist und die Simulationen vertrauenswürdig sind, werde ich mit dem zweiten Teil dieser Arbeit fortfahren.

Eine Schlüsselgröße für ausgedehnte Luftschauer ist das Schauermaximum oder $X_{\max}$, das Informationen über die Massenzusammensetzung des Primärteilchens liefert. Gut etablierte $X_{\max}$-Rekonstruktionstechniken nutzen die im Signalimpuls enthaltenen Informationen in Form seiner Fluenz und verwerfen Informationen in Bezug auf die Impulsform. Das Hauptziel dieses Teils der Arbeit besteht darin, ein Rekonstruktionsschema zu entwickeln, das nicht auf der Fluenz basiert, sondern ausschließlich auf den in der Pulsform enthaltenen Informationen. Ich werde ein solches Rekonstruktionskonzept hinsichtlich seiner Genauigkeit vorstellen und bewerten und mögliche Chancen aufzeigen.

Contents

Part I

Introduction: Cosmic Rays, Extensive Air Showers and Radio Emission from Extensive Air Showers

1 Introduction

Cosmic messengers from all over the universe arrive in Earth carrying valuable information in the form of high energetic particles. These particles traverse the vast cosmos and a small group of those can accelerate to reach tremendous amounts of energies before reaching Earth. These are called ultra-high-energy cosmic rays (UHECRs) and studying them provides the opportunity to investigate the most violent regions of the universe. At the same time, UHECRs exceed in energy even the most powerful particle accelerators and hence, allow for the study of particle physics at such extreme energy regime.

Cosmic Rays of the highest energies are very rare though with less than 1 particle per square kilometer and century reaching the Earth's atmosphere. The way to study them is indirect, through the cascade of secondary particles that they produce when interacting with the molecules of the atmosphere. This phenomenon is called an Extensive Air Shower (EAS). Large detectors all over the world are built for the purpose to study and measure EAS with different techniques. A very promising technique to measure EAS and reconstruct the properties of the primary cosmic ray particle is the radio detection technique. The radio waves emitted from the electromagnetic interactions taking place in EAS are measured with antenna arrays. To use these measurements and reconstruct key properties of the primary particle such as its mass composition and its energy, detailed microscopic simulations are essential. Formalisms have been developed to accurate describe the radio emission mechanisms and have been implemented in the form of algorithms in Monte Carlo simulation codes.

It has become evident though, that simulation codes such as CORSIKA 7 that has been the golden standard for the community for many decades, has far surpassed its initial use case, and has become insufficient to keep up with the growing experimental needs. Modern and flexible software that can be easily adjusted to more sophisticated scenarios, like showers crossing from air to dense media such as ice for example, aim to become a worthy successor and future-proof the simulation landscape. Such a candidate software is CORSIKA 8, a new C++ framework whose purpose is to address these challenges. One of the main goals of this work is to incorporate a radio module to CORSIKA 8 as an integral part. The module has been designed in a flexible way as to take full advantage of the capabilities of CORSIKA 8 and be easily upgradeable for potential use cases in the future. A direct comparison of formalisms that have been derived from

first principles in order to calculate the radio emission from EAS is made possible with the radio module as implemented in CORSIKA 8 and can solidify the idea that the radio emission is a well understood process. In this work, technical parameters of the simulation codes that affect the predictions of the radio emission formalisms are studied with respect to their relevance on energy and mass reconstruction of the primary cosmic ray particle. Finally, reconstruction techniques that utilize the information of the radio pulses are also developed in this thesis.

This thesis is organized in three main parts. Part I is a brief overview of cosmic rays and extensive air showers. Special focus is given to an overview of the radio emission mechanism from EAS, its physics, existing modelling approaches, detection capabilities and reconstruction methods. Part II marks the beginning of the work produced as part of this thesis. In section 5, the architecture and the philosophy of the radio module is being presented. In section 6, the radio module is validated for physical scenarios and cases where it acted as a diagnostics tool for the core code of CORSIKA 8 and the work done to improve CORSIKA 8 is discussed. In section 7, the parallelization procedure utilizing multithreading in an effort to speed up the radio calculations is shown. In section 8, the radio module in CORSIKA 8 is compared with existing well established simulation codes, namely CORSIKA 7 and ZHAireS. In section 9, a systematic comparison between the Endpoints and ZHS formalisms as implemented in CORSIKA 8 takes place for air showers. In section 10, the effect of technical parameters of simulations on the radio emission calculation and their connection to energy and mass reconstruction is investigated. Finally in part III, a novel X_{max} reconstruction procedure in the context of LOFAR and SKA is being presented.

2 Cosmic Rays & Extensive Air Showers

Cosmic rays are ionized nuclei, predominantly protons but also alpha particles and heavier nuclei that traverse the cosmos with velocities close to the speed of light. They arrive at the earth at a wide range of energies with an upper limit even exceeding 10^{20} eV [1]. The higher the energy the more rare cosmic rays are. Their energy also dictates how they can be detected. Lower energetic ones of 10 TeV or less are measured directly either outside or at the very edge of the atmosphere, while higher energetic ones of 1 PeV or more are measured indirectly through their interactions with the atmosphere. The atmosphere acts as a calorimeter, where the incoming cosmic ray while interacting with it produces a cascade of secondary particles, or an extensive air shower. Studying the development and the properties of extensive air showers is an indirect way to examine the cosmic ray by reconstructing its properties. The most important questions that cosmic rays pose to this day are "Where do they come from?" and "How do they accelerate to such high energies?". The answers to these questions are not yet fully known but progress towards answering them is being made for almost a century now. In this section, the basic properties of cosmic rays and extensive air showers are going to be briefly discussed. Detection techniques are going to be showcased and the radio emission from extensive air showers which is the main topic of this thesis is going to be summarized.

2.1 Cosmic Rays

To gain a better insight in cosmic rays it is instructive to look at their flux as a function of energy. The energy spectrum with data gathered from many experiments over the years is shown in fig. 1, and it extends from energies of 10^{13} eV all the way up to more than 10^{20} eV. The spectrum follows a power law with a varying spectral index (-2.5 to -3.3) which reveals the underlying structures in the measurements. To better visualize these structures the flux is multiplied by a factor of $E^{2.6}$. There are 4 clear features in fig. 1 where the spectral index changes at specific energies. They are annotated as "Knee" at an energy of $\approx 5 \times 10^{15}$ eV, "2nd Knee" at $\approx 10^{17}$ eV, "Ankle" at $\approx 5 \times 10^{18}$ eV and a "Suppression" region at energies above $\approx 5 \times 10^{19}$ eV where there seems to be no more incoming flux of ultra-high-energy cosmic rays.

The structure observed for energies below $\sim 10^{18}$ eV is likely associated with the transition between galactic and extra-galactic sources with the "Ankle" serving as the limit. According to the Hillas criterion ([2], [3]), cosmic rays are restrained by the galactic magnetic field and while their Larmor radius (r_L) is smaller than their enclosing region, they cannot escape. This is not the case for higher energies where cosmic rays can easily escape their enclosing region. This interpretation relies heavily on our current understanding of the galactic acceleration, propagation mechanisms and nearby galactic sources. However, the origin of the most energetic galactic cosmic rays close to the "Ankle" is not fully understood. Anisotropy studies which have been performed by IceCube [4] in the TeV to PeV range could potentially improve our understanding regarding the magnetic fields in our galaxy and hence the propagation mechanisms of galactic cosmic rays.

For ultra-high-energy cosmic rays ($E \gtrsim 10^{18}$ eV) the latest measurements have been published by Auger and TA [5]. The energy spectrum along with spectral features are shown in fig. 2 and in this figure the newly discovered "Instep" feature is visible between the "Ankle" and the "Suppression". It is not clear what causes the flux suppression at the highest of energies. A leading hypothesis for this behaviour is the Greisen, Zatsepin, Kuzmin cut-off (or simply GZK cut-off) published in 1966 [6], [7]. According to them, an ultra-high-energy proton interacts with photons from the cosmic microwave background (CMB) at the critical energy of $E = 5 \times 10^{19}$ eV and via the production of Δ^+ resonance decays losing roughly 20% of its energy into a nucleon and a pion via:

$$p + \gamma_{CMB} \rightarrow \Delta^+ \rightarrow p + \pi^0 \tag{1}$$

$$p + \gamma_{CMB} \rightarrow \Delta^+ \rightarrow n + \pi^+ \tag{2}$$

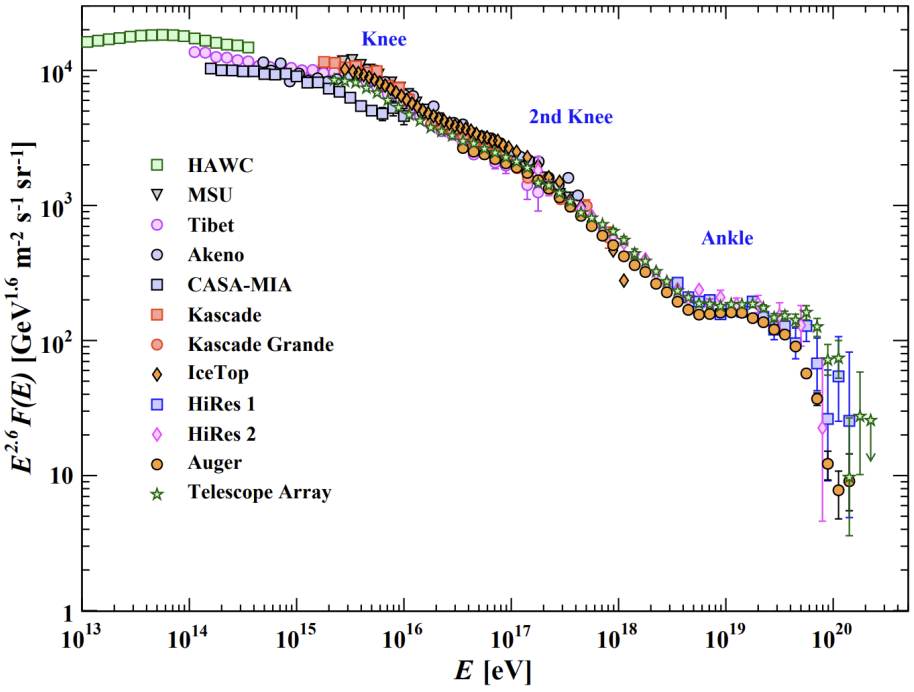


Figure 1: The all particle spectrum as a function of energy. The features of the spectrum as indicated by the knee and the second knee, along with the suppression at very high energies are shown. Figure taken from [8].

The energy losses occur until the proton falls bellow $E = 5 \times 10^{19}$ eV which limits the travelling distance of such a proton to roughly 50 Mpc to 100 Mpc around the earth. This is referred as the GZK horizon within which cosmic rays with energies up to $E = 5 \times 10^{19}$ eV are

able to be detected. Considering such photo-nuclear interactions for heavier nuclei leads to similar cut-offs which could potentially explain the suppression witnessed in fig. 1 and fig. 2. Another interesting explanation for the suppression would be the Peters cycle, where the maximum energy that can be reached by a nucleus is associated to the charge via its rigidity $R = \frac{pc}{Ze}$. Protons would experience then such a cut-off followed by heavier nuclei. A recent study seems to disregard this explanation though in favor of beyond the Standard Model processes [9]. In fig. 2, both TA and Auger data seem to agree within 9.5% for energies up to 10^{19} eV, while for larger energies the agreement drops to 20% per decade [5]. This is not fully understood yet.

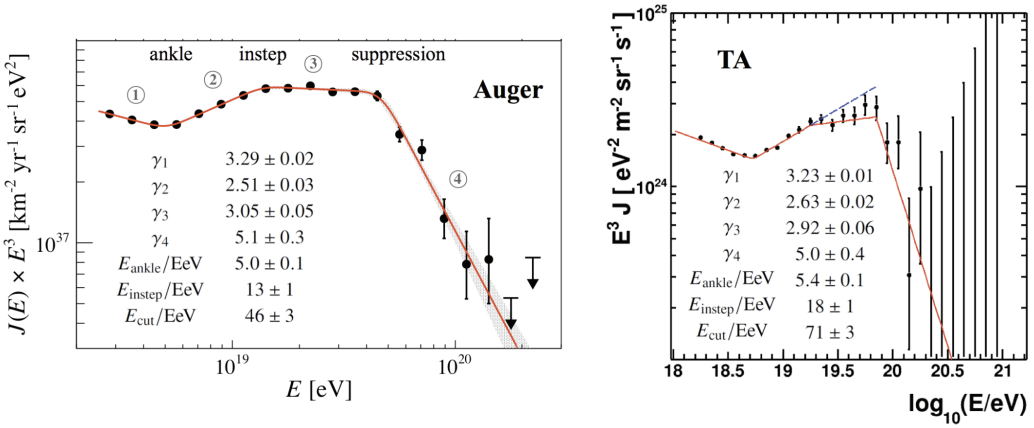


Figure 2: The energy spectrum zoomed into very high energies. The "instep" feature of the spectrum is shown for both Auger and TA. Figure taken from [5].

Another quite obvious remark from that figure is the fact that with increasing energies cosmic rays do become very rare. As an example, for an energy of the order of 10^{15} eV, roughly 1 particle per year per square meter is detected, while for an energy of the order of 10^{20} eV the integral flux drops tremendously to roughly 1 particle per century per square kilometer. To tackle this issue sparse arrays of detectors spanning thousands of square kilometers have been employed over the years for many experiments with prime examples in size being the Pierre Auger Observatory in Argentina and the Telescope Array in the United states. Inspecting and understanding in depth the cosmic ray energy spectrum especially in the key regions that were mentioned, it is evident that more detailed mass composition studies need to take place. Such studies will be possible with the upcoming Auger upgrade to Auger Prime [10].

For very high energy cosmic rays ($E \gtrsim 5 \times 10^{15}$ eV) indirect measurements are feasible through extensive air showers. The key information to extract is the energy, arrival direction and of course mass composition, which proves to be the most challenging parameter due to the highly statistical nature of hadronic interactions. The main focus of this thesis is extensive air showers and more specifically the electromagnetic radiation emitted from them, which are going to be discussed next.

2.2 Extensive Air Showers

When an ultra-high-energy cosmic ray (UHECR) impinges on the earth's atmosphere it interacts with its air molecules. This interaction creates a domino effect of secondary interactions and generates what is known as an Extensive Air Shower (EAS), or shorter, an air shower. The vast amount of energies of the incoming cosmic rays apart from studying its astrophysical aspects, give the opportunity to perform particle physics studies on an energy scale far beyond the scale of current man made collider experiments. The shower development is a highly complex process from which information for the primary cosmic ray particle such as mass composition, energy and arrival direction can be drawn. To do so, numerous Monte Carlo simulations are employed. At the start of the shower development, hadronic interactions are dominant producing mainly pions and kaons which comprise of what is known as the hadronic component of the shower [1]. These product particles interact further with the air molecules and give rise to the other two components of the shower via strong, weak and electromagnetic interactions, namely the muonic and the electromagnetic component. A rough picture of the shower components is drawn in fig. 3 (left). Decaying hadrons feed the muonic component with muons and muon-neutrinos. Neutral hadrons via the photon production produce electromagnetic particles ($e^{\pm}$ pairs) which in turn radiate more photons through bremsstrahlung. While most hadrons decay before they reach the ground, muons do reach the ground with a small chance of further interactions and their detection can provide crucial information about the mass composition of the primary particle. Additionally, the average electromagnetic particle energy increases until it reaches the critical energy of $E_c \approx 87$ MeV. After that energy losses due to ionization become substantial and the shower dies out. As a result, from $e^{\pm}$ pairs exciting atmospheric nitrogen molecules fluorescence light is emitted which can reveal the shower development with great accuracy. The $e^{\pm}$ pairs also radiate energy in the frequency band of a few MHz to a few GHz which are detected by ground based, sparse arrays of antennas.

This highly complex picture can be captured by the longitudinal profile shown in fig. 3 (right). This figure shows the longitudinal development of the air shower in terms of slant depth. The hypothetical straight line that connects the first interaction point in the direction of the primary particle's momentum to the ground is called the shower axis. The interception point of the shower axis with the ground is the shower core. Slant depth is defined as the amount of matter that a particle has traversed through along the shower axis. It is defined in eq. (3) below and it is measured in $\mathrm{g\,cm^{-2}}$.

$$X = \int_{\text{shower axis}} \rho(s)\mathrm{d}s \tag{3}$$

The point in the longitudinal development where the maximum number of particles is found is called the depth of the shower maximum, or simply $X_{\max}$. Furthermore, $X_{\max}$ reveals crucial information about the mass composition of the primary particle. According to the superposition principle for cascades for the same primary particle energy, $X_{\max}$ for lighter nuclei

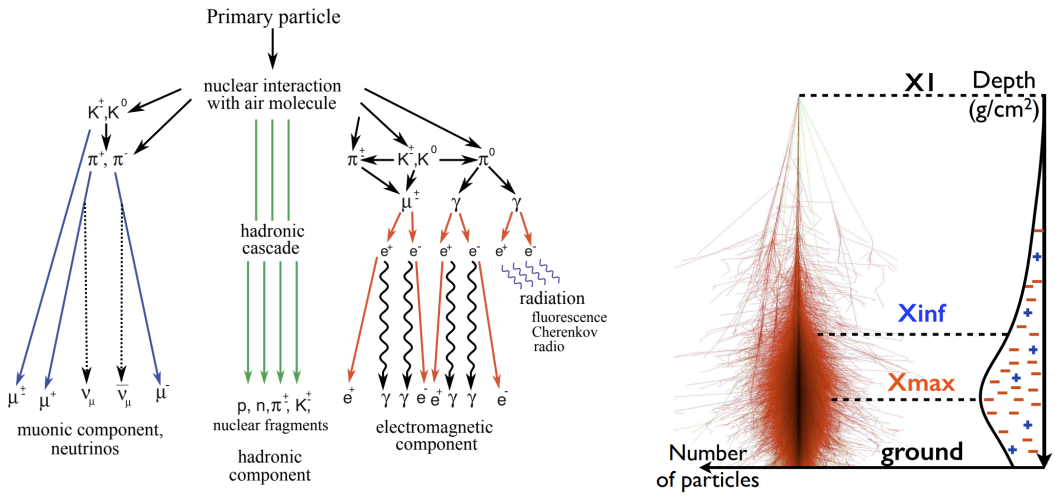


Figure 3: A schematic diagram of the components of an extensive air shower shown in the left. The longitudinal development of the number of particles of a shower is captured by its longitudinal profile. Figure on the right taken from [11] and figure on the left taken from [12].

is deeper than heavier nuclei [1].

2.2.1 The hadronic component

The hadronic component of the showers is the most complex and least understood part of it. Hadrons act as a source for the rest of the shower components through their decay channels and are mostly detected indirectly. What makes this component hard is the nature of hadronic interactions themselves. The theory in the Standard Model responsible for the strong interaction between partons (quarks and gluons) is Quantum Chromodynamics (QCD). A feature of QCD is asymptotic freedom, which means that the running gauge field coupling tends to vanish logarithmically at small distances [13]. In simpler terms, as the energy scale increases the coupling constant becomes smaller and the interactions weaker. When the coupling constant is small Perturbation Theory can be applied to QCD, while at lower energies or larger distances the coupling constant gets large and leads to confinement. In this regime, Perturbative QCD is no longer an option and one needs to resort to other techniques in order to make predictions for parton interactions. In EAS we are interested in hadron interactions. For these, one needs to first resort to parton level modelling in order to take into account parton scattering. From there one gets back to the final state hadrons, which is known as hadronization.

For EAS we rely on specialized simulation codes also known as hadronic interaction models. These are codes focused in calculating the final state particles from 2 initial hadrons and use LCH data from pp collisions to make extrapolations for the highest energies. Many

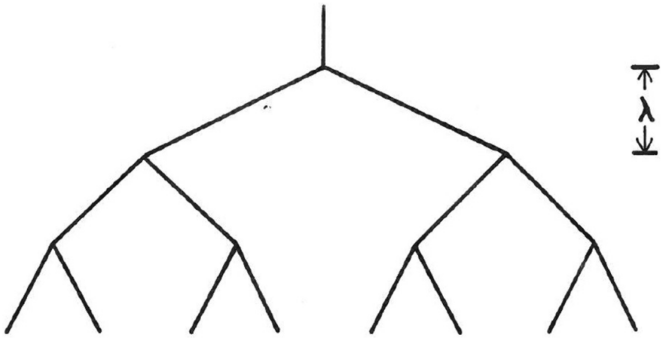


Figure 4: A sketch of the Heitler model [1].

such codes exist, with most common being SIBYLL 2.3d [14] which is designed for air showers, QGSJet-II.04 [15] and EPOS-LHC [16]. These 3 models do not satisfy the demands for lower energies < 50 GeV, but instead there are hadronic interaction models specifically aimed to lower energies. Some of them are FLUKA [17], [18] UrQMD [19] and HSA (Hillas Splitting Algorithm).

2.2.2 The electromagnetic component

As mentioned earlier, the hadronic component feeds the electromagnetic component mainly through short-lived neutral pion decay via the $\pi_0 \rightarrow \gamma + \gamma$ channel and then through pair production $\gamma \rightarrow e^- + e^-$. At energies above ≈ 30 MeV the total Compton scattering and photoelectric effect probability decrease rapidly, making photon absorption through pair production dominant. The $e^{\pm}$ pair suffers energy losses due to ionization from the air molecules and bremsstrahlung. The number of particles is increasing and at the critical energy $E_c = 87$ MeV the losses from bremsstrahlung and ionization are equal [1]. At this point, the number of particles is at the maximum ($X_{\max}$). Beyond the critical energy, ionization losses are completely taking over and the particle number decreases. It is also worth mentioning that at extreme densities and energies (usually beyond 10^{16} - 10^{18} eV) quantum mechanical treatment needs to be takes into account since collisions cannot be considered point-like anymore. Interference effects take place reducing the pair production and bremsstrahlung cross sections. This known as the Landau-Pomeranchuk-Migdal (LPM) effect [20], [21].

A simplistic toy model that describes roughly the above mentioned behaviour was devised by Heitler in 1944 [1] and is shown in fig. 4. The model assumes purely electromagnetic cascades, but can also be extended for hadron induced cascades. This branching process assumes an incoming charged particle with energy $E_0 \gg E_c$. It travels for exactly one collision length (λ) before branching out into 2 particles, where interactions are either due to pair production or bremsstrahlung. The branching process stops when critical energy E_c is reached. At a slant depth X the number of particles is given by:

$$N(X) = 2^{X/\lambda} \tag{4}$$

And the energy per particle is:

$$E(X) = E_0/N(X) \tag{5}$$

The maximum number of particles is reached at $X_{\max}$ and is:

$$N(X_{\max}) = E_0/E_c \qquad \text{where} \qquad X_{\max} = \lambda \frac{\ln(E_0/E_c)}{\ln 2}$$

It is the electromagnetic component which is responsible for emission of fluorescence light, Cherenkov light and radio emission, and that thus it is the most important one in the context of this thesis.

2.2.3 The muonic component

The muonic component of the shower is fed from the long lived ($c\tau$ = 7.8 m) charged pions decaying via the weak interaction into muons and neutrinos. Muons can decay further providing the electromagnetic component with electrons and positrons. The Heitler model can be extended to include hadronic interactions and help us gain more insight on the shower development. This extension is called the Heitler-Matthews [22] model as originally extended by Matthews. In this model, one considers a hadron with energy E interacting and producing n_{tot} particles. A third of those are neutral particles feeding the electromagnetic component and the rest (n_{ch}) are charged which interact further until they fall below their decay energy E_{decay}. Each interaction vertex is what is called a generation in the shower evolution and after n generations the energy is given by:

$$E_{\text{had}} = \left(\frac{2}{3}\right)^n E_0 \qquad \text{and} \qquad E_{\text{em}} = \left[1 - \left(\frac{2}{3}\right)^n\right] E_0$$

After only a few generations almost all of the initial energy is in the electromagnetic component, which dictates the depth of the shower maximum via:

$$X_{\max}^{(\text{had})}(E_0) = \lambda_{\text{int}} + X_{\max}^{(\text{em})}\left(\frac{E_0}{2n_{\text{tot}}}\right) \tag{6}$$

where λ_{int} is the hadronic interaction length accounting for the length of the first interaction in the atmosphere and the factor $\frac{1}{2}$ taking into account the neutral pion decaying into 2 photons. The number of muons (N_μ) can be approximated by the number of charged particles:

$$N_\mu = (\frac{E_0}{E_{\text{decay}}})^\alpha \qquad \text{with} \qquad \alpha = \frac{\ln N_{\text{ch}}}{\ln n_{\text{tot}}} \approx 0.82...0.9.$$

Since the interaction energies are much larger than the binding energy per nucleon of ~ 5 MeV, one is able to consider the superposition model. In this model, a nucleon of mass A can be approximated as A independent nucleons with energy of $E_{\text{h}} = E_0/A$. Using this one can arrive to the following relations:

$$N^A_{\text{em,max}}(E_0) = AN^{\text{h}}_{\text{em,max}}(E_{\text{h}}/E_{\text{c}}) \approx N_{\text{em,max}}(E_0) \tag{7}$$

$$X^A_{\text{max}}(E_0) = X_{\text{max}}(E_0/A) \tag{8}$$

$$N^A_\mu(E_0) = A\left(\frac{E_0/A}{E_{\text{decay}}}\right)^\alpha = A^{1-\alpha}\left(\frac{E_0}{E_{\text{decay}}}\right)^\alpha \tag{9}$$

From these we can realize a few important points. The number of particles in the electromagnetic component is not mass dependent. X_{max} and the number of muons *are* mass dependent. A heavier nucleus produces more muons than a lighter nucleus with the same energy. On the contrary, a lighter nucleus would be able to penetrate deeper in the atmosphere than a heavier nucleus with the same energy.

Hence, these shower observables can provide useful information about the primary cosmic particle. However, the production of muons is dependent on the hadronic interaction models. These are known to extrapolate cross sections to the highest energies due to the energy limitations of collider experiments. The interpretation of experimental data relies heavily on simulations and there has been observed a "deficit" in simulations when it comes to the number of muons detected. This is called the muon puzzle [23] and the hadronic interaction models widely used to this day, still predict a smaller number of muons than the one detected in experiments.

3 Radio emission from Extensive Air Showers and radio detection with LOFAR and SKA

The radio emission from EAS is the main topic of this thesis. Electrons and positrons from the electromagnetic component of a cascade radiate coherently since the emitted wavelength is larger than typical shower length scales. Muons being roughly 200 times more massive than electrons and positrons do not contribute to the radio emission process. The radio signal emitted can be distinguished from the background for energies $\gtrsim 10^{16}$ eV of the primary particle. Despite a period of intense research for detecting the radio waves from EAS in the 60s and 70s, the interest for it really flourished at the start of the 21st century due to the advent of modern technologies especially in the field of digital signal processing [24]. Many experiments like LOPES [25], LOFAR [26], AERA [27] and others have confirmed that EAS can be detected via their radio emission. This technique has been proven to be competitive in terms of $X_{\max}$ and arrival direction resolution to the well established fluorescence and particle detection techniques. To reconstruct the properties of the primary cosmic particle requires detailed knowledge of the radio emission and its relation to properties of the air shower. This is addressed by employing numerous detailed particle-level Monte Carlo simulations, in which the full air shower is simulated and the radio emission is calculated by dedicated algorithms. These "microcopic simulations" are used for analysis. In this section, the nature of the radio emission is going to be discussed along with a special focus on the mainly used modelling algorithms. Furthermore, some analysis techniques and relevant quantities will be showcased.

3.1 Radio emission mechanisms

For EAS in earth's atmosphere there are 2 main mechanisms that contribute to the emission of radio waves, namely the **geomagnetic emission** and the **charge excess** (or **Askaryan**) contribution [24], [28]. The first mechanism takes place due to the geomagnetic field, which deflects electrons and positrons in the shower that are interacting in the atmosphere to opposite directions inducing a time varying transverse current, polarized in the direction of the geomagnetic Lorentz force as they interact with the atoms in the atmosphere. The Lorentz force is defined by $\vec{F} = q\vec{v} \times \vec{B}_{\text{geo}}$, where q is the charge, $\vec{v}$ is the velocity vector of the particle and $\vec{B}_{\text{geo}}$ is the magnetic field vector of the Earth. The direction of the electric field vector due to the geomagnetic emission is constant along the $-(\vec{v} \times \vec{B}_{\text{geo}})$ direction. This contribution is dominant for air where the mean free path of particles is larger than in denser media such as ice. Additionally, the amplitude of the geomagnetic emission is proportional to the magnetic field B_{geo} and the geomagnetic angle α between the magnetic field direction and the shower axis as understood by the expression: $qvB_{\text{geo}} \sin\alpha$. Finally, the amplitude also depends on the number of electrons and positrons and most of the contribution comes from $X_{\max}$ where the radiation is the most coherent.

The second contribution is an excess of negative charge in the shower front. As the shower develops in the atmosphere, electrons are being ionized from atmospheric molecules and

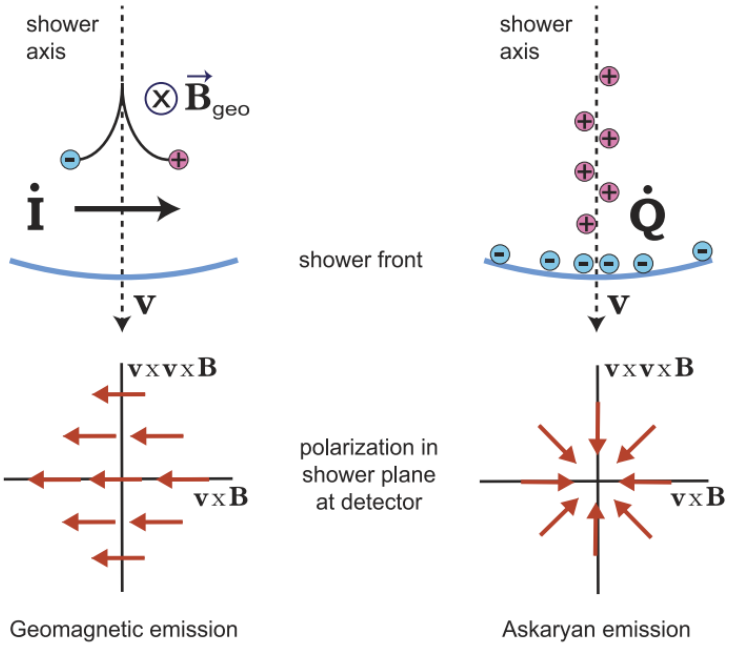


Figure 5: On the left, the geomagnetic emission due to the induced transverse current from Earth's magnetic field is shown. The polarization is in the direction of the geomagnetic Lorentz force. On the right, the radially polarized Askaryan emission due to the time variation of the net charge excess in the shower front is shown. Both of the mechanisms contribute to the total radiation emitted, with the geomagnetic emission being dominant in air. Figure taken from [29].

are accumulated close to the intersection of the shower axis with the shower front. The positive ions stay behind at rest. The resulting net charge varies over time and the electric field is not constant especially after $X_{\max}$ where most particles are absorbed by ionization. The charge excess emission is polarized radially. In air the charge excess contribution is subdominant to the geomagnetic emission accounting for roughly 10% . That is not the picture in dense media though. In this case, the mean free path gets smaller and the Askaryan emission dominates the electric field amplitude. Both mechanisms interfere constructively or destructively and depending on the observer's location this interference can be observed on their emission footprint on the ground. The resulting radiation emitted is forward beamed due to the relativistic motion of the particles, which leads to short radio pulses. In fig. 5, both mechanisms are sketched and their corresponding electric field vectors are drawn in the shower plane $[\vec{v}\times\vec{B}, \vec{v}\times(\vec{v}\times\vec{B})]$. It is worth pointing out that for observers in the $\vec{v}\times(\vec{v}\times\vec{B})$ axis the polarization vectors of each mechanism are orthogonal, which makes it convenient in order to decouple the two contributions and study them separately.

These two are not the only emission mechanisms that contribute to the total coherent sum but they are the most dominant ones in most cases. This picture is challenged by recent work

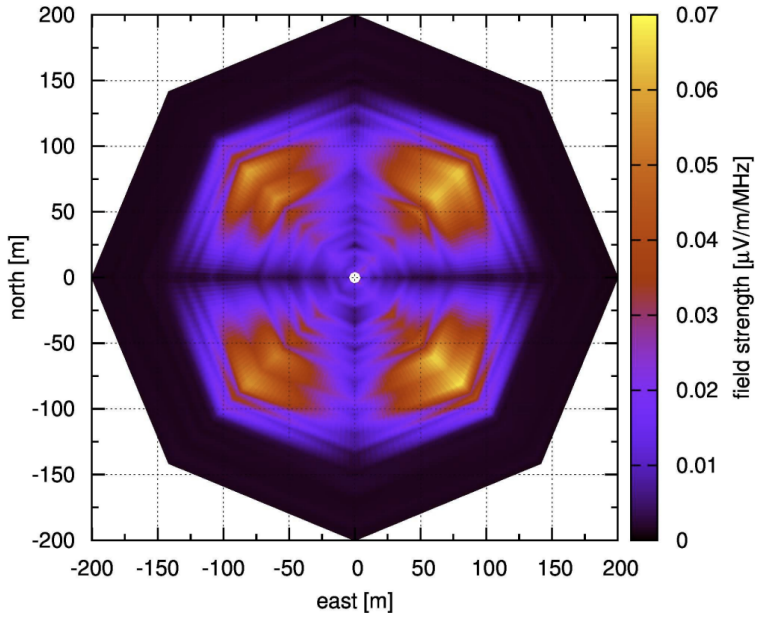


Figure 6: The clover leaf footprint on the ground is showcased for demonstration purposes as simulated with CoREAS in CORSIKA. Figure taken from [33].

[30] for the case of very inclined air showers which show the appearance of a synchrotron-like component to the emission. Initially, geo-synchrotron models that proved to be wrong as they neglected the time dependency of the radio emission would interestingly predict similar footprint patterns [31] on the ground. In more recent work [32] a transition from transverse current to synchrotron emission was predicted, where the latter was not negligible. The main parameters for this transition is observing frequency, the height of emission and the local geomagnetic field. Very inclined showers usually develop higher in the atmosphere where the density is smaller and if the local magnetic field is strong enough (like for example the GRAND site) a non-negligible synchrotron contribution is expected to be observed in the form of the "clover leaf" pattern. Observing such a footprint will be experimental proof that the radiation indeed shifts from being transverse-current-like to synchrotron-like. The "clover leaf" signature pattern is visible in the 3.4 GHz to 4.2 GHz frequency band in fig. 6 for a CORSIKA simulation of an iron induced 10^{17} eV vertical shower.

3.2 Cherenkov-like effects

Electrons and positrons are propagating in the atmosphere, where the refractive index changes with respect to height. As a result, the emitted waves may end up having velocities ($u = c/n$ where n is the refractive index and c the speed of light in vacuum) greater than the speed of light in this medium [28]. At a critical angle, $\theta_c = 1/\arccos \beta n$ assuming $v \approx c$ the radiation forms

what is known as a "Cherenkov cone" and the radio waves arrive at certain observer locations almost simultaneously with the signal being strongly coherent there. The amplitude is amplified there and the emission extends up to the GHz frequency band. In air, where the refractive index is close to 1, the Cherenkov angle (θ_c) is very small ($\simeq 1^o$) which corresponds to distances $\lesssim 150\,\mathrm{m}$ from the shower core for vertical showers and it gets larger for inclined ones. This creates a "Cherenkov ring" on the radio footprint where the energy deposited to the ground is at its maximum. Its diameter naturally depends on the observer's distance. In the case of denser media such as ice where the refractive index is much larger the Cherenkov angle corresponds to $\approx 56^o$. Interestingly, it has been shown [32] that the charge excess contribution is expected to be more Cherenkov-like in the GHz range rather than the MHz range and it may deviate from being Cherenkov-like below 100 MHz for dense media.

3.3 Relevant quantities for radio analyses

For radio experiments the measurements happen on the ground coordinate system. The analyses though usually happen to a more convenient coordinate system, known as the shower plane. Considering $\vec{v}$ as the direction of the shower axis and $\vec{B}$ the geomagnetic field vector one can align one axis to the $\vec{v}\times\vec{B}$ direction. In this coordinate one is able to decouple the contributions of geomagnetic emission mechanism and charge excess depending on the observer position relative to the shower axis [34], [35]. The electric field components can be expressed as a function of the polar angle ϕ, starting at 0° on the positive $\vec{v}\times\vec{B}$ axis and rotating counter clock-wise.

$$\begin{aligned} E_{\vec{v}\times\vec{B}}(\vec{r},t) &= E_{\mathrm{geo}}(\vec{r},t) + cos\phi E_{\mathrm{ce}}(\vec{r},t) \\ E_{\vec{v}\times(\vec{v}\times\vec{B})}(\vec{r},t) &= sin\phi E_{\mathrm{ce}}(\vec{r},t) \end{aligned} \tag{10}$$

where $E_{\mathrm{geo}}(\vec{r},t)$ is the modulus of the electric field due to geomagnetic emission and $E_{\mathrm{ce}}(\vec{r},t)$ due to charge excess. Another relevant quantity used widely in radio analyses is the energy fluence. It describes the coherent energy deposited to the ground in an array of observers, in this case antennas. For every antenna i the fluence (f_i) detected over a period of time $t_{\mathrm{end}} - t_{\mathrm{start}}$ sampled with a binning of Δt can be calculated according to the following formula:

$$f_i(\vec{r}) = \epsilon_0 c \Delta t \sum_{t_{\mathrm{start}}}^{t_{\mathrm{end}}} E_i((\vec{r}),t)^2 \tag{11}$$

with ϵ_0 being the vacuum permittivity, c the speed of light in vacuum and $E(t)$ the electric field. Energy fluence is a key parameter that is used in most analyses for X_{max} reconstruction. One can now switch to the shower plane and express the fluence in terms of $\vec{v}\times\vec{B}$ and $\vec{v}\times(\vec{v}\times\vec{B})$. In this coordinate system they can be written as:

$$
\begin{aligned}
f^i_{\vec{v}\times\vec{B}}(\vec{r}) &= \sqrt{f^i_{\text{geo}}(\vec{r})} + cos\phi\sqrt{f^i_{\text{ce}}(\vec{r})} \\
f^i_{\vec{v}\times(\vec{v}\times\vec{B})}(\vec{r}) &= sin^2\phi f^i_{\text{ce}}(\vec{r})
\end{aligned}
\tag{12}
$$

And finally, one is now able to decouple the contribution of the geomagnetic emission and the charge excess in terms of fluence in the following manner:

$$
\begin{aligned}
f^i_{\text{geo}}(\vec{r}) &= \left(\sqrt{f^i_{\vec{v}\times\vec{B}}}(\vec{r}) - \frac{cos\phi}{|sin\phi|}\cdot\sqrt{f^i_{\vec{v}\times(\vec{v}\times\vec{B})}(\vec{r})}\right)^2 \\
f^i_{\text{ce}}(\vec{r}) &= \left(\frac{1}{sin^2\phi}\cdot f^i_{\vec{v}\times(\vec{v}\times\vec{B})}(\vec{r})\right)
\end{aligned}
\tag{13}
$$

Fluence is dependent on the observer position relative to the shower axis, while the total energy carried in the electromagnetic component of the shower contained in the radio pulses is dependent on the height of observation. This is called radiation energy and is defined as such:

$$
E_{\text{rad}} = \int_0^{2\pi} \mathrm{d}\phi \int_0^{\pi} \mathrm{d}r r f(r,\phi) \tag{14}
$$

The radiation energy depends on the observation height because the parts of the shower up to that height can contribute. To showcase this notion the so called *clipping* effect [34] can be considered. If the observer site is too high above the ground, there is the possibility that the shower has not fully developed and hence not all of its parts have contributed to the radiation energy. In this case, the radiation energy is "clipped". The radiation energy also depends on more factors with the more prominent being the magnitude of the geomagnetic field strength. The same shower can produce a different amount of radiation energy on a different location on earth, where the geomagnetic field is different. A second order dependency comes from the density. An air shower that develops higher in the atmosphere that is less dense would produce slightly larger radiation energy than an identical shower developing lower where the density is larger. This happens mostly because of the charge excess dependency on density. Once all these corrections have been taken into account, it is shown that the corrected radiation energy scales quadratically with the electromagnetic energy in the shower which in turn is correlated with the energy of the primary particle. This makes the fluence and naturally, the radiation energy very crucial quantities that can be used to reconstruct X_{max} and the energy of primary particle. For the successful reconstruction, detailed Monte Carlo simulations are used and as part of this thesis, parameters that can affect the simulation of fluence and radiation energy are going to be discussed in section 10.

3.4 Microscopic modelling of the radio emission

Over the years, many modelling efforts have been made to capture and correctly describe the radio emission from air showers. At the start of these efforts, models were divided between the ones that would predict unipolar pulses and the ones that predicted bipolar ones. It was later realized that the unipolar models were missing a radiation term due to the changing number of radiating particles during the shower development [36]. An important role for this realization was played by the bipolar pulses predicted by another model, MGMR [37], [38]. This is a *macroscopic* model, in the sense that the coherence emission is the outcome of the motion of currents averaged over appropriately chosen distances in the shower. One advantage of such macroscopic models is that they can provide qualitative insights on the radio emission within short amount of computational time. However they lack in accuracy, which renders detailed *microscopic* models that keep track of the emission of each individual particle crucial.

Microscopic models are widely used in Monte Carlo simulation codes like CORSIKA and ZHAireS. Since these Monte Carlo codes generate particle tracks, it is only natural that models that treat individual particle tracks to be suitable. The microscopic models make no assumption for the nature of the radio emission. The calculation starts from first principles, namely Maxwell's equations for linear, isotropic, homogeneous and non dispersive media. The scalar field and vector potential of the electric field rising from a charged particle in the relativistic case are provided by the Liénrad-Wiechert [39] potentials:

$$\begin{aligned} \Phi(\vec{x},t) &= \left[\frac{q}{(1-n\vec{\beta}\cdot\hat{r})R}\right]_{\text{ret}} \\ \vec{A}(\vec{x},t) &= \left[\frac{q\vec{\beta}}{(1-n\vec{\beta}\cdot\hat{r})}\right]_{\text{ret}} \end{aligned} \tag{15}$$

where q is the particle charge, n the effective (average) refractive index along the line of sight from emission point to observer in the medium, $\vec{\beta}$ is defined as $\vec{\beta} = \vec{v}/c$, with c being the speed of light, $\hat{r}$ is the line of sight between particle and observer position defined as $\hat{r} = \vec{R}(t)/|\vec{R}(t)|$ and $R = \vec{R}(t)$ is the distance between particle and observer. This equations needs to be evaluated at the retarded time $t' = t - nR/c$ for the observer at location $\vec{x}$ and time t, hence the "ret" subscript. From eq. (15), one can derive the corresponding electric field:

$$\vec{E}(\vec{x},t) = -\nabla\Phi(\vec{x},t) - \frac{\partial}{\partial t}\vec{A}(\vec{x},t) \tag{16}$$

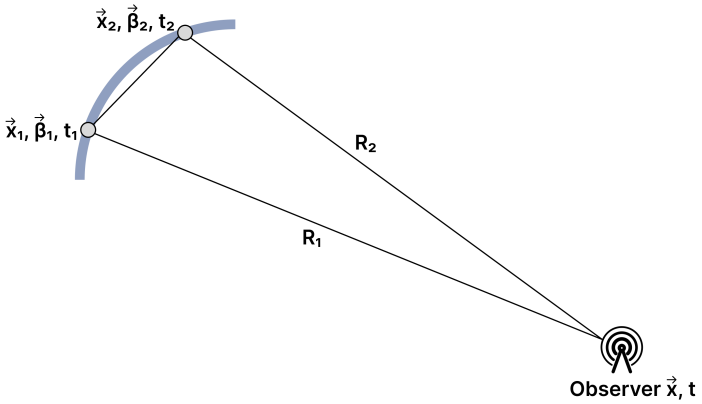


Figure 7: A sketch of how "Endpoints" treats a particle track.

3.4.1 The Endpoints formalism

From the Liénrad-Wiechert potentials, the electric field in a dielectric, non-magnetic medium due to a charge q can be expressed [40]:

$$\vec{E}(\vec{x},t) = q\left[\frac{\hat{r}-n\vec{\beta}}{\gamma^2(1-n\vec{\beta}\cdot\hat{r})^3R^2}\right]_{\text{ret}} + \frac{q}{c}\left[\frac{\hat{r}\times[(\hat{r}-n\vec{\beta})\times\dot{\vec{\beta}}]}{(1-n\vec{\beta}\cdot\hat{r})^3R}\right]_{\text{ret}} \tag{17}$$

where the notation is the same as eq. (15), with the addition of $\dot{\vec{\beta}}$ being the time derivative of $\vec{\beta}$ and $\gamma = 1/\sqrt{(1-\beta^2)}$. The first term decreases as R^{-2} and affects the electric field for close distances. It is called as the near-field term and in practical applications for our purposes is being neglected. The second term decreases as $1/R$ and is referred to as the radiation term.

For the "Endpoints" formalism [40] eq. (17) is being calculated directly. The assumption being made is that radiation happens due to acceleration of a charged particle to a velocity $\vec{\beta} = \vec{\beta}^*$, or due to deceleration from velocity $\vec{\beta} = \vec{\beta}^*$ to rest. The time-averaged electric field over a chosen time window Δt, with "+" indicating the acceleration point where the acceleration vector $\vec{\beta}^*$ is parallel to the velocity vector $\vec{v}$, and "−" indicating the deceleration point where $\vec{\beta}^*$ is anti-parallel to $\vec{v}$ is:

$$\vec{E}_{\pm}(\vec{x},t) = \pm\frac{1}{\Delta t}\frac{q}{c}\left(\frac{\hat{r}\times[\hat{r}\times\vec{\beta}^*]}{(1-n\vec{\beta}^*\cdot\hat{r})R}\right) \tag{18}$$

A schematic sketch of how this works for a particle track is shown in fig. 7. One can assume as the "starting" point of a track, the point indicated with 1, and the "stopping" point the point indicated by 2. The velocity vector is defined as $\vec{v} = \frac{\vec{x}_2-\vec{x}_1}{\vec{t}_2-\vec{t}_1}$. The formalism assumes

that the track is a straight segment. The contributions of both these points can be calculated by eq. (18), as "+" for the "starting" point, and "−" for the "stopping" point. In Monte Carlo simulations a particle trajectory consists of numerous of these sub-tracks. This calculation takes place for each sub-track of each electron or positron particle trajectory.

The pitfall of the formula eq. (18) is close to the Cherenkov angle where it becomes numerically unstable. This is because of the $(1 - n\vec{\beta}^* \cdot \hat{r})$ term in the denominator, which goes to infinity. To overcome this obstacle, an approximation takes place often referred as the "ZHS-like" approximation. In this context, the radiation of the track is considered as a whole by calculating the middle point of the track and feeding this to eq. (18). The "Endpoints" formalism is implemented as a plug-in known as CoREAS in CORSIKA 7.

3.4.2 ZHS

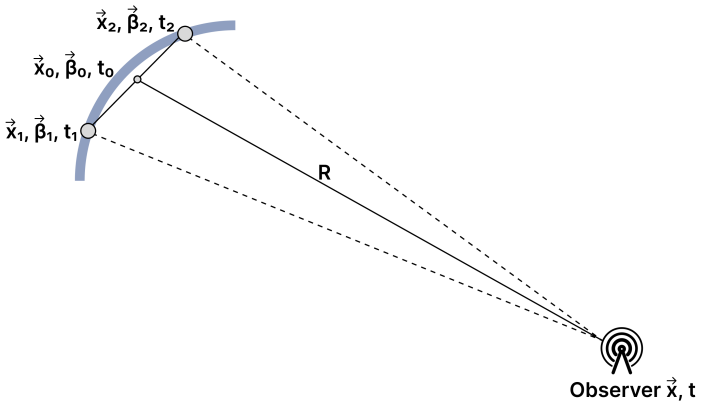


Figure 8: A sketch of how ZHS treats a particle track.

For the ZHS formalism, one starts from Maxwell's equations assuming a track corresponding to a particle motion between two points x_1 and x_2 at times t_1 and t_2 respectively with constant velocity. The track connecting these 2 points is assumed to be straight. The middle point at position x_0 and time t_0 on that particle track is chosen. The assumption here is different than the one presented in the previous section. Radiation is not the result of instantaneous acceleration or deceleration at the ending points of the track, rather than the track itself. The track provides a constant vector potential and its time derivative will produce an electric field behaving similarly to a δ function. The contributions at the start and end of the track will always be exactly antisymmetric. Given a large distance between observer and particle track the vector potential can be shown to be [41] as follows:

$$\vec{A}(\vec{x}, t) = \frac{\mu q}{4\pi R} \frac{\vec{v} - (\vec{v} \cdot \hat{r})\hat{r}}{|1 - n\vec{\beta} \cdot \hat{r}|} \tag{19}$$

$$R\vec{A}(t,\theta_c) = \left[\frac{e\mu_r}{4\pi\epsilon_0 c^2}\right] \delta\left(t - \frac{nR}{c}\right) \vec{v}_\perp \delta t \tag{20}$$

where μ is the permeability, $\frac{\vec{v}-(\vec{v}\cdot\hat{r})\hat{r}}{|1-n\vec{\beta}\cdot\hat{r}|}$ represents the projection of the displacement vector to a perpendicular plane towards the observer's direction ($\vec{v}_\perp$) and $\hat{r}$ is the direction vector to the observer. A schematic sketch for an observer at $(\vec{x},t)$ and how a particle track is being treated by ZHS is shown in fig. 8. Contrary to the "Endpoints" formalism, ZHS does not diverge at the critical Cherenkov angle θ_c, as for this value it can be shown to correspond to eq. (20). The evaluation of the delta function $\delta\big(t - \frac{nR}{c}\big)$ takes place. Its numerical evaluation is dictated by choosing a sufficiently small window, which is well below the inverse high-frequency limit of the system. As a final remark, ZHS works on the Fraunhofer limit. In order for the coherence condition to be met, tracks need to be small when compared to the wavelength and the observer should not be too close to the observer. If these are not true, then the track is sub-divided to smaller tracks.

Figure 9: The dense core of LOFAR, also known as Superterp. Picture taken from [42].

3.5 The Low Frequency Array (LOFAR)

The Low Frequency Array or LOFAR [26] for short, is a large scale radio telescope which spans throughout many European countries with the latest addition being in Bulgaria. It is a multi-purpose sensor network that serves many science projects. For this work, I focus into the Cosmic Ray Key Science Project. Its antennas operate in a wide frequency range of 10 MHz to 240 MHz and are split into 2 groups. The low-band antennas (LBAs) that are sensitive in the 10 MHz to 90 MHz band and the high-band antennas (HBAs) which are sensitive in the 110 MHz to 240 MHz band. 96 LBAs and 96 HBAs form a station in the international locations. LOFAR's dense core, an area of roughly 320 m in diameter, also called "superterp", is located in the Northern part of the Netherlands, and a core station there consists of 96 LBAs and 48 HBAs. The "superterp" is also shown in fig. 9

For cosmic ray detection the LBAs are mainly used in the 30 MHz to 80 MHz band, since the signal emitted from air showers is stronger in the region below 100 MHz. The LBA antennas amount to 2 orthogonal dipoles Λ-shaped each, oriented southwest to northeast and southeast to northwest. Distributed in the same area as the Dutch stations is the LOFAR Radboud Air Shower Array (LORA) which is an independent array of particle detectors. LORA amounts to 20 detectors each containing two scintillators read out by a photomultiplier tube to detect the incoming particles. LBAs, HBAs and the LORA scintillators make up the center of LOFAR which is used for the cosmic ray detection. The layout is shown in fig. 10, where with black and orange crosses the outer and inner LBAs are annotated, respectively. The blue open squares show the HBA tiles and the purple squares indicate the LORA particle detector positions.

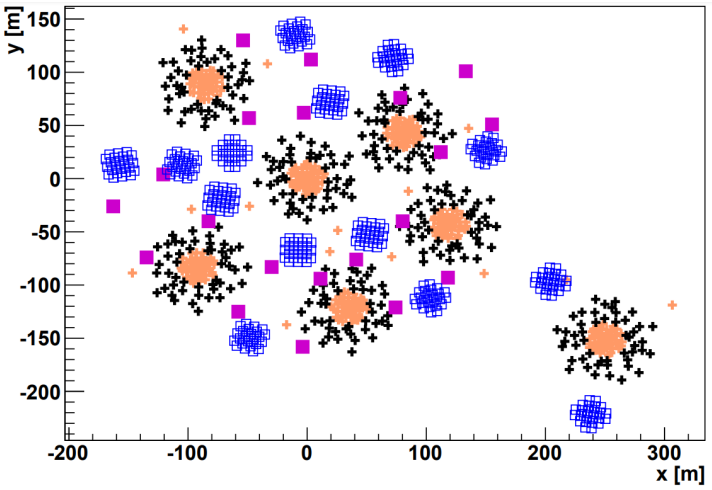


Figure 10: The LOFAR layout center. Figure taken from [26].

LORA's operation is twofold. For an event above 4σ from the noise the detectors get triggered and perform a reconstruction procedure on the go. The arrival direction, energy and core position of the primary particle are reconstructed along with the arrival times. If 13 or more LORA detectors are triggered, then LOFAR is also triggered proceeding to storing 2.1 ms of data which corresponds to 77 MB per station. This trigger threshold of 13 LORA detectors amounts to a minimum primary energy of 2.4×10^{16} eV with an average trigger-rate of 0.8 events per hour. This procedure happens with the help of the Transient Buffer Board (TBB). A TBB is connected to each antenna dipole where raw data of up to 5 seconds is continuously stored. When triggered, TBBs are frozen and the specified data block is saved to disk for further processing.

For the data processing and analysis a custom made object oriented software named PyCRTools was developed. This framework is responsible to select and read out the data blocks as indicated by the relevant timestamps, take into account timing and phase offsets and perform RFI cleaning. Additionally, the data is calibrated using the galactic background emission and finally after unfolding the antenna response by using the LOFAR antenna model and a successful direction fit, the voltage traces are transformed to electric field x,y and z components which are returned from the pipeline along with the peak amplitude and integrated power per antenna. More recent efforts have been made by the LOFAR cosmic rays group to modernize this existing software. Initially, a new framework that would use most of the existing functionality was done under the name Kosmic-ray Radio Analysis TOolS (KRATOS). After a few months into development it was realized that the community in general and the LOFAR group as well would benefit more by the integration of KRATOS within NuRadioReco [43]. This is a radio specific software initially designed for ARIANNA [44], [45], but later came to include more radio experiments. Hence the inclusion of the LOFAR radio experiment there, with a community of physicists/developers to support it seemed only natural.

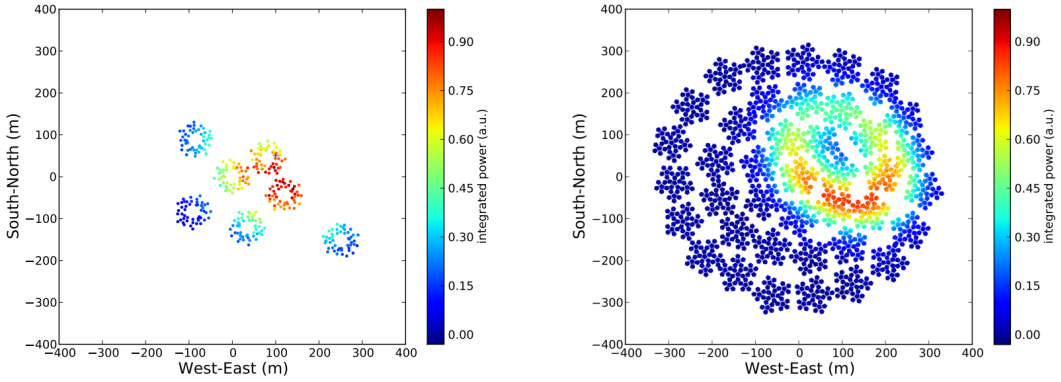


Figure 11: The layout of SKA (right) is compared with the layout of LOFAR (left). The big difference in the number of antennas is evident. This figure was taken from [46].

3.6 The Square Kilometre Array (SKA)

The natural evolution of LOFAR is the low frequency part of the Square Kilometre Array or SKA for short, an extremely dense detector consisting of over 60.000 antennas packed in an area of $1\,\mathrm{km}^2$ [46]. Such an extreme number of antennas is expected to push the radio detection precision for cosmic rays with an energy of around $10^{17}eV$ and allow for an expected $X_{\max}$ resolution of $\leq 10\,\mathrm{g\,cm}^{-2}$ [47]. This will allow more thorough studies of cosmic rays that are expected to be in the transition region between galactic and extra-galactic sources.

There are 3 main advantages for detecting cosmic rays with SKA over LOFAR. The first one is the increased antenna density and homogeneity. The antenna layout of LOFAR is rather inhomogeneous and hence, is able to capture the radio emission footprint characteristics (like the Cherenkov cone for example) on a few favorable events to yield maximum $X_{\max}$ resolution. With SKA on the other hand, the dense array of antennas provides a vary favorable sampling which is able to capture the footprint characteristcs very precisely. This is shown in fig. 11 where a CoREAS based footprint simulation is sampled by LOFAR (left) and by SKA (right). Simulation studies where the LOFAR-style $X_{\max}$ reconstruction scheme (topic of the next section) is being utilized show a $X_{\max}$ resolution of $\sim 6\,\mathrm{g\,cm}^{-2}$ [47].

The second advantage of SKA over LOFAR is the increased frequency coverage from 30 MHz to 80 MHz to 50 MHz to 350 MHz. Simplified simulation studies [46] have shown that in the 30 MHz to 80 MHz band $X_{\max}$ resolution of $\sim 10\,\mathrm{g\,cm}^{-2}$ is reached, while for the 50 MHz to 350 MHz the resolution gets better to $\sim 6\,\mathrm{g\,cm}^{-2}$. Finally, the third advantage of SKA over LOFAR comes from the increased event statistics. Unfortunately, LOFAR uses only the "superterp", an area of $0.1\,\mathrm{km}^2$ for cosmic ray detection. On top of that, the duty cycle for cosmic ray detection is not 100% due to clashes with other astronomical experiments that use the LOFAR detectors and because cosmic ray detection was not initially designed as a feature. Finally, while the 30 MHz to 80 MHz range is the nominal band of LOFAR, in reality the antennas are very

resonant near 60 MHz. These will not be issues for SKA, as the fiducial area is larger amounting to 1 km^2 and the goal is to get a 100% duty cycle.

3.7 $X_{\max}$ reconstruction with LOFAR

The exact energy at which the origin of cosmic rays shifts from galactic to extra-galactic origin is an open question. To tackle this, accurate measurements of the mass composition of cosmic rays are essential. The mass of the primary particle initiating EAS in Earth's atmosphere can be studied indirectly by the shower maximum $X_{\max}$, the point where an air shower reaches the maximum number of particles. The higher the $X_{\max}$, the more the shower has penetrated in the atmosphere. For heavier nuclei, like iron, the nucleus-air cross section is larger and the point of first interaction is higher in the atmosphere resulting in lower values of $X_{\max}$ when compared to lighter nuclei. The effect of different $X_{\max}$ values can also be seen in the radio footprint. Showers that penetrate deeper into the atmosphere have smaller footprints than the ones that penetrate less.

An $X_{\max}$ reconstruction method has been developed for LOFAR that relies heavily on CORSIKA/CoREAS simulations. Once an event is measured with LOFAR, its radio dataset is processed by the analysis pipeline [26]. When all quality tests have been passed, an accurate measurement for the incoming direction, a first estimate for $X_{\max}$ and pulse energy per antenna per triggered station is given [48], [49]. These estimates are then used to produce simulations. Initially, fast CORSIKA 7 simulations using CONEX [50] of 600 showers are produced to find suitable random number seeds in order to have enough showers around the first $X_{\max}$ estimate. It is crucial to have many simulations around the $X_{\max}$ estimate in favor of precision. The total number of simulations is ~ 30 per measured event and the simulations are protons and iron induced showers.

The next step is for every simulated pulse of each shower to pass through a band-pass filter and the LOFAR LBA antenna model [26], in order for the simulations to be comparable to LOFAR measurements. The pulse energy per antenna is matched to LOFAR data, with the core position and overall scaling factor being free parameters. Consequently, to each simulated shower corresponds a χ^2 value defined as such:

$$\chi^2 = \sum_{\text{antennas}} \left(\frac{P_{\text{ant}} - f_r^2 P_{\text{sim}}(x_{\text{ant}-x_0}, y_{\text{ant}} - y_0)}{\sigma_{\text{ant}}} \right)^2 \tag{21}$$

where P_{ant} is the measured signal energy and σ_{ant} its uncertainty. P_{sim} is the simulated signal energy, f_r^2 is the overall scaling factor and (x_0, y_0) is the position of the fitted shower core. It is worth noting that such a reconstruction scheme is self sufficient, in the sense that it relies solely on information from the radio signals [51]. All the χ^2 values for an event can be plotted with respect to their corresponding $X_{\max}$. A parabola fitting procedure can then take place and

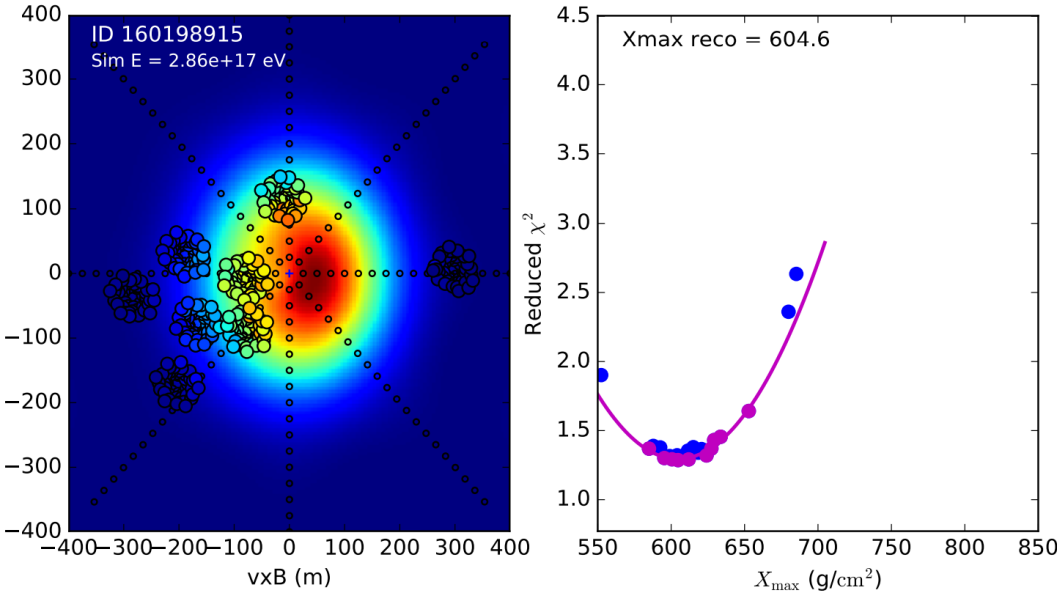


Figure 12: On the left the circles represent the antenna locations of the triggered LOFAR stations. Their color is an indicator of fluence. On the background, the best fitted simulation is plotted. A good indication between measurement and simulation is characterized by the colors within the circles. On the right, an example of the parabola fitting of χ^2 values is showcased. The figure was taken from [49].

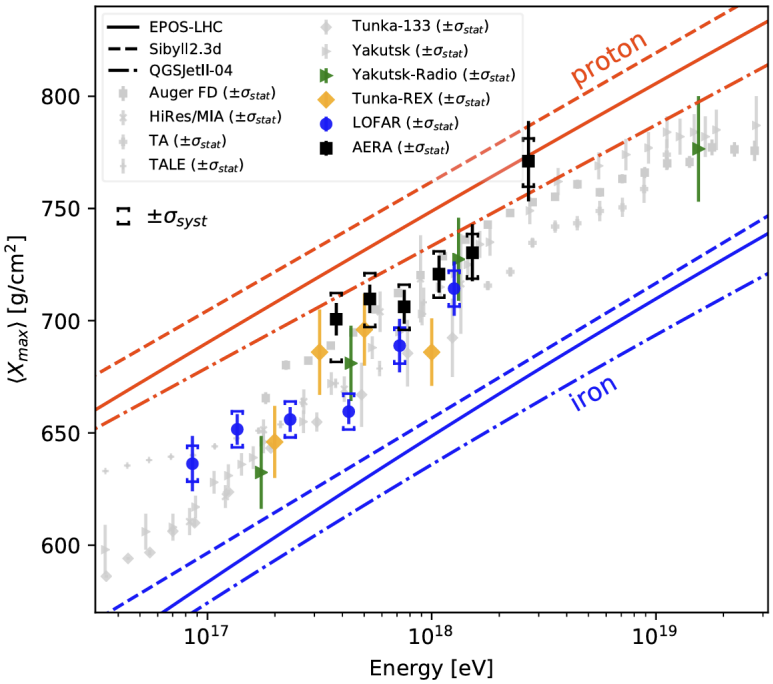


Figure 13: The LOFAR $X_{\max}$ results are compared with the results of other experiment. The discrepancy between LOFAR and AERA is clear. This figure was taken from [53].

its minimum consists the reconstructed $X_{\max}$. In fig. 12 (right) an example of the χ^2 minimization scheme is shown. For the best fitting shower the fluence footprint is shown in fig. 12 (left), where the circles are the antenna positions of the triggered LOFAR stations. The colors of the circles blend in the simulated background, which indicates good agreement between measurement and simulation. From the measurement and simulation the energy of the electromagnetic shower energy can be estimated, to which the radiation energy scales quadratically [34]. The square root of the scaling factor acts as a correction to the simulation energy. Another correction to the simulated radio pulses that is taking place is due to the step size used in CORSIKA 7 and amounts to 11% [52]. This particular effect is also found in the CORSIKA 8 in the form of the track length used by the tracking algorithm and is going to be discussed in section 10. The results of the $X_{\max}$ reconstruction are shown in fig. 13 with respect to other experiments. It is worth pointing out that there is a discrepancy between LOFAR and AERA results which is not yet understood.

As a final remark on this section, one can notice that this reconstruction procedure is making use of the pulse intensity, i.e. the pulse amplitude. The rest of the information that is hidden in the radio pulses is being discarded. A topic of discussion of this thesis is to associate the pulse shape, instead of the pulse amplitude to $X_{\max}$ and develop a reconstruction scheme that it taking advantage of the information hidden there. This is discussed in part III.

Part II

Simulating the Radio Emission from Extensive Air Showers with CORSIKA 8

The limiting factors posed by legacy software like CORSIKA 7 [54] and ZHAireS [55] affect naturally the capabilities of simulations needed for modern experiments that involve complex physical scenarios. The radio emission from air showers crossing media from air to ice for example are already being studied and are proposed to be studied in experiments like ARA [56], ARIANA [45], RNO-G [57] and others. More proposed experiments like SKA [46] and GRAND [58] intend to detect cosmic rays using the radio technique employing thousands of antennas. This will have a huge impact on computing times in order to produce such simulations. Even more exotic cases in the likes of showers produced by τ neutrinos, earth skimming and upward-going showers already push the capabilities of existing software, if not even prohibiting such simulations. Hence, the design and implementation of a modern, flexible and easily upgradable *radio module* early in the development procedure of CORSIKA 8 seems only natural. Such a module aims to takes advantage of modern technologies in order to have a generic enough structure, have a clear design for users but also for potential developers and exploit contemporary computing technologies to minimize running times. While being developed as an integral part of CORSIKA 8, the *radio module* needs to be able to work also as a stand-alone module. In more detail, it needs to be able to simulate and examine phenomena for isolated cases of radiation being emitted by charge particles not necessarily in the context of air showers. Another important use case of the module given its extreme sensitivity to timing information is that of the diagnostic tool for the rest of the CORSIKA 8 code. To put it more clearly, the *radio module* needs to achieve 3 things: 1) Successfully simulate the radio emission from Extensive Air Showers, 2) have the capability to perform comparison studies to gain a better insight into the nature of radiation and radio signal propagation and 3) with its validity established, act as a diagnostic tool for other parts of CORSIKA 8. Such a *radio module* did not exist in CORSIKA 8 before this work. A major goal of this thesis, was to design and develop it successfully in order to use it for radio simulation studies which is the topic of later sections.

This part of the thesis covers the work that was made towards the realization of the *radio module* and the studies that were performed with it. First, the CORSIKA 8 framework is going to be summarized briefly in order for the reader to become familiar with its philosophy. Next, the discussion is moved to the *radio module* itself. Its design logic and how it works internally in CORSIKA 8 will be explained. Specific parts of the module focused on its ability to be easily updated to simulate complex cases are going to be extensively documented. Validation studies for single and pairs of charged particles will be presented. Cases where the module acted as a diagnostics tool and was essential to find logical errors and bugs within the CORSIKA 8 core code will be discussed along with the solutions found. The introduction of its multithreading

capabilities acting as a proof of concept that new technologies can be used will be showcased.

Once the technical work has been extensively presented, the discussion will shift to comparison simulation studies. First, radio simulations with CORSIKA 8 will be compared with existing software that have served the community until today. A systematic study between the "Endpoints" formalism as originally implemented in CoREAS and the ZHS formalism for air showers is going to be made. Finally, the effect of details like different track lengths of the simulated particle tracks and different multiple scattering approaches on radio simulations are going to be put into perspective. Such comparison studies aim to assess the understanding of the radio emission mechanism. A good understanding as can be showcased by simulations indicates that the radio detection technique is indeed reliable for reconstructing $X_{\max}$ and the primary particle energy.

4 The CORSIKA 8 framework

CORSIKA 7 originally developed for the KASCADE experiment [59] has been highly successful in serving a major fraction of the atroparticle physics community for over 30 years. Its FORTRAN monolithic structure along with hand-optimized code design makes its maintenance, the implementation of modern features and utilization of modern technologies like multithreading and GPUs a challenge. To tackle these obstacles the decision was made to make a new, modern and flexible framework written in C++, CORSIKA 8 [60], [61]. The two key design aspects of CORSIKA 8 is *modularity* and *flexibility*. It is comprised of *modules*, and each module executes an individual simulation task with its algorithms being able to exist and be developed independently of other *modules*. As a consequence, *modules* can be exchanged or combined as long as they meet some requirements in a *flexible* way.

The structure of this framework is shown in fig. 14. The main loop is in the *Cascade* class, where a particle from the stack is being propagated and if secondaries are produced they are also put in the stack. *Cascade* has access to the *Tracking* algorithm, the *Environment*, the *Process List* and the stack. The particle trajectories in the *Environment* are dictated by the chosen *Tracking* algorithm. The *Environment* entails quantities such as density, composition, refractive index and the process can query properties from it. The user is able to create the *Environment* to their needs. The *Process List* contains user arranged processes that may or may not affect the development of the shower. The final outcome of the processes put in the *Process List* is being gathered and saved by the *Output*.

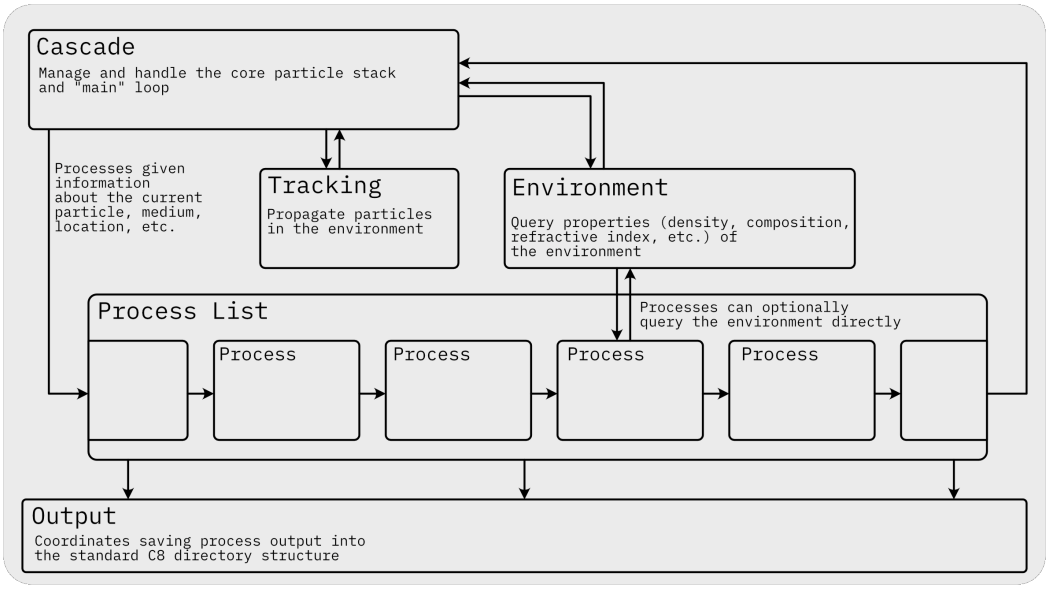


Figure 14: The CORSIKA 8 high-level architecture.

5 Architecture of the radio module

In a simplified way, the main operations of CORSIKA 8 during a simulation are two. The first one is the propagation of particles under the effect of electromagnetic fields and energy losses and the second one is the handling of stochastic processes, interactions and decays that either generate new particles or discard old ones from the particle stack. The accurate event generation such as electromagnetic or hadronic interactions is being handled by external codes that are configured to work within CORSIKA 8. Such computations entail an enormous amount of "book keeping". Modules that can save, process and pass on information to other modules are an integral part of CORSIKA 8 process logic. A clear distinction for the processes that make up the process sequence of a CORSIKA 8 simulation can be made, the "interaction processes" and the "observing processes". The interaction processes are the ones that are capable of altering the trajectory, lifetime and nature of a simulated particle such as energy losses or electromagnetic interactions etc., while observing processes can register and process for their own purposes a simulated particle without altering any of its properties. The radio module and the *radio process* it creates naturally falls into the second category, because the energy losses due to the radio emission can be completely neglected.

The main objective of the radio module is the calculation of the electric field vector received from an electron or a positron given an observer's location. To achieve that, it needs to be able to instantiate a *radio process* in the CORSIKA 8 standard in order for the process to be part of the Cascade algorithm and have access to all the relevant information. Therefore, it has to work with particle tracks which are received sequentially from the running Cascade simulation

and access the constructed simulation environment to query properties relevant for the radio emission calculation. On a more technical level this is achieved by using public inheritance in C++ from the `ContinuousProcess` class to the `RadioProcess` class. The two main methods that each class inherits from `ContinuousProcess` are the `doContinuous()` and the `getMaxStepLength()`. `doContinuous()` feeds of particle tracks and then forwards them further into the *radio process* and `getMaxStepLength()` returns the maximum length allowed of that particle track, which for the case of radio since no constraints are present at the present, is set to infinity. As it is going to be shown in this thesis though, setting a limit to the maximum track length might turn out to be beneficial, since the track length of the particle track can affect the radio simulations.

In order to achieve flexibility and modularity the radio emission calculation is split into conceptually independent, interchangeable parts. These are the *particle track filtering*, the *formalism* responsible for the radio emission calculation, the *radio signal propagator* and finally the *antenna* instance which acts as the observer. The *radio process* being part of the *process sequence* receives as input the particle tracks and is able to get the necessary environmental information from CORSIKA 8 as mentioned above. Then, it proceeds to filter the particle tracks so as to treat the desired ones for the radio emission calculation. The filtered particle tracks are pushed forward to the chosen formalism that is the main workforce of the calculation. It is important to note that this procedure happens sequentially and every particle track is treated separately. Once the necessary information of the particle track is gathered, the particle is looped over all the user configured antennas that form the antenna collection. At this point, within the formalism the chosen propagator is called to calculate the signal propagation from the particle track to the specific antenna location under examination. Through the propagator, the radio module can query environmental information like the refractive index for instance, and calculate the propagation time of the radio signal. Once the calculation for that particle is over, the electric field and timing information are binned and stored in the antenna instance. The same procedure takes place for the same particle track to the next antenna instance in the antenna collection. Finally, this series of steps happens for all the particle tracks that are generated during a simulation. A schematic diagram of this course of action is shown in fig. 15.

Such a module design provides the opportunity to isolate parts of the radio emission procedure and study specific effects without the need to alter the rest of the code. Technically, this is achieved with the use of template parameters in C++. They allow the development of generic code that works with different data types and is evaluated during run time. For the time being, a *radio process* requires 3 template parameters for the formalism, propagator and antenna collection as the track filtering is being handled internally by the `doContinuous()` method. In simple terms, this means that one is able to instantiate multiple *radio processes* during the same simulation with different characteristics. For instance, the same track filters, antennas and propagators can be used with the sole exception being the formalism which is studied in this work. This allows to directly compare different formalisms like "Endpoints" and ZHS eliminating any uncertainty coming from the underlying air shower code, since they are evaluated on an identical air shower. Time domain versus frequency domain studies on the same air shower, once a frequency domain implementation is in place, will be made possible with the instantiation of 2 *radio processes*, one

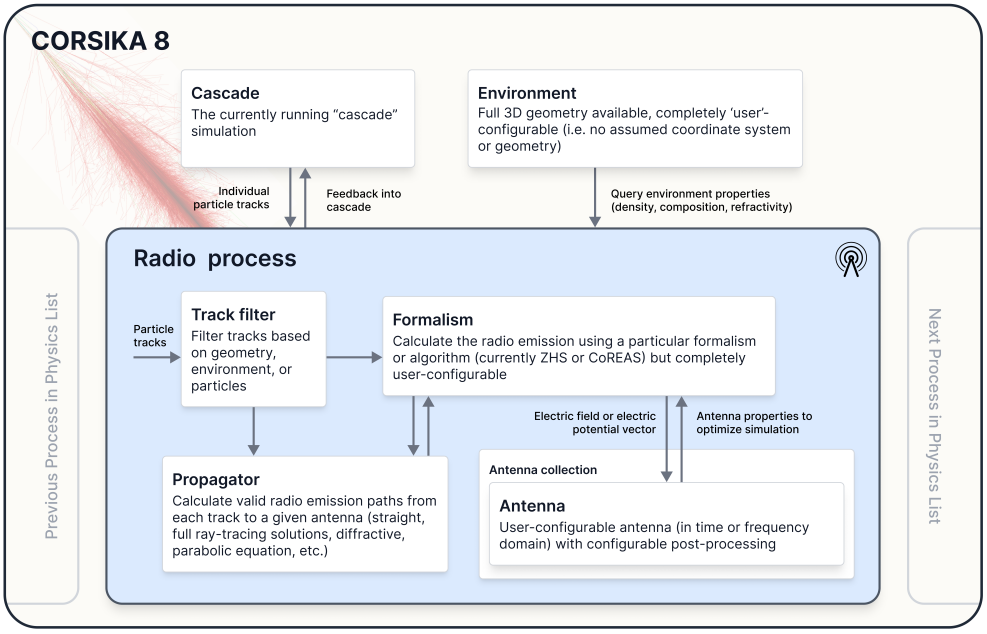


Figure 15: A schematic diagram of the radio process currently implemented in CORSIKA 8 and how it integrates with the CORSIKA 8 framework.

with time domain antennas and one with frequency domain. One could also evaluate the effect on the radio signal simulated by different signal propagation techniques by instantiating *radio processes* with different propagators. All of this structure is going to be illustrated in the next sections using a top-down approach. Starting from the core *radio process* class, all the way to the final step which is the output that the user obtains will be thoroughly discussed. The radio module code can be found in detail at the KIT gitlab page.

5.1 The Radio Process

The Radio Process class is the base interface for any *radio process* created during a simulation. The class constructor requires 3 template parameters for the antenna collection, the formalism and the propagator. As mentioned earlier it has to fit within the CORSIKA 8 framework. Therefore, it has to follow the `ContinuousProcess` and output logic of CORSIKA 8. In listing 1, the main structure of the class is showcased with some methods being omitted for the sake of simplicity. Being a `ContinuousProcess` the class inherits the `doContinuous()` and `getMaxStepLength()` methods, with the `doContinuous()` pushing forward the actual radio emission calculation. The methods `startOfLibrary()`, `endOfShower()` and `getConfig()` exist to facilitate and follow the general output format guidelines used widely in CORSIKA 8.

Listing 1: The general structure of the Radio Process class.

```
template <typename TAntennaCollection, typename TRadioImpl, typename TPropagator
    >
class RadioProcess : public ContinuousProcess<
                         RadioProcess<TAntennaCollection, TRadioImpl,
                             TPropagator>>,
                     public BaseOutput {
  \\ [ ... ]
  template <typename Particle>
  ProcessReturn doContinuous(Step<Particle> const& step, bool const);
  \\ [ ... ]
  template <typename Particle, typename Track>
  LengthType getMaxStepLength(Particle const& vParticle, Track const& vTrack)
      const;
  \\ [ ... ]
  void startOfLibrary(boost::filesystem::path const& directory) final override;
  \\ [ ... ]
  virtual void endOfShower(unsigned int const) final override;
  \\ [ ... ]
  YAML::Node getConfig() const final;
  \\ [ ... ]
}; // END: class RadioProcess
```

Currently the `doContinuous()` method is a wrapper method for the `simulate()` method, which is the main method of the formalism. There the particle is identified and if it is an electron or a positron it is pushed further to the `simulate()` method to be treated by the desired radio calculation algorithm. This is the particle track filtering taking place and the distinction of the `doContinuous()` and `simulate()` methods allows for further improvements of the filters. Ideally, in a future upgrade of the code the filters will be a user configurable `std::vector` with `bool` values that allow particle tracks with specific features to be pushed into the `simulate()` method. These features can be particles that have traversed a specific amount of matter, or exist in a specific part of the atmosphere, or are within a specific energy range. As mentioned earlier, the `getMaxStepLength()` method is a mandatory method of the continuous processes that in the case of radio is set to return infinity as radio does not currently constrain in any way the maximum step length of the particle track.

The remaining 3 methods are related to the output manager that has been developed for CORSIKA 8. The output uses the parquet format [62] for large files and YAML format [63] for smaller files like configuration files. `startOfLibrary()` is the method that initializes the parquet streamer for the radio output files. It creates a directory named after the user configured name of the *radio process* instance and within it a single file called `antennas.parquet`. For filling this file the `endOfShower()` method is responsible. It creates a single matrix where the first column is the *time* information and the other 3 are the electric field vector components, namely E_x, E_y and E_z. All the antenna information of an antenna collection is stored there. The number of rows is dictated by the sampling frequency and the detection duration of each antenna. Finally, the `getConfig()` method produces a YAML configuration file with all the relevant information for the *radio process* and the antenna collection. In more detail, it states which formalism was used for the *radio process* and the units of time, frequency, electric field and distance. Next, it

provides antenna information like antenna type (i.e. time domain or frequency domain), name and location for all the antennas of the antenna collection.

5.2 Formalism

The formalism is the main work force of the radio module. It is the underlying algorithm with which the contribution to the electric field of each particle track is calculated. At the moment, 2 formalisms are fully implemented. These are the CoREAS algorithm [64] based on the "Endpoints" formalism [40] which is also used in CORSIKA 7 and the "ZHS" algorithm [41] ported as closely as possible from ZHAireS [55]. From now on, these formalisms will be simply referred also as CoREAS and ZHS respectively. The main feature of the formalism logic is the `simulate()` method which is a `ProcessReturn` type as defined in the `ContinuousProcess` class. This means that once the method is successful it returns to the Cascade run a process type `ProcessReturn::Ok`, which allows for the rest of the simulation to go on. Apart from the `simulate()` method, helper functions that construct *radio process* objects that explicitly use a chosen formalism have been implemented to hide the template complexity from the user.

It is worth touching upon the train of thought of the `simulate()` method, as it is independent of the formalism chosen to calculate the electric field vector (or vector potential for the case of ZHS) and it might facilitate the understanding of potential future developers. Its input are the particle tracks from which it is able to get information like velocity, position, direction and time. Next, it performs checks on whether the particle track is valid to proceed with the calculation. For the time being it checks whether the particle ID corresponds to an electron or a positron. If yes, the particle track is looped over all antennas of the antenna collection and is hence treated for each antenna separately. At this stage, the chosen propagator is called which determines the signal propagation from the position of the particle track to the antenna location and returns a *signal path collection*. The code design allows for a ray to have multiple propagation paths whose number is dictated by the propagator. This design choice is being made in order for the module to be able to treat cases such as dense media with refractive index gradients where two propagation paths are typically expected. For every signal path contained within the *signal path collection*, quantities like the propagation time and relevant environment attributes are picked up from the *signal path collection*. Once all the information is at hand, the formalism proceeds to calculate the electric field vector and the detection time and arrival direction vector of the radio signal which are then forwarded to the antenna. This procedure is showcased in listing 2.

Listing 2: A simplified layout of the key steps of the simulate method.

```
template <typename TRadioDetector, typename TPropagator>
template <typename Particle>
inline ProcessReturn FormalismName<TRadioDetector, TPropagator>::simulate(
    Step<Particle> const& step) {
  \\ [ ... ]
    // loop over each antenna in the collection
    for (auto& antenna : antennas_.getAntennas()) {
```

```
        // ask for signal path collection thorugh the .propagate() method for
        // a particle at its position through the antenna location
          auto Paths{this->propagator_.propagate(particle, particlePosition,
                                                 antenna.getLocation())};
      \\ [ ... ]
          // Loop over signal path collection
          for (size_t i{0}; i < Paths.size(); i++) {
          \\ [ ... ]
          Electric Field Vector or Vector Potential formula
          \\ [ ... ]
          // pass the information to the antenna through the .receive() method
          antenna.receive(detection time, electric field vector or vector
          potential, signal arrival direction vector);
          } // END: loop over signal path collection
    } // END: loop over antennas
    return ProcessReturn::Ok;
  }
} // END: simulate()
```

5.3 Propagator

5.3.1 Generic structure of the propagator class

Crucial for the electric field contribution of a particle traversing through matter is the behavior of the radio wave propagated to the antenna. The exact signal propagation behavior is dictated by the `propagator` object which is completely decoupled from the formalism. This way, one is allowed to make approximations and studies for their own experimental needs. In the propagators the simple case of the straight ray approximation is considered, in which the signal is propagated in a straight line from the source to the antenna. This is common practice for all air shower simulations that make use of the radio detection technique so far. To develop a future proof and efficient radio module means introducing the inclusion of a design that takes full advantage of the CORSIKA 8 volume tree environment building. As a result, propagation phenomena like diffraction, refraction, reflection and even more complex phenomena that occur in specific media can be easily included in the propagator logic without the need to alter the rest of the radio module. The generic structure of a `propagator` class is shown in the listing 3 below.

Listing 3: Generic structure of the propagator class.

```
// class constructor
template <typename TEnvironment>
inline Propagator<TEnvironment>::Propagator(TEnvironment const& env)
    : RadioPropagator<Propagator, TEnvironment>(env) {
      \\ [ ... ]
} // END: class constructor

// propagate method responsible for the signal propagation calculation
// between 2 points
  template <typename TEnvironment>
```

```
template <typename Particle>
inline typename Propagator<TEnvironment>::SignalPathCollection
Propagator<TEnvironment>::propagate(Particle const& particle,
                                    Point const& sourcePosition,
                                    Point const& destinationPosition) {
  \\ [ ... ]
  return {SignalPath(time, averageRefractiveIndex_, refractiveIndex_source,
                     refractiveIndex_destination, emitionVector,
                     receiveVector, geometricalDistance_, points)};
} // END: propagate()
```

The class inherits directly from the `RadioPropagator` parent class, whose sole argument is the user constructed environment in a template parameter to make sure it works with any kind of constructed environment. This way, any implementation of a `propagator` class has direct access to the environment to get all the information it needs from that. In the constructor, the developer is free to implement the propagator specific properties. The main method of the class is the `propagate()` method. Its input is the particle whose signal is being propagated and is passed on by the formalism. Along with it are 2 points, one being the position of the source particle and the position of the observer, or in our case the antenna. In this method, the way the signal is going to be propagated is fully treated. The return object is a `SignalPathCollection`, which as the name suggests is the collection of paths that the signal has traversed through its propagation. In more technical terms it is a `std::vector` of `SignalPath` instances depending on the number of the paths that the signal propagated through. This particular design choice is what allows the radio interface to handle multiple paths that include reflection, refraction etc. Cases like showers crossing media can be successfully simulated with this design [65].

5.3.2 The Signal Path

Regarding the `SignalPath`, it is simply a `struct` that acts as a container of information that the `propagator` returns to the formalism for the radio emission calculation. The information it holds for the time being, is the propagation time of the signal, the average refractive index between source and observer locations, the refractive index for source and observer locations separately, launch direction unit vectors at the point of emission and receive direction at the point of reception, the geometrical distance along the path of these two points and finally a `std::deque` of points that the signal has traversed. This can be very easily upgraded as shown in listing 4 in case some formalism in the future needs to have access to more information, but having in mind CoREAS and ZHS, this information is sufficient.

Listing 4: Information within the SignalPath

```
\\ [ ... ]
struct SignalPath final : public Path {
\\ [ ... ]
/**
* Create a new SignalPath instance.
*/
```

```
SignalPath(TimeType const propagation_time,
          double const average_refractive_index,
          double const refractive_index_source,
          double const refractive_index_destination,
          Vector<dimensionless_d> const& emit,
          Vector<dimensionless_d> const& receive,
          LengthType const geometrical_path_length,
          std::deque<Point> const& points);
}; // END: class SignalPath final
```

5.3.3 Propagators implemented so far

Currently, there are three fully tested and functional `propagator` classes. They all use the straight ray approximation, which means that they assume that the signal is being propagated along a straight path from the emission to the detection point. Under this assumption the propagation time is being calculated. Even though all 3 propagators use the same assumption they are fairly different and have different use cases. The first one is the `DummyTestPropagator` which is meant to be used for idealized scenarios like for instance an environment consisting of a uniform refractive index. The advantage of using such a `propagator` is that it is very fast, since it only has to access the CORSIKA 8 universe to get the refractive index twice and calculates the propagation time through the formula $t = n\frac{R}{c}$, where n is the refractive index, R is the distance between the 2 points and c is the speed of light. To use this `propagator` in the *radio process* the user can simply utilize the corresponding helper function `make_dummy_test_radio_propagator(environment)` which takes only the environment as an argument.

The second `propagator` is the `NumericalIntegratingPropagator` which as the name suggests performs a numerical integration using a user-configurable stepsize along the propagation path. For this, the standard Simpson's rule is used. The `propagator` depending the stepsize figures out the number of sub-paths that it has to subdivide the initial straight path from the emission to the detection point. For the last sub-path it makes sure to stop exactly on the observer's location by either adding the missing distance or subtracting the additional. The advantage of this particular `propagator` is that it can be used with any setup in the environment (e.g. where Eath's curvature needs to be taken into account) and has the potential to be very accurate, but this comes at a great cost in terms of performance. The relevant helper function is the `make_numerical_integrating_radio_propagator(environment, stepsize)`, with the environment and integration stepsize as arguments.

The final fully implemented `propagator` is called `TabulatedFlatAtmospherePropagator`. This is the best of both worlds solution when it comes to a balance in accuracy and performance. It comes with its limitations because it assumes the atmosphere to be flat, and it only works for relatively vertical air showers (zenith angle $< 60°$). In its constructor, a table for refractivity is created between a user defined lower and an upper limit with the user desired step. The lower limit is automatically set 1 km below the user given value to take into account particles penetrating the ground. From this table another table for the integrated refractivity is generated which is later

used for the calculation of the propagation time. This is quite fast because it needs only initialize the tables during construction. In the `propagate()` method the table indexes that correspond to the positions of the emission and detection points are worked out and the value of refractivity is simply accessed through the table rather than the CORSIKA 8 universe which is costly. If one of the points (or both) lie outside the edges of the table, then the value is extrapolated via an extrapolation scheme which is defined in the class constructor. This propagator can be utilized via its helper function `make_tabulated_flat_atmosphere_radio_propagator(environment, upperLimit, lowerLimit, step)`, with given environment, upper, lower limit and step.

5.4 Antenna Collection

Part of the radio detection technique is the detector itself. Many cosmic ray experiments use all sorts of antenna models depending on their hardware to fold the simulated traces and match as closely as possible their detected signals. In the simulation realm, we work with ideal detectors that store the individual electric field contributions to a user configured detection window. On the top level of the radio module, the `AntenaCollection` is found. This is a container class for `antenna` objects. The user can place the desired number of antennas anywhere in the 3 dimensional environment constructed. The antennas are then put into the `AntenaCollection` container, where they can be accessed, called or reset one by one. All together, they form the detector of the simulation where all the radio signals emitted by the simulated air shower are going to be detected.

The parent class `Antenna` that fills the `AntennaCollection` provides the main characteristics of an antenna and uses the `<typename TAntennaImpl>` as a template parameter to allow the inclusion of antennas in different domains, i.e. time or frequency domain. Its characteristics are its name, location and coordinate system that are initialized in the class constructor. Each antenna can have its own coordinate system and the orientation of x, y and z can be determined by it. The class methods are shown in the listing 5 below. The user has access to all the attributes of the antennas via the corresponding *getter* methods. `getAxis()` returns a reference to the x-axis label, namely time or frequency information depending on the antenna implementation and the x, y and z polarizations are accessed through their `getWaveform(X,Y,Z)()` method. The antenna is also able to reset itself via the `reset()` method which erases all the information gathered. The `implementation()` returns the antenna implementation in order to be able to identify its type. Finally, one of the most important methods is the `receive()` one. For this, variadic templates have been used in order to provide the flexibility to its derivative classes to receive types of data suitable based on their implementation.

Listing 5: Overview of methods for the antenna class.

```
// class constructor
Antenna(std::string const& name, Point const& location,
        CoordinateSystemPtr const& coordinateSystem);
\\ [ ... ]
template <typename... TVArgs>
```

```
void receive(TVArgs&&... args);
\\ [ ... ]
Point const& getLocation() const;
\\ [ ... ]
std::string const& getName() const;
\\ [ ... ]
void reset();
\\ [ ... ]
axistype getAxis() const;
\\ [ ... ]
std::vector<double> const& getWaveformX() const;
\\ [ ... ]
std::vector<double> const& getWaveformY() const;
\\ [ ... ]
std::vector<double> const& getWaveformZ() const;
\\ [ ... ]
TAntennaImpl& implementation();
```

For the time being, a time domain antenna where the electric field contribution is recorded as a function of time is implemented. This way information is being thrown away about the direction of emission. As long as the direction of emission does not vary too much, and the antenna response is reasonably the same for all relevant signal arrival directions, this approximation is acceptable. If this is not the case though, then the antenna characteristics would need to be taken into account already when collecting the electric field contributions. This is not done at the time being, but such a modification in the future is foreseen by the code design. The time domain antenna inherits attributes from the `Antenna` parent class, but also has a few more parameters. These are the start detection time, the time duration it collects data, its sampling frequency and a timing information on the earliest possible signal that can be detected by that antenna. This is used to make sure that the window of detection given in the output is correct and the antenna does not miss any contribution from the air shower particles. The `receive()` method dictates how the signal is being stored in the antenna. It is an overloaded method so as to be able to accommodate the reception of both electric field vectors and vector potentials depending on what the formalism calculates. It gets as input the timing information of the signal and its contribution in the form of a vector object whose x, y and z components are determined by the coordinate system of the antenna. Then it figures out the correct time bin and stores all polarizations of the signal. Finally, the `TimeDomainAntenna` class has a built in `getConfig()` method that writes out in the *radio process* configuration YAML files the antenna specific information like type, start detection time, detection duration, sampling frequency and its location.

5.5 Handling and reading the output

CORSIKA 8 has its own output manager which provides an interface for the *processes* to build upon and dictate how the data is going to be written to disk. This interface was used in order for the radio module to write out the output in a consistent way with the rest of the *processes*.

Every *radio process* produces a subdirectory inside the default simulation directory named after the name given to the *radio process* by the user. There exist the `antennas.parquet` binary files and the YAML configuration files. In the `antennas.parquet` files the timing information and the electric field components of all antennas are stored. Similarly, the relevant antenna information like type, name, location etc of all antennas are stored in the YAML configuration files.

To read the output files one can use the CORSIKA 8 python library. By using this library one can choose to read the radio output using `pandas` [66]. By doing so a multi-dimensional `pandas` dataframe is generated. The first column is the shower number to accommodate the case of multiple shower runs in a single CORSIKA 8 simulation. The second column is the antenna name, the 3 next columns are the X,Y and Z polarization components labeled as 'E_x, E_y and E_z' and the final column is the timing information of the antenna labeled as 'time'. Helper functions are implemented to give more information about the *radio process*. These are the `get_antenna_list()` which returns a list with the names of all the antennas in the simulation, the `get_units()` which as the name suggests returns the relevant information of the units of the antennas.

A standard way to get the module output is shown in the listing 6 below. The `corsika` and `pandas` libaries have to be imported. One first needs to initialize a library where the CORSIKA 8 output directory exists. Then, the `corsika` library provided getters can be used given the *radio process* name and desired output format. Currently, when using the CORSIKA 8 python library, the radio module output can be read solely utilizing `pandas` dataframes, but the interface allows for the inclusion of more formats if needed. Apart from accessing the data, helper functions are implemented that return important information like units and antenna attributes to the user.

Listing 6: Python script example on how to access radio output.

```
# import the necessary libraries
import corsika
import pandas as pd

# access the output of the shower simulation by providing the
# directory it exists
library = corsika.Library('./directory_to_corsika_output/')

# get the radio process data in the form of a pandas dataframe
radio_data = library.get('radio_process_name').astype('pandas')

# access time and polarization information of a specific
# antenna of the specified simulation run
time = radio_data['shower_number']['antenna_name']['time']
Ex = radio_data['shower_number']['antenna_name']['Ex']
Ey = radio_data['shower_number']['antenna_name']['Ey']
Ez = radio_data['shower_number']['antenna_name']['Ez']

# return a pandas dataframe on the antenna information of
# the specified radio process - each antenna can also be
# accessed separately by providing its name
```

```
antenna_info = library.get('radio_process_name').get_antennas()

# return a list of all antenna names for quick reference
antenna_list = library.get('radio_process_name').get_antenna_list()

# return the units for all antennas
units = library.get('radio_process_name').get_units()
```

6 Radio module validation and diagnostics

The radio module is designed to be able to simulate the radio emission from air showers for both standard and more complex scenarios. Apart from that, as shown in section 6.1 it is perfectly capable of acting as a stand-alone module and inspect very specific physical cases that involve as few as one particle. This can help our understanding of the radio emission with small scale simulation studies, where one is able to study the dependency of the radio emission on properties like the geomagnetic field value and the atmospheric density. In recent work [30], it has been shown that for very inclined showers and depending on the geographical site the well established paradigm that the radio emission consists of two mechanisms, namely the transverse current and the charge excess contributions is no longer valid. With CORSIKA 8, contributions of separate electrons and positrons in frequencies up to 1 GHz are studied here, along with their radiation pattern on the ground.

Being part of the early development phase of CORSIKA 8, the opportunity arose to use radio as a diagnostics tool for the rest of the core code. Although significant testing has been done and is still ongoing with any major update of the code, the fact that simulating the signal pulses is extremely sensitive to sensitive to details of the spatial, energy and momentum distributions of electrons and positrons, reveals potential unwanted behaviors or even bugs in the core code. The module has been validated and gives correct results, hence the logical next step is to proceed and simulate air showers. As will be shown, this was a challenge due to the early development stage of CORSIKA 8 as a whole. By simulating radio pulses we were able to identify and fix logical errors and bugs within CORSIKA 8 and further improve the main interface. In the next subsections, the benefits of developing radio as an integral part for CORSIKA 8, along with a few case studies will be discussed.

6.1 Module validation

CORSIKA 8 poses strict guidelines when it comes to testing and validating its modules and all the included classes. Extensive unit testing has been made to ensure that individual classes and methods have the desired functionality. The successful outcome of those is automatically documented on the coverage report. However, before proceeding into full blown air shower simulations it needs to be established that the radio module gives correct results in a well understood physical scenario both as a stand-alone module and as an integral part of CORSIKA 8. For this validation test,the case shown in fig. 16 is considered.

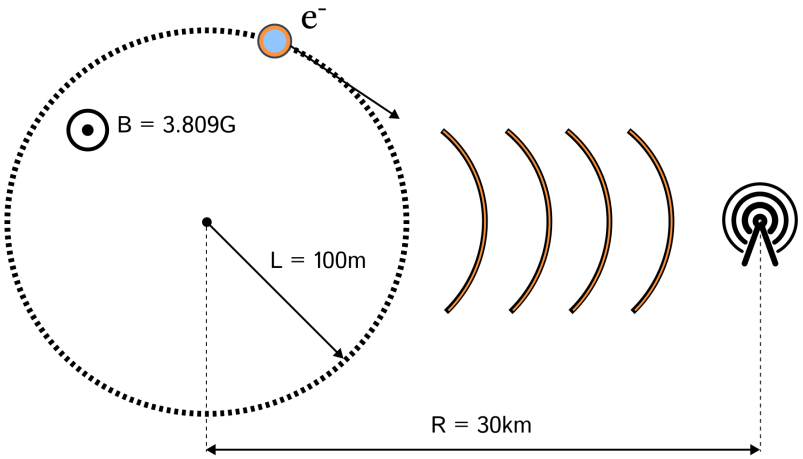


Figure 16: An ultra-relativistic electron performing a circular loop in a uniform magnetic field. Synchrotron radiation is emitted and detected by an antenna placed at a large distance of 30 km.

An ultra-relativistc electron with $\beta = \frac{u}{c} = 0.999$ and energy 11.4 MeV performs exactly one circular loop in vacuum ($n = 1$) in a uniform magnetic field of strength $B = 3.809$ G. The circular track with a radius of $L = 100$ m lies on the $x - y$ plane and the magnetic field is aligned in the $z - axis$. Such a motion produces synchrotron radiation which is detected by an antenna placed at a distance $R = 30$ km from the center of the circular track to ensure that the calculation takes place in the far field regime. The goal of such a simulation is two-fold. First, we want to simulate the synchrotron radiation with both formalisms implemented in the radio module and compare the outcome with an analytical solution. And second, this scenario needs to be simulated in two ways. One is with the radio module being isolated from the rest of the CORSIKA 8 code, propagating the electron on the circular path manually, and another way is with the inclusion of all the machinery provided by CORSIKA 8, where the tracking algorithms within CORSIKA 8 do the particle propagation in the magnetic field. Both simulations need to agree within reasonable limits of course. This way we establish that the radio module produces physical results both by itself and as an integral part of CORSIKA 8.

The same test was also considered in [40]. The most common analytical solution for the

power spectrum of the synchrotron radiation is:

$$P(v) = \frac{\sqrt{3}e^3 B}{m_e c^2} F(x) \tag{22}$$

$$F(x) = x \int_x^\infty K_{\frac{5}{3}}(\sqrt[3]{2.25x^2}) \tag{23}$$

Where $P(v)$ is the product of the normalization constant and the dimensionless function $F(x)$. The ratio of the frequency v to the critical synchrotron frequency $v_c = \frac{3\beta c\gamma^3}{2\pi r}$ is $x = \frac{v}{v_c}$. And finally, K is a modified Bessel function of the second kind. This validation test is performed in the time domain, as the formalism and antenna have been implemented in the time domain within the radio module.

As already mentioned, the purpose of this test is two-fold. The radio module runs with both formalisms without using most of the machinery of CORSIKA 8. The electron trajectory is manually constructed, in which the electron performs the circular loop. This is achieved by introducing points of a circle in the $x-y$ plane and connecting those with straight track segments. Two points connected by the straight track, constitute the *particle track* that is the building block of CORSIKA 8 simulations. For each of these tracks, the `doContinuous()` method of the radio process is manually called and the resulting pulses are stored in the antenna instances. This way, the radio module does not rely on CORSIKA 8 to create a physical scenario and we can validate that the module as a stand-alone entity behaves as expected. This validation test is referred as "manual tracking". The second aspect that needs validating is whether the radio module works correctly when introduced to the CORSIKA 8 machinery. This is important to validate as this is the way it is intended to be used for air shower simulations, i.e. as another physics process in the *process sequence*. In this test the same environment is constructed the same way as before. The electron is set to have a gyroradius of 100 m and with known charge and mass the angular frequency, momentum and energy are calculated through the well known formulas:

$$\omega_g = \frac{|q|B}{m} \tag{24}$$

$$p = |q|Br_g \tag{25}$$

$$E = \sqrt{pc + (mc)^2} \tag{26}$$

The electron with this momentum is then injected at the circle and is allowed to gyrate for 1 period. This is dictated by the `TimeCut process` which forces the CORSIKA 8 *process sequence* to start at 0 sec and stop exactly after 1 period. The circular track is then constructed by the CORSIKA 8 magnetic field tracking algorithm called "Leapfrog Curved". This case is referred as "C8 tracking".

To approximate the analytical solution sufficiently it is important to make the circular trajectory fine enough, i.e. approximate the circle with enough particle track segments so as

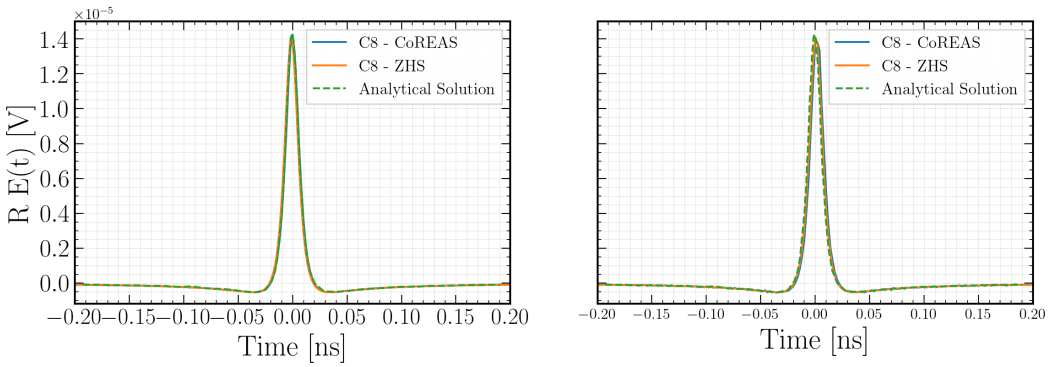


Figure 17: Radio pulse from an electron in a uniform magnetic field performing a circular loop. “Manual tracking” algorithm (left) and “C8 tracking” (right). Both tracking algorithms are compared with the analytical solution of that system presented in [40]. In this case the "Manual tracking" 100000 points have been used, while the maximally allowed deflection angle for the "C8 tracking" is set to 0.0001 rad. Both CoREAS and ZHS agree with the analytical solution.

the resulting pulse appears continuous, which depends on the highest frequency one wants to correctly describe [40]. For the "manual tracking" one is able to chose the number of particle tracks. For the "C8 tracking", the same outcome can by achieved by adjusting the maximally allowed deflection angle (maxRad) for one tracking step within the "Leapfrog Curved" algorithm. The smaller the maximally allowed deflection angle, the more particle tracks are produced making the circular trajectory finer. In fig. 17, a very fine circle has been produced for both tests. In the case of "manual tracking", 100000 points have been set to construct the circle and for "C8 tracking" the maximally allowed deflection angle is set to 0.0001 rad. The pulses produced match very closely the analytical solution, from which we conclude that both formalisms deliver correct results.

For the case of 1000 points for the "manual tracking" and 0.1 rad for the "C8 tracking" the resulting pulse is shown in fig. 18. The signal is very noisy and even the pulse amplitude is off with respect to the analytical solution. This illustrates that the circle needs to be approximated finely enough.

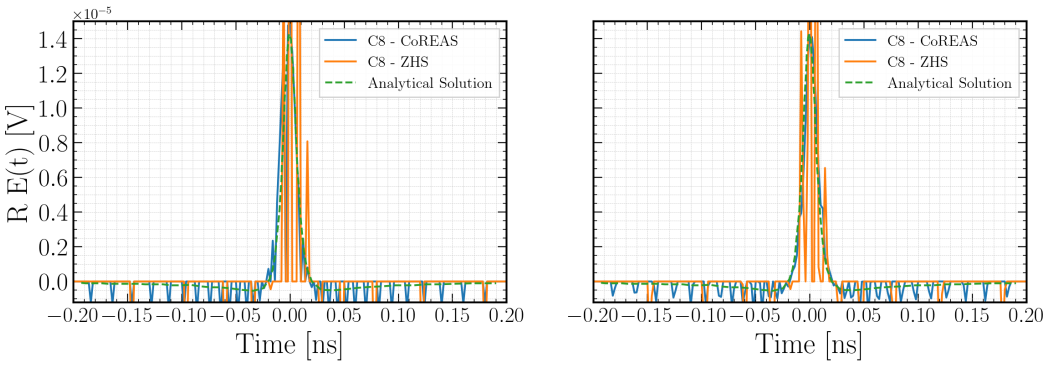


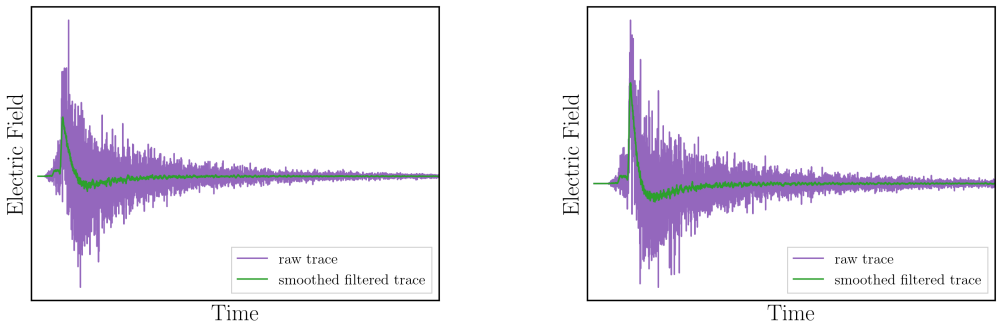
Figure 18: Radio pulse from an electron in a uniform magnetic field performing a circular loop. “Manual tracking” algorithm (left) and “C8 tracking” (right). Both tracking algorithms are compared with the analytical solution of that system presented in [40]. In this case, for the "Manual tracking" 1000 points have been used, while the maximally allowed deflection angle for the "C8 tracking" is set to 0.1 rad. The simulated pulses appear noisy due to the insufficient number of particle tracks simulated.

6.2 Tracking algorithm checks

For the first air shower simulations that the new radio module within CORSIKA 8 was used, the simulation complexity was kept to a minimum. Low energy 100 TeV electron induced vertical showers in the US standard atmosphere were simulated, with the refractive index set to be uniform ($n = 1.000327$) and the geomagnetic field of 50 µT to be constant and horizontally aligned in the x direction.

As can be seen in fig. 19, the dominant polarization of the raw traces originating from CORSIKA 8 for two antennas at 100 m and 200 m north have an unexpected behavior. For illustrating purposes the pulses are plotted in arbitrary units as the main issue studied at this level was the fact that they appear to be rather noisy, see fig. 19. The structure of these noisy pulses indicates that there is signal. To confirm, a Savitzky-Golay filter [67] is applied to smooth out the noisy traces, and indeed there is a "hidden pulse" in the noise. Given the successful validation tests in section 6.1, some extra checks in the magnetic field "Leapfrog" tracking algorithm that is used in CORSIKA 8 are performed. These checks are aimed towards inspecting the accuracy of the tracking algorithm regarding position and time.

The noise in the raw traces might indicate either a problem of accuracy or timing in the tracking algorithm. The key variable to control accuracy in the tracking algorithm is the *maximally allowed magnetic deflection angle* (maxRad). The default value in both CORSIKA 8 and CORSIKA 7 is 0.2 rad, however it is worth noting the tracking algorithms between the two simulation codes are *not* the same. Utilizing the algorithms devised in section 6.1 for the synchrotron emission one can compare how close the CORSIKA 8 "LeapFrog" tracking is performing with respect to the manual tracking which is strictly dictated by the equation of



(a) 100 m. (b) 200 m.

Figure 19: Two pulses at different antenna locations plotted in arbitrary units to showcase the initial behavior. The raw traces resulting from CORSIKA 8 are plotted in purple and the same traces smoothed out with a Savitzky–Golay filter [67] are plotted in green.

circle and not the magnetic field. It is clear from fig. 20 that the larger the maximally allowed magnetic deflection angle the less accurate the circular track is depicted, as it deviates from the set gyroradius of 100 m.

In fig. 21 the deviation of the set gyroradius value for the maximally allowed magnetic deflection angle being set to $0.2, 0.1$ and 0.01 rad is shown. As expected, the smaller the angle the more points the "LeapFrogCurved" algorithm produces and hence increases accuracy.

Next, the difference in $\beta = \frac{u}{c}$ between the end and the start of the track for different values of the maximally allowed magnetic deflection angle is also inspected. Both CoREAS and ZHS assume that the track segment is straight and hence the velocity remains unchanged on both its ends. So, if a big difference was observed here that could explain the noise in the simulated pulses. However, in fig. 22 it is shown that the difference regardless the deflection angle value is extremely small.

As a final consistency check for the tracking algorithm, a circular trajectory consisting of the same number of points using both the CORSIKA 8 "LeapFrogCurved" and "manual" tracking is constructed. The duration the charged particle needs to get from one point to another with both tracking methods is calculated. Then, the duration calculated with "LeapFrogCurved" is subtracted from the duration calculated with the "manual" tracking. Ideally, this difference must be extremely small to prove that the CORSIKA 8 tracking does not diverge from the analytical solution. And as can be seen in fig. 23, this is indeed the case. However, the offset for 0.1 rad appears to be larger than 0.2 rad, which is counter intuitive, given that smaller tracks should be able to approximate the analytical solution better.

The CORSIKA 8 tracking algorithm given the changes in accuracy and the strong magnetic fields that were studied behaves as expected especially for very small values of the maximally allowed deflection angle. It would be beneficial for CORSIKA 8 though, that the

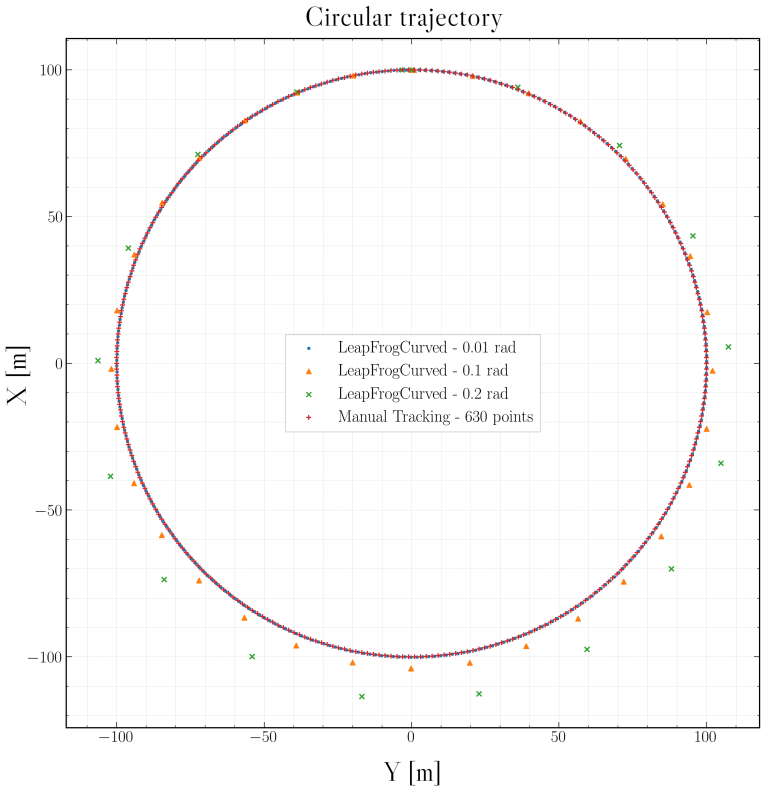
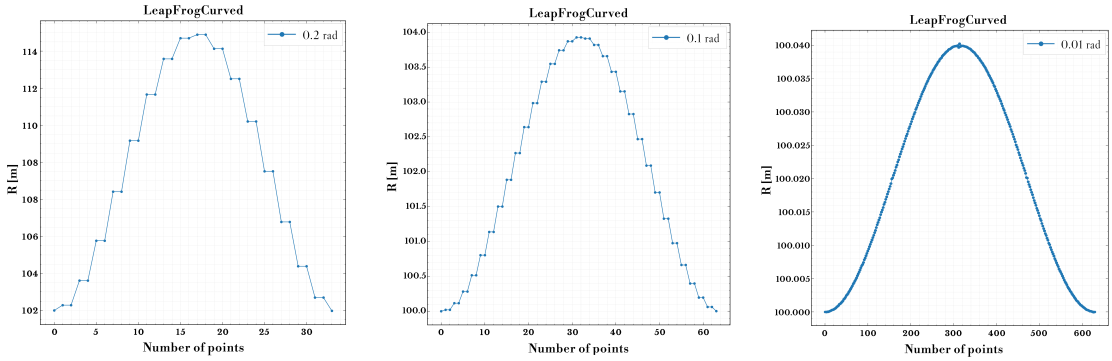


Figure 20: A circular trajectory produced by the LeapFrogCurved CORSIKA 8 tracking algorithm for different values of the maximally allowed magnetic deflection angle. The same circle has been constructed using the "manual tracking" with 630 points.

tracking algorithm is revisited at some point and probably even tested further. In this section, a validation test devised initially for the radio module was used for other parts of the CORSIKA 8 code. Coming back to the noisy pulses though, the maximally allowed magnetic deflection angle was set to 0.0001 rad and a 10 TeV electron induced air shower was simulated and compared with CORSIKA 7 and ZHAireS [68]. Setting such a fine resolution mitigated the noise issue (fig. 24) but a heavy penalty in computation times had to be paid. This combined with the disagreement in amplitude for antennas further away from the shower core clearly indicated that the issue was not resolved. Especially when considering the much lower computing times of CORSIKA 7 where the maximally allowed magnetic deflection angle is set to 0.2 rad and the noise there is absent.

While inspecting the effect of the value of the maximally allowed magnetic deflection angle an interesting question occurred. In CORSIKA 7 this parameter is set by default to 0.2 rad. Changing this value affects the shower development, so even by using the same random number seed, two showers with a different maximally allowed magnetic deflection angle setting will be different. As will be extensively studied in chapter 10, this has some implications for the



(a) 0.2 rad. (b) 0.1 rad. (c) 0.01 rad.

Figure 21: The values of R for different values of maximally allowed magnetic deflection angle. The smaller the angle the less R deviates from the set value of 100 m. Note that the y-scale is different for all three plots.

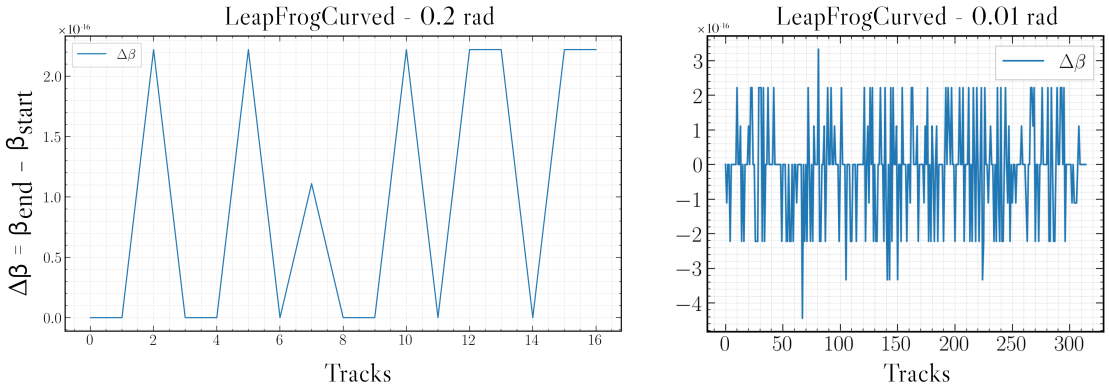


Figure 22: Difference in β between start and end point of each track for maximally allowed magnetic deflection angle set to 0.2 rad (left) and 0.01 rad (right).

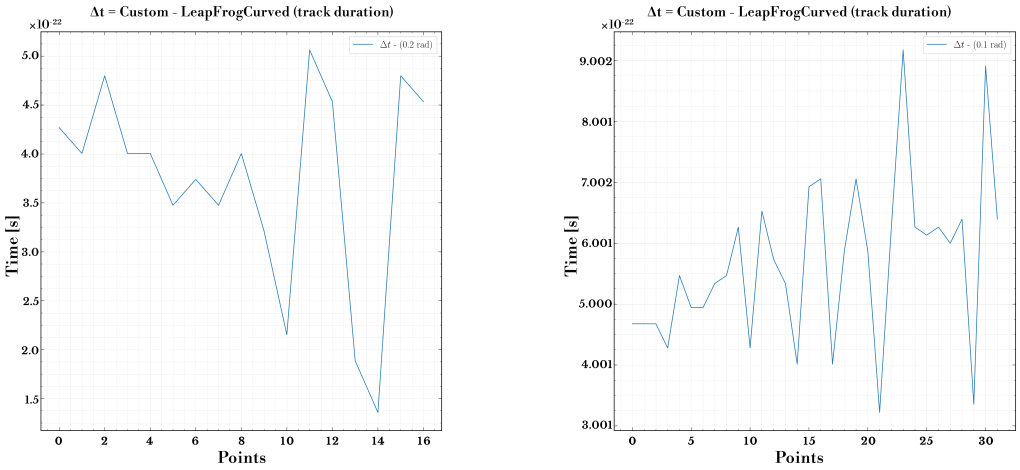


Figure 23: Difference between the track duration of the same tracks calculated with "manual" and CORSIKA 8 "LeapFrog" tracking for maximally allowed magnetic deflection angle set to 0.2 rad (left) and 0.1 rad (right).

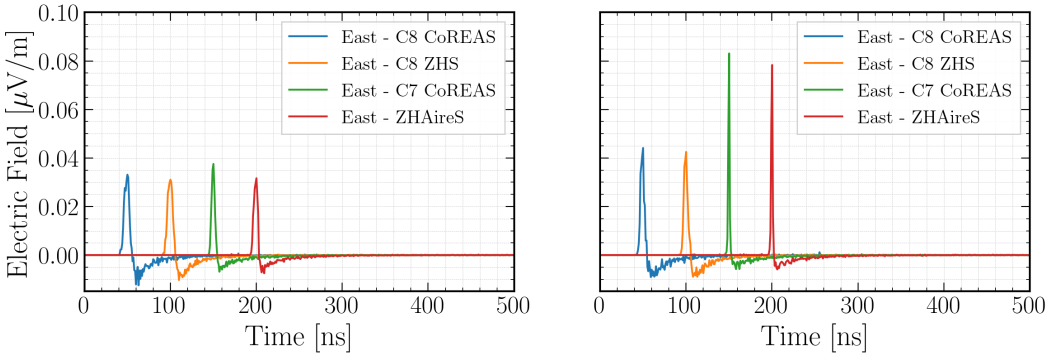


Figure 24: Signal pulse comparison of dominant polarization for antenna at 50 m (left) and at 200 m (right) from the shower core. Maximally allowed magnetic deflection angle was set to 0.0001 rad and the time offsets are arbitrary. The simulation details can be found in [68]

radio simulations. Another important remark, is that since the tracking algorithms are different in CORSIKA 8 and CORSIKA 7, setting the same value for the maximally allowed magnetic deflection angle in both simulation codes might still not guarantee "identical" outcomes.

6.3 Cascade bug and Step object

The tracking algorithm validation checks together with the corresponding unit tests solidified the idea that the "LeapFrogCurved" magnetic field tracking algorithm works as intended and was not the cause for the excessive "noise" in the simulated radio pulses. The next logical step was to inspect the main algorithm loop that "runs" the simulation, the *Cascade*. All the processes that are a part of the *process sequence* in the CORSIKA 8 universe are characterized as `ContinuousProcesses`, and as mentioned earlier all have the `doContinuous()` method to indicate that. There is a specific part of the code in *Cascade* that is responsible for applying all the `ContinuousProcesses` to all particles sequentially.

Initially, for the `ContinuousProcesses` a rather odd interface had been implemented in which every particle track was two separate objects, namely a `particle` and a `track`. What this means, is that a `particle` object can and does have information about time, position and energy among others but so does the `track` object. One has to be very careful especially when implementing a module that fits this interface to keep these two separate objects that refer to the same entity synchronized and up-to-date simultaneously. While this information is being calculated within the specified `ContinuousProcesses`, the update on the `particle` and `track` happens separately in the *Cascade* level. At the time of inspection (icrc-2021-b tag), it was realized that the position, direction and time information of the particle was updated through the `particle` and `track` before the application of `ContinuousProcesses`. Especially for the timing information which needed not only to access the `particle` time but also the `track` duration on which the `ContinuousProcess` took place, this was inconsistent. Updating the time information right after the application of `ContinuousProcesses` for each particle and not before, showed some dramatic changes. Most notably the maximally allowed magnetic deflection angle needed not be set to extremely low values anymore in order to produce plausible results and the runtimes decreased considerably.

Detailed comparison simulations between CORSIKA 8, CORSIKA 7 and ZHAireS were presented in [69]. In short, a much better agreement was observed especially when considering the dominant polarization (fig. 25). Unfortunately, when considering the other 2 polarizations shown in fig. 26 a bipolar structure in the radio pulses simulated with CORSIKA 8 was observed. This was not the case in CORSIKA 7 nor ZHAireS. This behavior was traced back to the rather odd interface choice were the `particle` and `track` were 2 separate objects with overlapping information. In more detail, the `particle` object would get updated through the `track` object in terms of energy, position, time and direction. This was not being updated correctly for the case of multiple scattering which is calculated through the PROPOSAL [70] interface for the `particle` object. In simpler terms, the direction of the particle would be calculated throughout

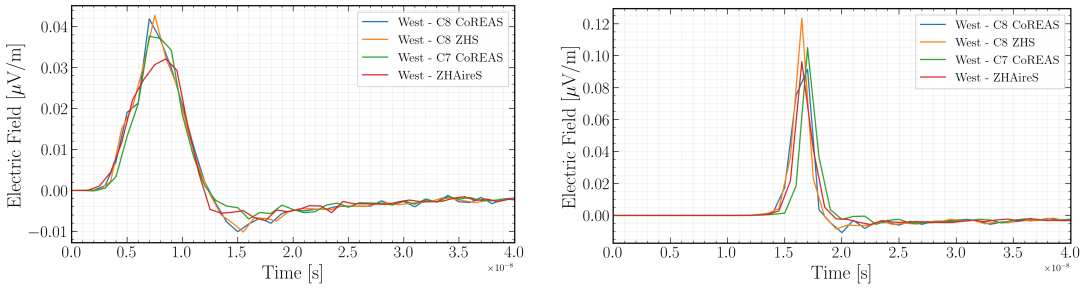


Figure 25: Signal pulse comparison of dominant polarization for antenna at 50 m (left) and at 200 m (right) from the shower core.

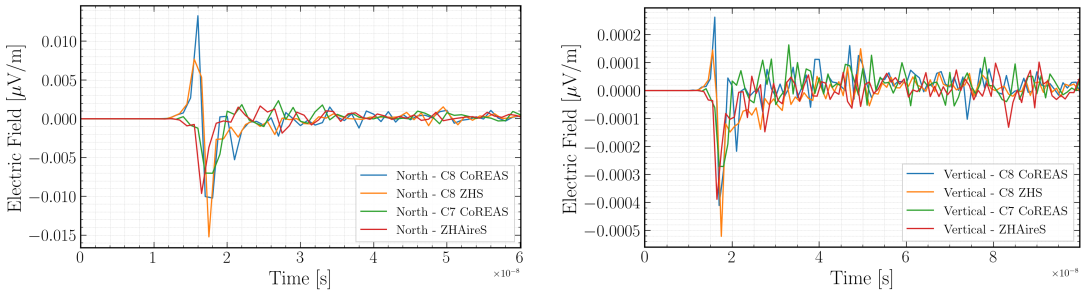


Figure 26: Signal pulse comparison for antenna at 200 m from the shower core. North (left) and vertical polarization (right) are plotted.

PROPOSAL but would then be overwritten by the `track` object in the *Cascade* algorithm.

To tackle this problem, it was decided to merge in a wrapper class the `particle` and `track` objects in order for the `ContinuousProcess` class to be able to process and return the differential change of state. This wrapper class is called `Step` and it creates and returns a `step` object which contains information about the differential change in energy, time, position, direction and velocity of a particle. Using this class the `ContinuousProcesses` get as input one `step` object instead of a `particle` and `track` and modifies the `step` accordingly. Then in the *Cascade* algorithm the `particle` object gets updated through the `step` object in which all the `ContinuousProcesses` have been applied. Most of the modules needed to be adapted to this interface change and especially the ones that were using the `doContinuous()` method. With this change though, multiple scattering is taken into account as intended this time.

An identical air shower simulation and comparison as presented in [69] was performed with the new interface. In fig. 27 it is shown that in the non-dominant polarizations the bipolar structure that was present in the CORSIKA 8 pulses (fig. 26) utilizing the old interface, is not present anymore and the agreement with CORSIKA 7 and ZHAireS is now within acceptable limits. The 2 dimensional maps of the energy deposited to the ground in the 30 MHz to 80 MHz band is depicted in fig. 28 in the same manner as in [69]. In [69] the fluence maps for $\vec{v} \times (\vec{v} \times \vec{B})$ and $\vec{v}$ polarizations were symmetric in contrast to the ones simulated with CORSIKA 7 and

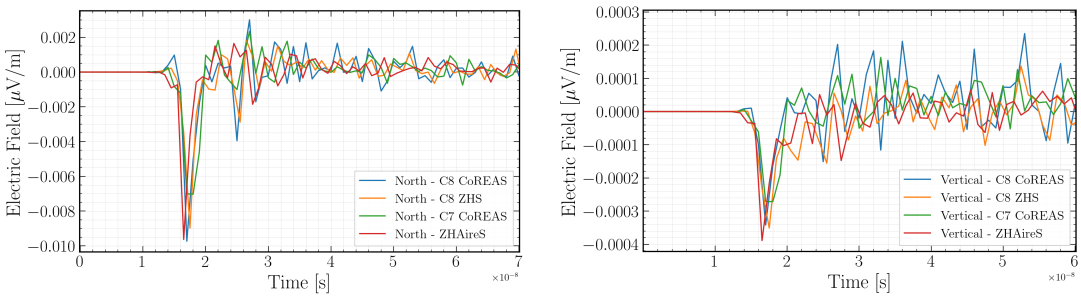


Figure 27: Signal pulse comparison for antenna at 200 m from the shower core. North (left) and vertical polarization (right) are plotted with the `step` object being used for the simulation.

ZHAireS. This is no longer the case and the overall agreement is within acceptable limits given the shower to shower fluctuations.

Summing up the work that has been presented so far in this chapter, some milestones are worth mentioning. First, it was proved that the radio module calculates the correct radio emission for both CoREAS and ZHS. The tracking algorithm was tested and found to have an expected behavior. The issues of noisy and bipolar pulses due to inconsistent handling of the time information were tracked down because of to the radio simulations. A solution for that was presented and major architectural changes in the CORSIKA 8 core code were made to tackle this. These changes were successful and the CORSIKA 8 radio simulations displayed good agreement with CORSIKA 7 and ZHAireS. This proves that the radio module can also be used as an excellent diagnostics tool and can contribute to improvements in the core code.

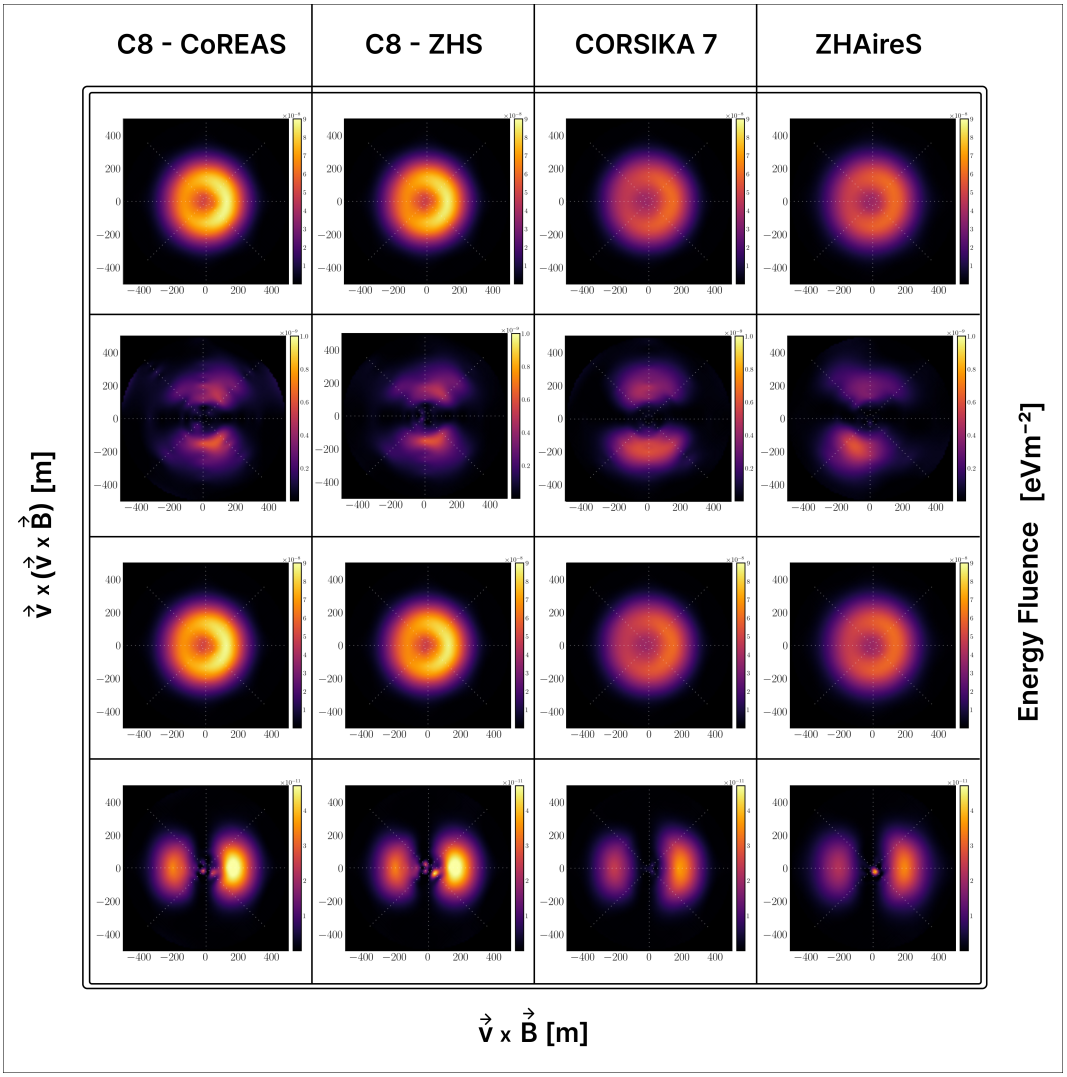


Figure 28: Table of energy fluence in different polarizations of the electric field in the 30 MHz to 80 MHz frequency band for C8 CoREAS, C8 ZHS, C7 CoREAS and ZHAireS. The order of the polarizations we see starting from top to bottom is: all polarizations, $\vec{v} \times (\vec{v} \times \vec{B})$, $\vec{v} \times \vec{B}$ and $\vec{v}$.

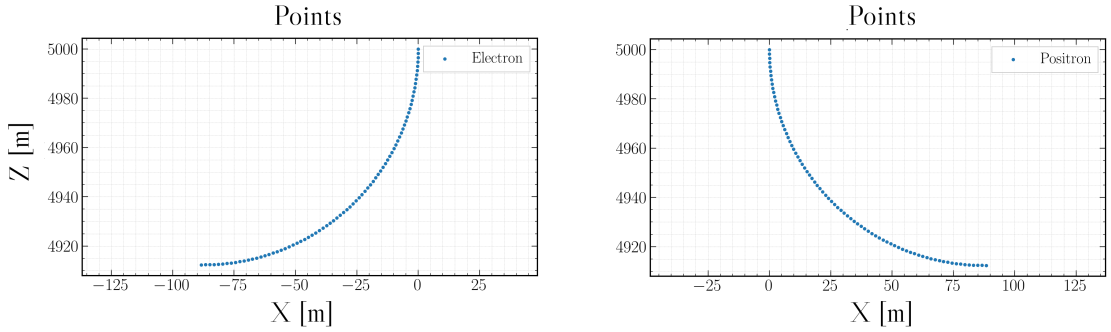


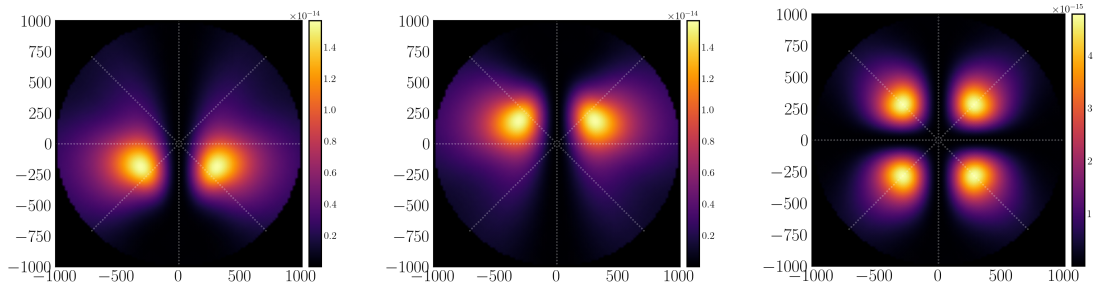
Figure 29: Points that consist the particle trajectory for the electron (left) and the positron (right).

6.4 The "Clover Leaf" pattern

Inclined air showers develop higher in the atmosphere, where the air density is lower. At these altitudes, the radiation pattern we witness on the ground from simulations is often called as the "clover leaf" pattern. There have been efforts for the current macroscopic descriptions of the radio emission to be refined [30]. This is the case, since the pattern under discussion cannot be described by the standard paradigm of a transverse current and a charge excess contribution. Such a polarization pattern (fig. 6) could be potentially explained by the geo-synchrotron models [31] which have predicted the "clover leaf" pattern in the past. Since the atmospheric density is low, no sufficient amount of charge excess can be built up on the shower front, and given a strong geomagnetic field the synchrotron emission can dominate over the charge excess. This can also be understood from the fact that the particles because of the lower density and the stronger magnetic field can propagate on appreciable arcs and are not stopped immediately by the interactions with the air molecules. Except for the "clover leaf" pattern, under the same circumstances the amount of radiation energy is significantly lower due to coherence loss [30].

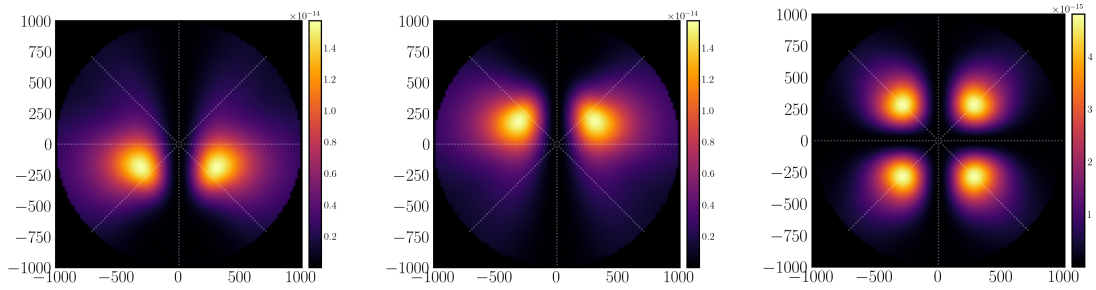
To test this, CORSIKA 8 and the radio module built in it provides an excellent opportunity given its modularity. An environment with a uniform horizontal magnetic field roughly ten times stronger than the Auger site and a refractive index equal to 1 is simulated. A pair of an electron and a positron is set to a downward vertical motion performing a quarter-circle with their energy set to 10 MeV. The injection altitude for both particles is 5 km. The trajectories of each particle under the influence of the magnetic field are shown in fig. 29. For the antenna array, an antenna is put to the center and around it 40 concentric rings spaced equally per 25 m with 8 antennas distributed azimuthally in each ring, covering a radial distance up to 1 km. For the radio emission both the CoREAS and ZHS formalisms were used.

The fluence contribution to the $\vec{v} \times (\vec{v} \times \vec{B})$ component, for an electron, a positron and an electron-positron pair is shown for CoREAS (fig. 30) and for ZHS (fig. 31). The emission resulting from an electron or a positron separately do not result in the clover leaf pattern, but



(a) Electron (b) Positron (c) Electron-Positron pair

Figure 30: The resulting fluence to the $\vec{v} \times (\vec{v} \times \vec{B})$ component for an electron (left), a positron (middle) and an electron-positron pair (right) as simulated with CoREAS. The radio pulses were filtered in the 10 MHz to 1 GHz frequency band.

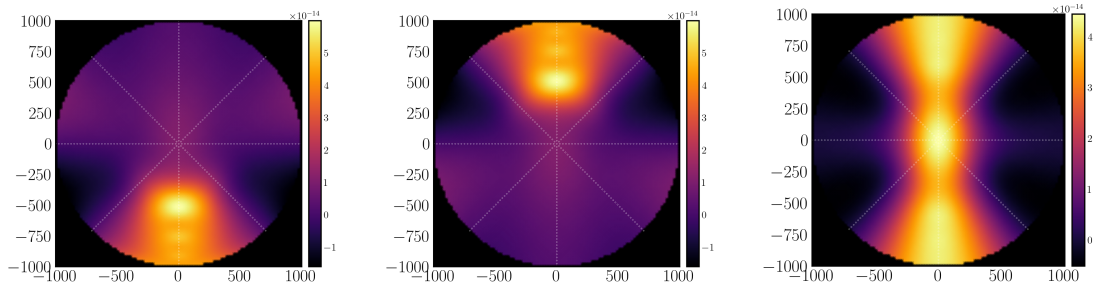


(a) Electron (b) Positron (c) Electron-Positron pair

Figure 31: The resulting fluence to the $\vec{v} \times (\vec{v} \times \vec{B})$ component for an electron (left), a positron (middle) and an electron-positron pair (right) as simulated with ZHS. The radio pulses were filtered in the 10 MHz to 1 GHz frequency band.

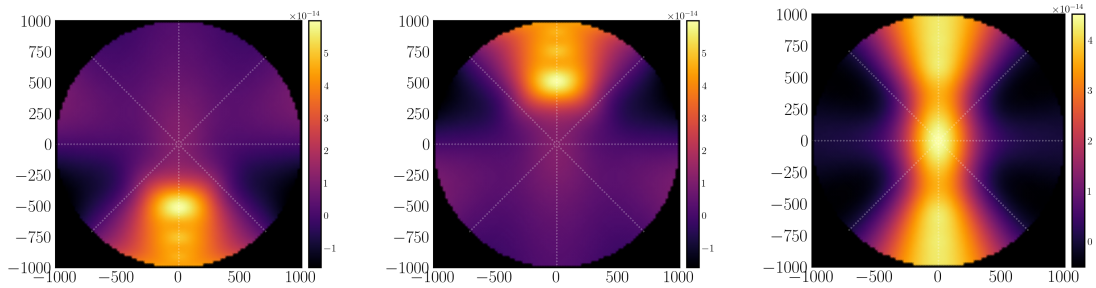
the superposition of both their emissions does. For the pair, destructive interference is observed on the North-South and East-West axis and the pattern shows close resemblance to fig. 6. This signature pattern is a strong indication that clover leaf patterns observed in air shower simulations could be explained by the synchrotron emission model [30]. Looking into the dominant $\vec{v} \times \vec{B}$ component for the CoREAS (fig. 32) and ZHS (fig. 33) formalisms, it is evident that the pair results from constructive interference along the North-South axis with the fluence peaking at the center.

On a more technical note, this lab experiment showcases the capabilities of the module and of CORSIKA 8. Simulations like this can be used as a tool to inspect the dependence of the clover leaf pattern on many parameters such as density or magnetic field strength in a very straightforward way. This is an easy to run and tweak example code that can be found on the CORSIKA 8 physics examples. Additionally, it is shown that the CoREAS and ZHS formalisms do agree quite well and both predict the same patterns. It is also worth noting, that to properly simulate the fluence patterns discussed above, the maximally allowed deflection angle in the



(a) Electron (b) Positron (c) Electron-Positron pair

Figure 32: The resulting fluence to the $\vec{v} \times \vec{B}$ component for an electron (left), a positron (middle) and an electron-positron pair (right) as simulated with CoREAS. The radio pulses were filtered in the 10 MHz to 1 GHz frequency band.



(a) Electron (b) Positron (c) Electron-Positron pair

Figure 33: The resulting fluence to the $\vec{v} \times \vec{B}$ component for an electron (left), a positron (middle) and an electron-positron pair (right) as simulated with ZHS. The radio pulses were filtered in the 10 MHz to 1 GHz frequency band.

tracking algorithm had to be set to a very small value (0.0001 rad). This effectively simulates the particle trajectory as a set of many, very small, fine tracks. Using the tracking algorithm's default value which is 0.2 rad did not simulate enough points to calculate the synchrotron radiation properly. It produces blurry patterns which indicate the lack of proper resolution. As a final remark, the CoREAS formalism was able to reach the desired resolution with a larger value of 0.001 rad, whereas the ZHS formalism would produce blurry patterns at that value. This hints to dependency of the agreement between the 2 algorithms on the effective simulated track length, which is the topic of discussion of section 9.

7 Parallel calculation of radio signals using multithreading

Computing performance is a major issue when it comes to simulating and processing air showers from ultra-high-energy cosmic particles with CORSIKA 8. This is the case for CORSIKA 7 and ZHAireS too, where these simulation codes have been hand-optimized to tackle this with the downside of losing flexibility. Hurting flexibility is not an option for CORSIKA 8. Computing times are very long when it comes to the radio calculation, especially when one needs to simulate hundreds or even thousands of antenna instances. In the modern era of computers, multicore systems with an ever increasing core count are available for the mainstream audience. Developing CORSIKA 8 from scratch using the latest C++ standards allows for the inclusion of technologies like multithreading. Instead of performing calculations in a sequential manner, available threads on the system handle the operations on a parallel fashion which can dramatically increase performance of a single job. The job runtime boost, or speed up factor (S) is dictated by Amdahl's law [71] which is shown in eq. (27) below.

$$S = \frac{1}{f + \frac{1-f}{P}} \tag{27}$$

Where f is the sequential workload and P is the number of processors, in our case threads. This formula states that even with an infinite number of threads there will always be a threshold $\frac{1}{f}$ that job runtime cannot be boosted beyond. In simple terms, employing raw computing power in terms of threads (or cores) will not make the code run infinitely faster; optimizations and efficient code writing is still crucial. Regarding CORSIKA 8, the ideal scenario of parallelization would be to parallelize all the modules that take place in the *process sequence* and also the particle stack. This way particles would be assigned and treated in parallel by different cores and the performance jump would be significant. Such a task is not at all trivial as it entails both technical and conceptual challenges. On a more technical level, the code needs to apply the concept of thread safety which ensures that it behaves as intended and operations do not go through unintended interactions. This is up to the developer to figure out and eventually choose wisely the levels of parallelism employed.

The radio module is the first module in CORSIKA 8 to make use of multithreading technologies. The reasons that radio was chosen for this were that an "inner loop" consumes a large fraction of computing time and to parallelize over antennas is conceptually easy. In the following section the library used and the implementation of it in the radio code is going to be discussed. Benchmarks with idealized scenarios and air shower simulations will be shown that confirm the benefits of multithreading. The numerical consistency of the implementation will also be validated by simulating identical showers many times on a different number of threads.

7.1 Gyges and adjustment to the radio module

The most obvious and probably one of the most efficient way to parallelize the radio calculation is on the antenna level. As discussed in section 5 every particle track is pushed into the formalism where it is looped over all antennas available in the `AntennaCollection`. At this level, concepts of C++ concurrency are employed in order to create a suitable thread pool that allocates the computational load for each antenna. The Gyges library [72] that handles multithreading concepts has been used by the radio module.

Gyges is a lightweight C++17/20 header-only library for easing the creation, task submission and management of thread pools on multicore systems. By deploying Gyges, the user avoids the implementation details involved in the creation, synchronization and destruction of thread pools, as well as the submission of tasks to it. For instance, by using Gyges, the computational costs associated to the creation and destruction of threads can be paid just once in the program lifetime, since threads of the pool pick-up tasks as they become available. If there are no tasks to be processed, the threads just go sleeping, releasing resources for other processes that might be running on the machine. A simple interface is available for the task submission to multiple threads. The submitter gets a `std::future` for monitoring and collecting the result of the task in-place. Additionally, the assignment of submitted tasks to threads, and the processing of assigned and running ones, can be "politely interrupted" at anytime acting over a `std::stop_token` instance passed to the task from the Gyges interface.

For certain workloads, it is convenient to submit all tasks to the pool before starting the processing. Gyges covers such cases as well. Indeed, it is possible to create a thread pool and put in a “hold-on” state, which postpones the processing of the tasks until the thread pool is put back on “unhold” status. In Gyges, the threads are never kept busy-waiting, i.e., the whole pool is put to sleep until an “unhold” command is sent. These features make Gyges very suitable for workloads consisting of many tasks with different times of completion. Given the tasks are picked up by threads of the pool in an out-of-order way, the overall time required to process a bunch of tasks tends to be limited by the task taking the longest to complete. Resource starvation at the pool is automatically prevented, once threads that cannot find available tasks are put to sleep.

The parallelism offered by Gyges is exploited to accelerate the processing of the radiation emitted by each particle, parallelizing the calculations of the antenna array. For each formalism, a new callable `runner` object was implemented which abstracts away the implementation details for the actual radio calculation, and is convenient for the Gyges library. This object together with the `AntennaCollection` are distributed to the worker threads as indicated by the `gyges::gang` instance. More specifically, a task is defined for each antenna in the array detector. The set of tasks is distributed to the thread pool managed by the `gyges::gang` and the corresponding calculations are performed in parallel. The size of the thread pool is specifiable at the *radio process* construction time and should be chosen in order to maximize performance, hiding the latency inherently associated with the computational costs associated to keep a large number of threads active, even if for a short period of time. To hide all this complexity from users simple

and comprehensive interfaces have been implemented as shown in listing 7.

Listing 7: The radio module interface

```
// the detector (aka antenna collection) for CoREAS and ZHS
AntennaCollection<TimeDomainAntenna> detector;

//helper function for creating a propagator which takes the environment as a
    parameter
auto propagator = make_dummy_test_radio_propagator(environment);

//helper functions for creating CoREAS and ZHS instances, taking as parameters
    the detector and propagator object, along with the number of threads
auto coreas = make_radio_process_CoREAS(detector, propagator, nthreads);
auto    zhs = make_radio_process_ZHS(detector, propagator, nthreads);
//the coreas and zhs objects are radio processes which are put in the process
    sequence
```

Once an `AntennaCollection` (or detector) has been initialized, one is free to construct and add there the specified type of antenna instances. Using the helper functions for the propagators, a `propagator` object is returned which along with the `AntennaCollection` and the user designated *number of threads* act as variables in the formalism helper functions. Each formalism has implemented a helper function that returns a *radio process* type object which is the final object that needs to be put in the *process sequence* of the CORSIKA 8 simulation. By default, the number of threads for the formalism helper functions is set to `std::thread::hardware_concurrency` which is the optimal number depending on the hardware at hand. It is worth noting that more threads do not always imply better performance. There is a latency cost to be paid for thread synchronization which makes multithreading suitable for heavier loads. These aspects along with performance studies are going to be discussed next.

7.2 Performance benchmarks

In this section, we aim to benchmark and validate the multithreading capabilities of the radio module. For this task, a modern CPU with 24 physical cores and 48 available threads has been employed. The runtimes are documented and the overall speed-up factor is investigated as a function of the number of threads. The 2 main observations are that the speed-up is heavily dependent on the computational load that is assigned to each thread and the number of antennas. The more the load for each cycle the better the speed-up factor. Additionally, the more antenna instances within the `AntennaCollection` the better the performance boost. These observations are confirmed by the following tests.

For the first test, a single charged particle travelling on a finite straight trajectory is considered. For the calculation of the radio emission from that track both formalisms are used. A `gyges::gang` with up to 48 worker threads is employed. In fig. 34, the benchmark plots are shown for CoREAS (left) and ZHS (right) on an antenna collection consisting of 200 antennas. The highest speed-up is observed for 12 threads in the case of CoREAS and for 8 threads for

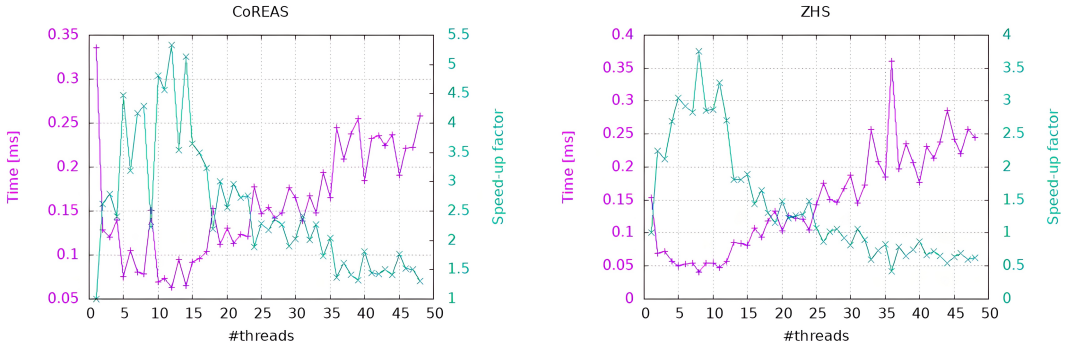


Figure 34: Performance to process a single particle as a function of number of threads for an array detector containing 200 antennas.

ZHS. For more threads, the performance is saturated and in the case of ZHS we even observe the multithreaded performance getting worse for more than 26 threads compared to the single core performance. This happens due to computing tasks not being able to occupy the CPU enough to hide the latency and costs associated with the management of multiple threads. Additionally, the 48 threads correspond to logical cores due to hyper-threading, but the physical cores of the system are 24. Hence, beyond 24 cores no boost in computation times is expected. Next, in fig. 35 a larger antenna collection consisting of 1000 antennas is inspected. In this case, peak performance is observed for 16 threads and 10 threads for CoREAS and ZHS respectively. For a larger number of threads the performance is saturated again but this time it is always better than the single core performance. Finally, in fig. 36 we perform the same test for 10000 antennas where we see a consistent increase in performance as the number of threads is increased. In this test case, the performance dependency on the number of antennas is investigated and it is shown that the multithreading capabilities of the module truly shine on detectors that consist of antennas in the order of thousands. Such a performance boost gives an advantage and a viable technical solution for simulating air showers for experiments like SKA and GRAND that utilize thousands of antennas.

For the next test case the synchrotron emission similar to the calculation described in section 6.1 is considered. Here the performance dependency on the number of antennas but also on the number of particle tracks is inspected. In fig. 37 and fig. 38 it is confirmed that the performance boost is not correlated with the number of simulated particle tracks. Performing the radio calculation on a larger antenna collection though, clearly shows the benefits of multithreading. This is evident when looking at how the runtimes behave on different number of threads. In fig. 37 the runtimes decrease for up to 5 threads but quickly saturate and then the task assignment and thread management hurt performance due to the small number of antennas. In fig. 38 where instead of 160 antenna instances, 1600 are utilized for the same calculation the runtimes considerably drop and saturate at roughly 10 threads for both formalisms. In all benchmarks cases inspected so far the behavior is expected and follows Amdahl's law. Similar results would have been achieved, albeit leading to performance peaking at different number of threads, when

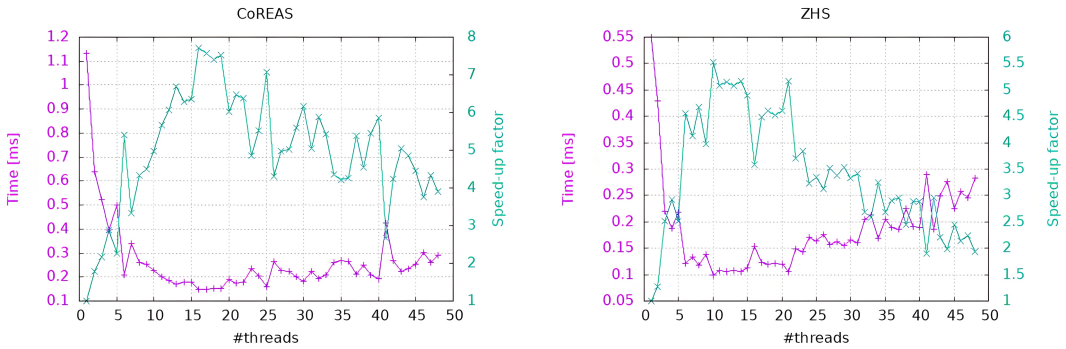


Figure 35: Performance to process a single particle as a function of number of threads for an array detector containing 1000 antennas.

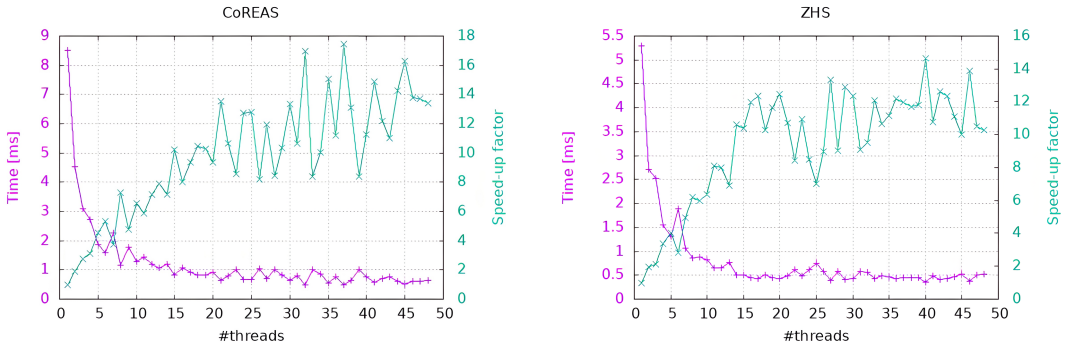


Figure 36: Performance to process a single particle as a function of number of threads for an array detector containing 10,000 antennas.

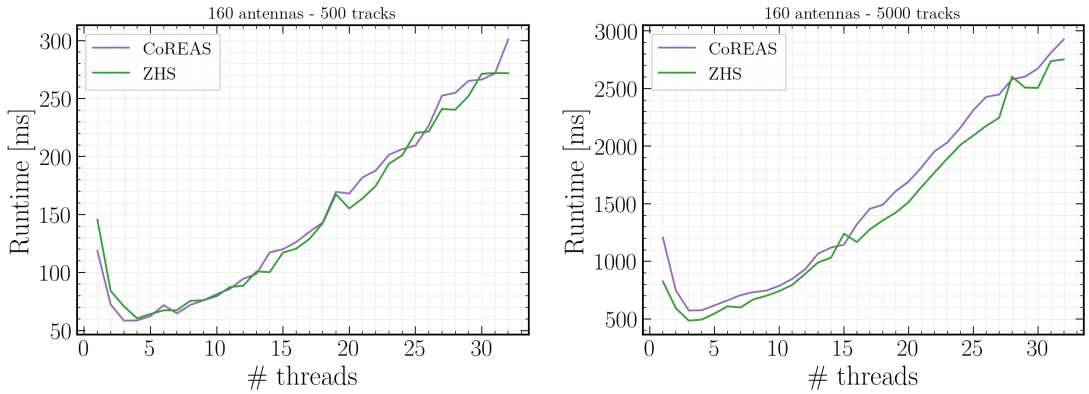


Figure 37: Performance for different number of threads calculated for the synchrotron emission case for 160 antennas. The circular trajectory consists of 500 tracks (left) and of 5000 tracks (right).

deploying propagators performing heavier calculations.

Performing these benchmarks especially the ones in figs. 34 to 36 provides the opportunity to compare the runtimes between the CoREAS and ZHS formalisms. Given an identical scenario, the one of processing a single particle track on different number of antennas in an environment with a uniform refractive index, ZHS is almost twice as fast as CoREAS. This is to be expected though, because as already mentioned ZHS treats every particle track as one point internally whereas CoREAS for every particle track considers the start and end point, hence two points. Furthermore, CoREAS needs to call the propagator twice while ZHS calls it only once. As a result, CoREAS gains more performance boost by multithreading as can be seen by the larger speed-up factors. Interestingly, looking at figs. 37 and 38 one can realize that for 500 tracks ZHS is actually slower than CoREAS and for 5000 tracks is again faster regardless the number of antennas. In this case, since the circular trajectory is fixed in size, when considering 500 tracks, the tracks are larger. Here, ZHS needs to subdivide each track to smaller ones to fulfill the Fraunhofer limit, which makes the algorithm process more track segments. For 5000 tracks though, the tracks are smaller and the subdivision does not need to happen anymore and hence CoREAS is slower as explained earlier. This observation highlights once again the capabilities of the radio module which is able to handle both formalisms on the exact same footing and as a result, reveal the intricacies of both the CoREAS and ZHS formalisms.

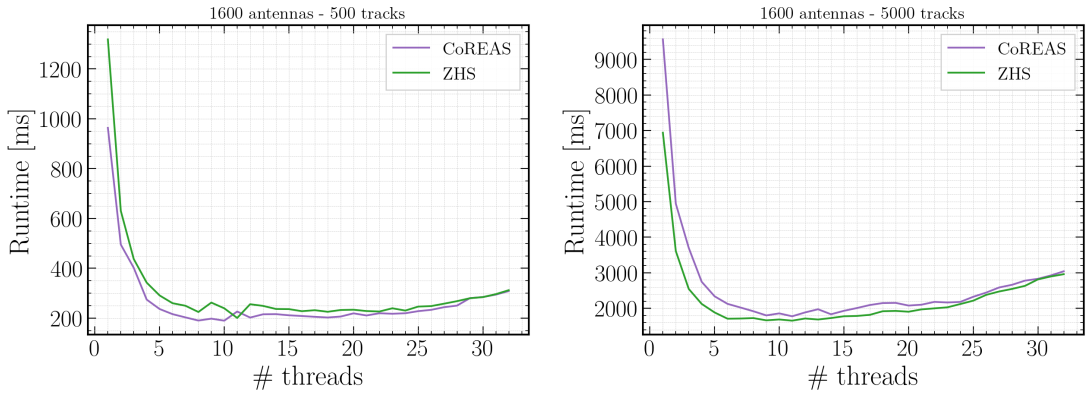


Figure 38: Performance for different number of threads calculated for the synchrotron emission case for 1600 antennas. The circular trajectory consists of 500 tracks (left) and of 5000 tracks (right).

7.3 Air Shower validation

It has been confirmed that by employing `Gyges` and adjusting the radio module accordingly, multithreading works as intended and follows Amdahl's law. In this section, the overall impact of parallelization on an air shower simulation is examined. The *radio process* is the only one in the *process sequence* that employs multithreading which means that the rest of the processes run sequentially and the speed-up is limited on the radio level. For this reason, the simulation is restricted to an electron-induced air shower with hadronic interaction models being omitted. It is worth noting that in a full shower simulation due to the Gyges design, the overhead for creating, managing and submitting tasks to the thread pool is negligible in comparison to the other initialization routines called up-front by all the modules relevant to the simulation.

The air shower is set to be vertical with primary energy for the electron being 10 TeV at a constant horizontal geomagnetic field of 50 µT aligned in the x direction and a uniform refractive index ($n = 1.000327$). The antenna array is configured in 8 concentric rings distributed azimuthally with a 25 m spacing totalling 160 antennas per *radio process*. 2 *radio processes* with different formalisms, but identical propagators, antenna collections and filters are initialized. Each of the *radio processes* is submitted to `Gyges` separately and although in total CORSIKA 8 handles 320 antennas, `Gyges` assigns threads to 160 antennas at a time.

The electromagnetic shower configuration is benchmarked and the results are shown in fig. 39. First, the electromagnetic shower with the radio module being turned off is simulated in order to highlight the computational cost of the module itself. Then the same electromagnetic shower with identical setup and identical number sequence is simulated for a varied number of threads, starting from 1 thread all the way up to 32. The performance profile shown in fig. 39 confirms the behavior observed in the previous performance plots. The peak speed-up factor is observed at 10 threads; for more threads the CPU load is not high enough to conceal latency

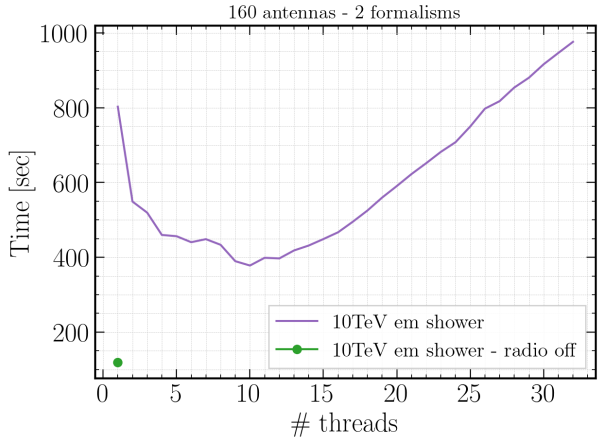
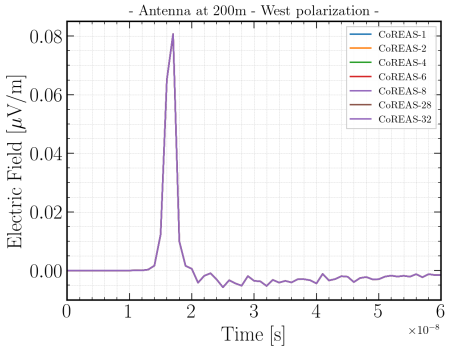


Figure 39: The parallelized radio module running on an electron induced air shower processing a detector array of 160 antennas. 2 formalism, namely CoREAS and ZHS are activated and use 160 antennas each. The performance peaks at 10 worker threads, beyond which the performance degradates.

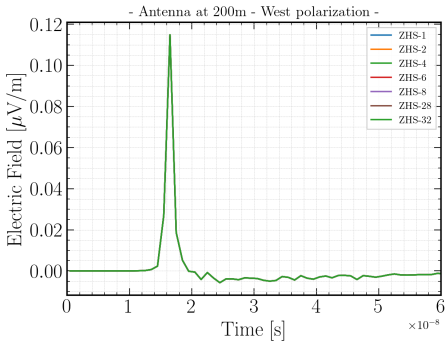
costs and thread management, as expected. In single-core simulations the computing cost of the radio module is also evident leading to a $\simeq 8$ times increase in computing time. This highlights once again the necessity of the module being able to reduce its computing times. In this case, this is achieved to roughly a half by taking advantage of multithreading.

Finally, the numerical consistence of the radio pulses simulated using both formalisms is validated for a different number of threads. In fig. 40 the pulses simulated with CoREAS and ZHS respectively are plotted. The same calculation has been performed by a varying number of threads multiple times. A random antenna at 200 m from the shower on the arm at 90° is picked and all the polarizations are plotted separately. It is evident in the plots that regardless the number of threads employed the resulting pulses are identical with no measurable numerical impact. This proves that the radio calculation employing different number of threads is carried out consistently and accurately.

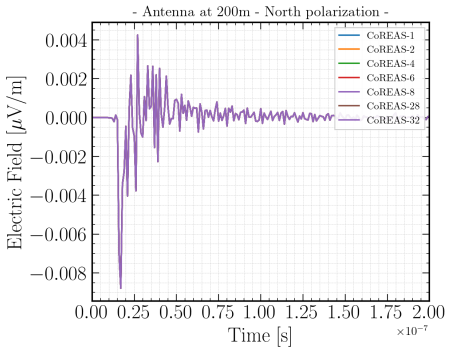
As a final remark, the successful parallelization of the radio module acts as a proof principle that the module is thread safe and written in a modern way that allows the inclusion of modern technologies. Ideally, the parallelization strategy followed can be extended by the rest of CORSIKA 8 developers on the other modules and significantly improve runtimes. Given the modularity and flexibility associated with the CORSIKA 8 code a penalty in performance is paid. The way to overcome this, is parallelization not only in the CPU level but also by employing GPU cores. Modern GPUs come with thousands of cores which can help to accelerate radio emission calculations even further and thus CORSIKA 8 as a whole.



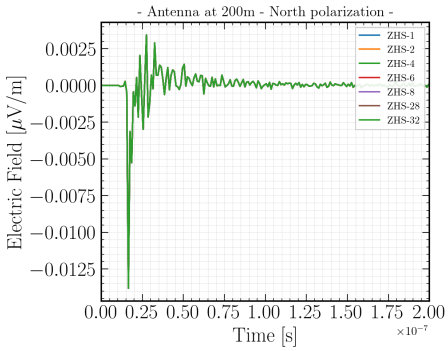
(a) West polarization - CoREAS.



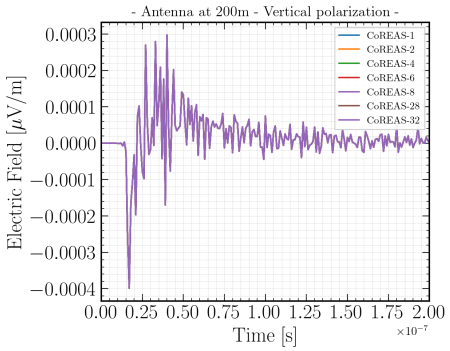
(b) West polarization - ZHS.



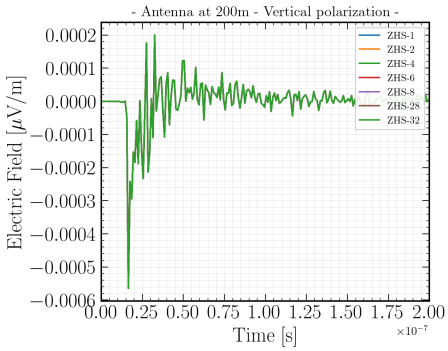
(c) North polarization - CoREAS.



(d) North polarization - ZHS.



(e) Vertical polarization - CoREAS.



(f) Vertical polarization - ZHS.

Figure 40: Pulse comparisons in all three polarizations using the CoREAS and ZHS formalisms. The pulses have been simulated and processed on `gyges::gangs` of different sizes. Different number of threads produce identical pulses for both formalisms, as expected.

8 Air Shower validation with existing software

Having presented the radio module's architecture, capabilities and various use cases it is only natural to showcase its full potential on a realistic air shower simulation. To validate its results, radio simulation comparisons take place with well established software like CORSIKA 7 and ZHAireS for similar air showers. In addition, an extensive comparison with a library of simulations takes place between CORSIKA 8 and CORSIKA 7 where it is showcased that the effective track length simulated for particle tracks can alter their agreement.

8.1 A CORSIKA 8 versus CORSIKA 7 versus ZHAireS comparison

Many extensive air showers were simulated with CORSIKA 8 using both the CoREAS and ZHS formalisms via two separate *radio processes*, as well as with CORSIKA 7 and ZHAireS. The settings used for all shower codes were set to be identical wherever possible. It is also worth mentioning that for the CORSIKA 8 simulations the `beta_release_pre_v1.1` tag was used. 3 of those showers in which the longitudinal development of the shower was very similar were chosen and compared in terms of their radio emission below.

The setting for the simulation for all software was a vertical iron induced 100 PeV extensive air shower. Energy cuts were set for both electromagnetic and hadronic components of the shower. For the electromagnetic cuts the value was set to 0.5 MeV, and for the hadrons and muons the cuts were set to 0.3 GeV. Regarding hadronic interaction models, Sibyll 2.3d [14] was used for high energies and for the low energy particles for C8 and C7 FLUKA [18] was used and Sibyll 2.3d [14] for ZHAireS. Concerning the electromagnetic interaction model, in contrast to CORSIKA 7 and ZHAireS, for CORSIKA 8, PROPOSAL v7.6.1 [73, 70] was used. In addition, thinning was employed to reduce computing times. All 3 shower codes use different thinning algorithms so the parameters were set in a way to produce comparable effects. For CORSIKA 8, the thinning algorithm is applied to the electromagnetic cascade only [74]. One more thing that is not the same in the shower codes is the tracking algorithms. For CORSIKA 8 and CORSIKA 7 the default value of the maximally allowed deflection angle used is 0.2 rad, while for ZHAireS the default value is there too. Taking all these technical differences into account together with the fact that Monte Carlo simulations rely on random number sequences, it is unrealistic to expect a 100% agreement.

The environment was configured in all simulations to be the "US Standard Atmosphere", with a refractive index at sea level of $n = 1.000\,327$ and for higher altitudes scaling following the Gladstone-Dale law [75], i.e, refractivity scaling proportionally with air density. A constant geomagnetic field of 50 µT aligned in the x direction was set. A star-shaped pattern of 160 time domain antennas for the antenna collection was set. The sampling period of each antenna is set to 0.1 ns. They are located at sea level, in 20 concentric rings spaced equally from 25 m to 500 m from the shower axis with 8 antennas distributed azimuthally in each ring. The precise locations

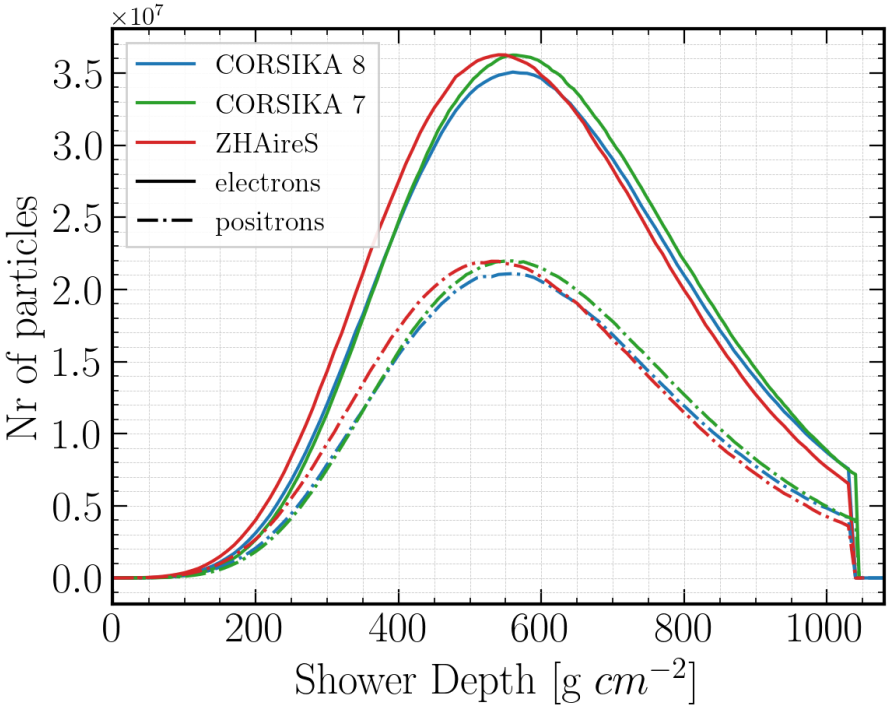
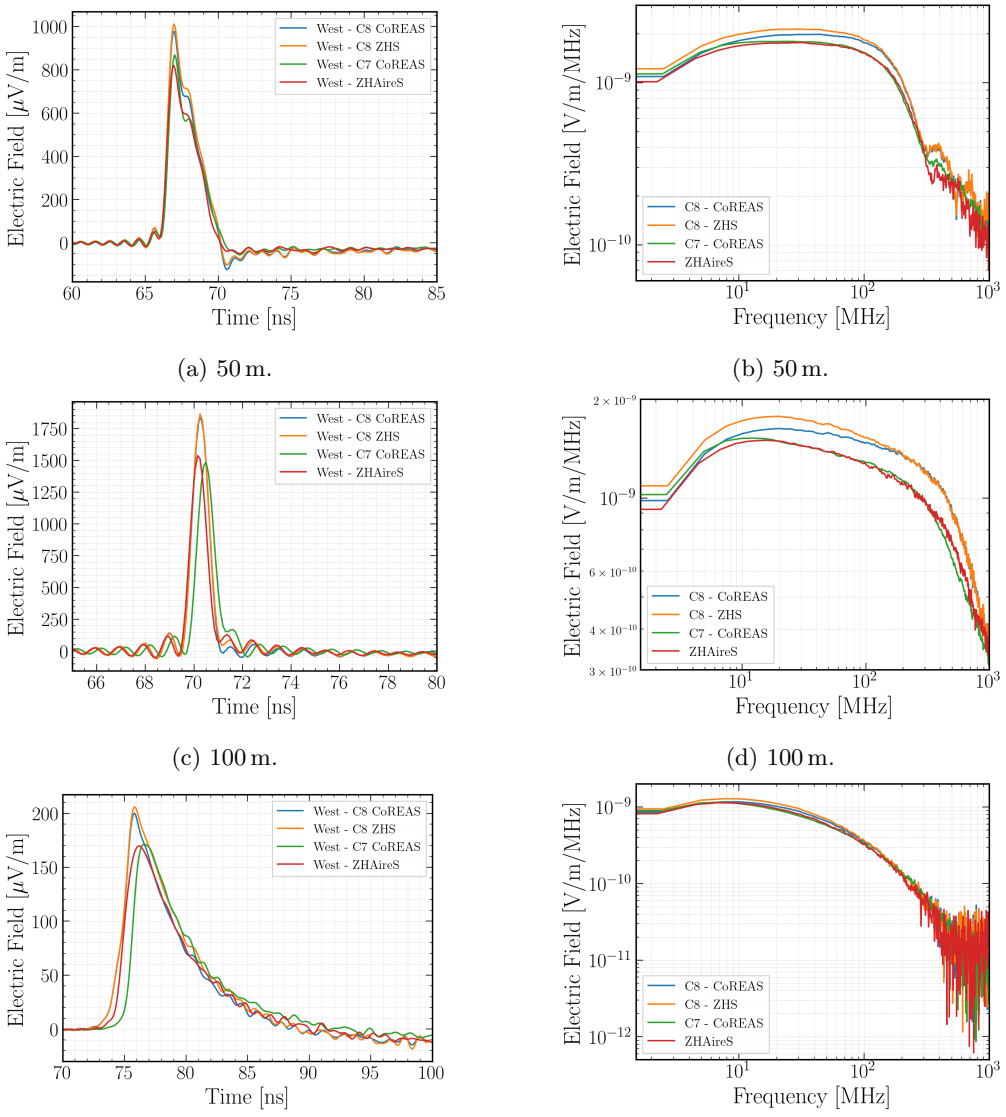


Figure 41: Longitudinal profiles of air showers simulated with CORSIKA 8, CORSIKA 7 and ZHAireS. The number of electrons and positrons with respect to grammage is shown.

can be found in the 2 dimensional fluence maps in fig. 44 and fig. 45. It is worth pointing out once more, that for C8 we run one shower simulation which consisted of 2 *radio processes*. Both used the same filters, propagator, identical antenna collections and differ only in the formalism, one used CoREAS and the other one ZHS.

In fig. 41 the longitudinal profiles of all 3 showers simulated with CORSIKA 8, CORSIKA 7 and ZHAireS are shown. The number of electrons and positrons are plotted with respect to atmospheric depth in $\mathrm{g\,cm^{-2}}$. For the CORSIKA 8 shower, X_{max} is $\sim 560\,\mathrm{g\,cm^{-2}}$, for ZHAireS it is $\sim 550\,\mathrm{g\,cm^{-2}}$ and finally for CORSIKA 7 it is $\sim 570\,\mathrm{g\,cm^{-2}}$. In terms of number of particles, from fig. 41 it is evident that for both electrons and positrons all 3 showers are very close. Hence, given that both number of particles and X_{max} are similar, it is safe to proceed and compare the resulting radio emission.

To directly compare the radio emission from all air showers an arm is chosen in the $\vec{v}\times(\vec{v}\times\vec{B})$ direction, at 90° from the positive x-axis on the star-shaped pattern. This choice, given that a purely vertical shower is examined decouples the geomagnetic emission from the charge excess contribution by comparing the different polarizations on the antenna level. Antennas on 3 distances from the shower core are picked for the detailed comparisons, namely one closer to the shower core at 50 m, one at the Cherenkov ring where the pulses peak in amplitude at 100 m and



(a) 50 m. (b) 50 m.

(c) 100 m. (d) 100 m.

(e) 200 m. (f) 200 m.

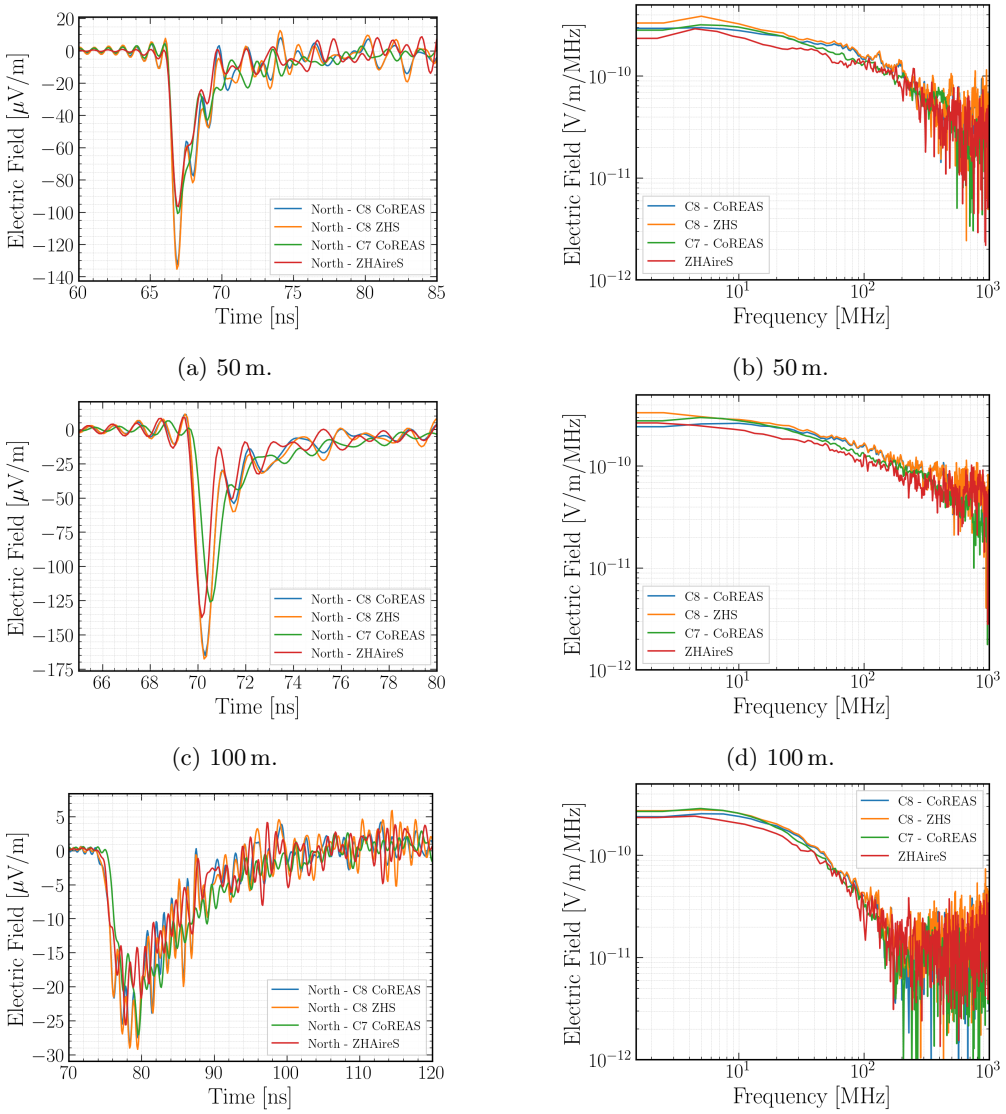
Figure 42: Signal pulse and frequency spectra comparison in the 0 GHz to 1 GHz band, for various antenna distances from the shower core - Geomagnetic contribution.

one further away at 200 m where the signals start to die out. I inspect both the time domain and frequency domain information on all these locations. For the time domain the pulses are filtered in the 0 GHz to 1 GHz frequency band.

In fig. 42, the comparison starts from the geomagnetic contribution which is dominant in air showers. This is shown in terms of antenna polarization which is $\vec{v} \times \vec{B}$ or also referred as "West" as dictated by the geomagnetic field vector orientation. Close to the shower core at 50 m (fig. 42a and fig. 42b) the agreement for the pulse amplitudes is quite good within 20% with CORSIKA 8 pulses producing slightly more power in the whole spectrum. Near the Cherenkov ring at 100 m (fig. 42c and fig. 42d) in terms of pulse amplitude the CORSIKA 8 pulses are stronger by $\leq 15\%$ and this also the case for the corresponding frequency spectra. Finally, for an antenna at 200 m only for the pulse amplitude the CORSIKA 8 pulses are stronger by $\leq 15\%$ while in terms of spectral amplitude the difference is much smaller.

The subdominant charge excess contribution is showcased in fig. 43 for the same antennas as previously. This contribution is studied on the $\vec{v} \times (\vec{v} \times \vec{B})$ or "North" polarization. The agreement in this case is similar to the one found in the geomagnetic contribution both in the time and frequency domain. The most prominent difference of roughly 25% is spotted closer to the shower core, where the CORSIKA 8 signals are again stronger. As I compare the pulses and spectra moving away from the shower core the agreement gets better. In both the geomagnetic and charge excess contributions apart from the amplitude and power differences that were discussed, the radio signals in the time domain agree also in terms of pulse shape. Moreover, a slight inconsistency in the timing information is observed for all 3 showers. This is not a CORSIKA 8 issue as we can see that CORSIKA 7 times are off even when compared with ZHAireS. The reason for this is unclear at the time of writing.

As a final comparison, I calculate and plot in two-dimensional maps the energy fluence of the 3 showers in all polarizations as defined in eq. (11). For the comparison, two frequency bands are considered, namely the 30 MHz to 80 MHz (fig. 44) and 50 MHz to 350 MHz (fig. 45). In both fluence maps the exact antenna locations are shown where the signal has been simulated, whereas in the in between regions the signal has been interpolated using the radial basis function interpolation through `scipy` [67]. In both frequency bands and all polarizations the symmetry and shape of the footprint match very well between the different codes and formalisms. One can also notice that the 2 formalisms match better in the 50 MHz to 350 MHz band as can be seen in fig. 45. However, there is an excess of fluence for the CORSIKA 8 pulses of roughly 20 % to 25 % especially close to the shower core. A similar behavior was observed for the pulses in the time domain where the frequency band considered was 0 GHz to 1 GHz though. It is worth noting once more that the simulated showers are not identical so a 100% agreement is not expected. Systematic studies that have been performed on electromagnetic air showers regarding the agreement of CORSIKA 7 and CORSIKA 8, might indicate why even though the CORSIKA 8 shower has less electrons and positrons as seen in the longitudinal profiles (fig. 41) more energy is being deposited on the ground in the form of radio waves. Looking at the bottom of fig. 47, the ratio of the lateral profiles between CORSIKA 8 and CORSIKA 7 is plotted. The straight



(a) 50 m. (b) 50 m.

(c) 100 m. (d) 100 m.

(e) 200 m. (f) 200 m.

Figure 43: Signal pulse and frequency spectra comparison in the 0 GHz to 1 GHz band, for various antenna distances from the shower core - Charge excess contribution.

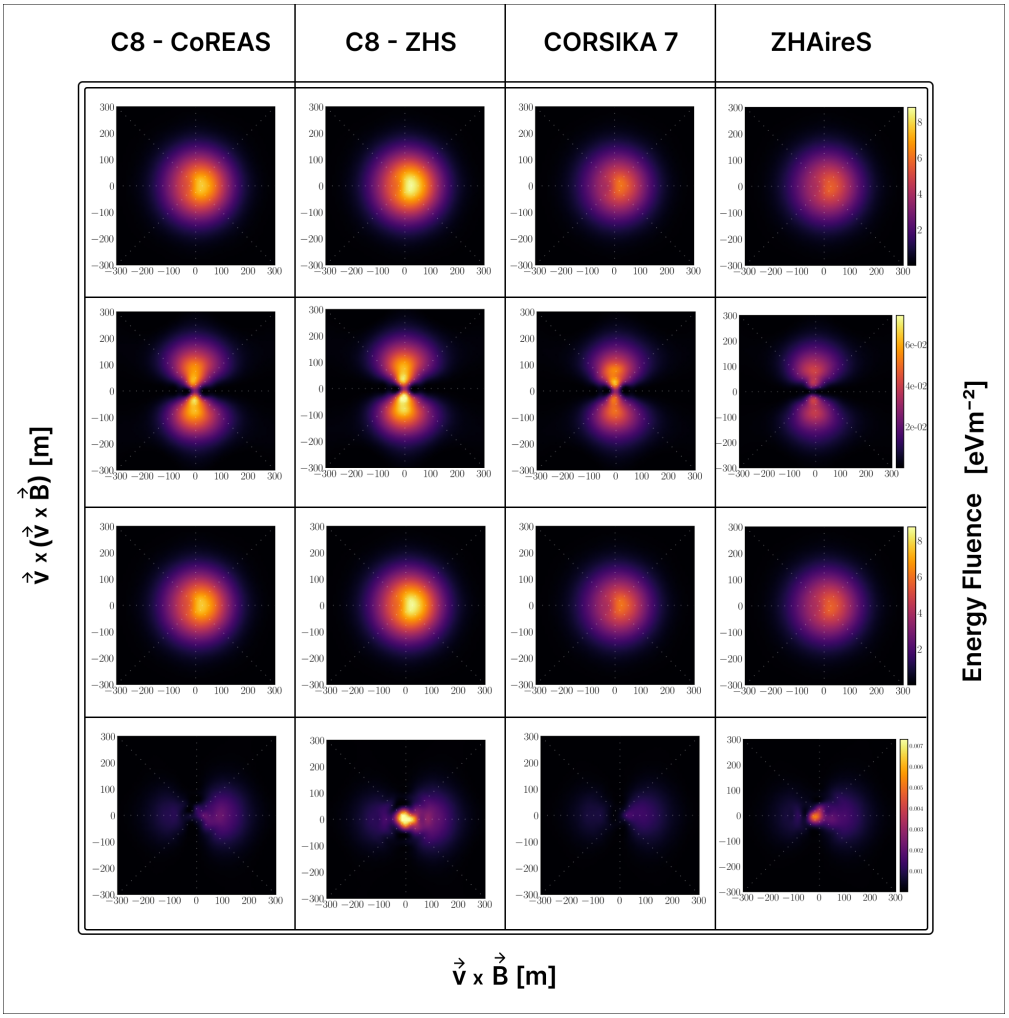


Figure 44: Table of energy fluence in different polarizations of the electric field in the 30 MHz to 80 MHz frequency band for C8 CoREAS, C8 ZHS, C7 CoREAS and ZHAireS. The order of the polarizations we see starting from top to bottom is: all polarizations, $\vec{v} \times (\vec{v} \times \vec{B})$, $\vec{v} \times \vec{B}$ and $\vec{v}$.

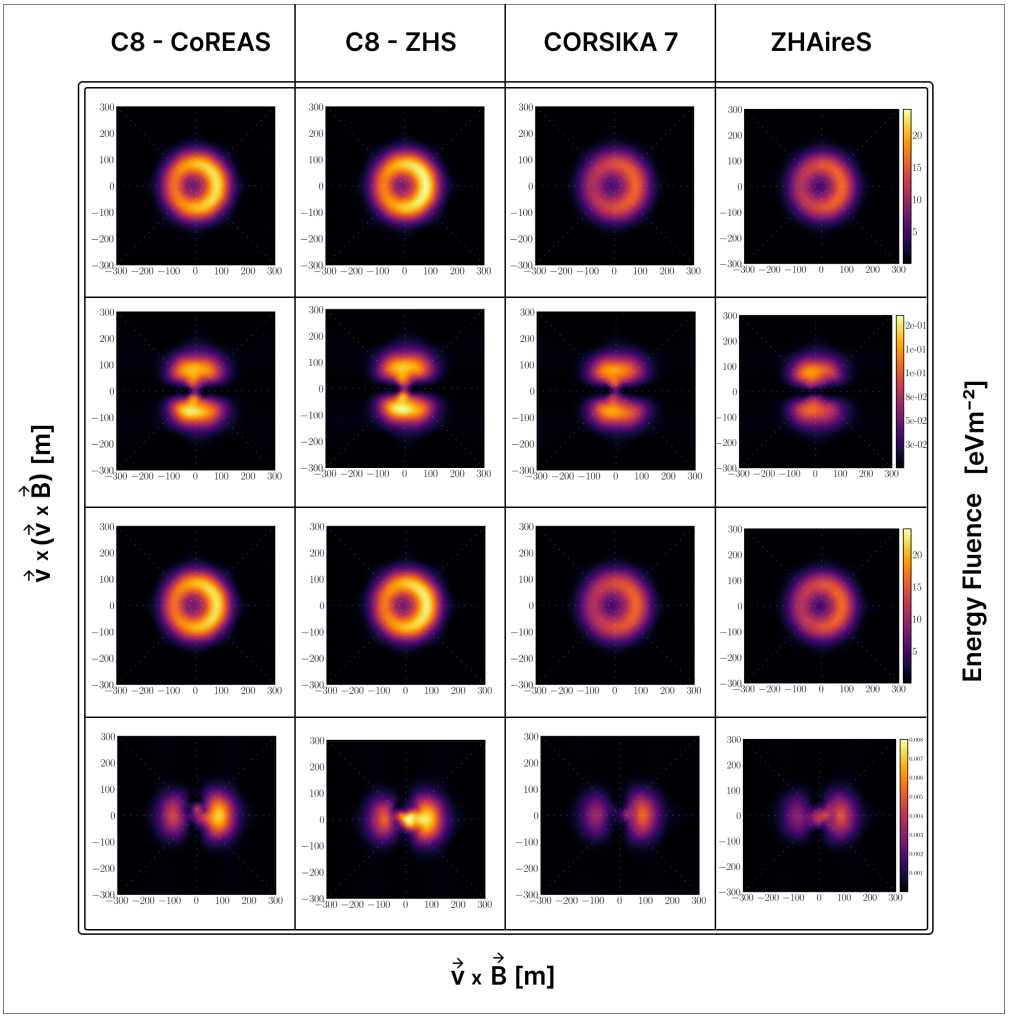


Figure 45: Table of energy fluence in different polarizations of the electric field in the 50 MHz to 350 MHz frequency band for C8 CoREAS, C8 ZHS, C7 CoREAS and ZHAireS. The order of the polarizations we see starting from top to bottom is: all polarizations, $\vec{v} \times (\vec{v} \times \vec{B})$, $\vec{v} \times \vec{B}$ and $\vec{v}$.

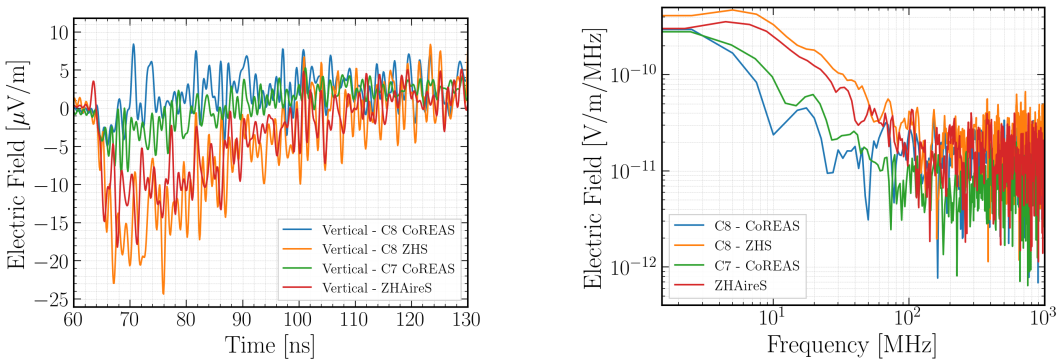


Figure 46: Signal pulse and frequency spectra comparison in 0 GHz to 1 GHz band, for antenna at 25 m from the shower core - Vertical polarization.

line is the electron and positron ratio and the dashed line is the photon ratio. For electrons and positrons which are contributing to the radio emission, it is shown that they are systematically located closer to the shower core for CORSIKA 8 (PROPOSAL v7.6.1). Up to 1 m, there are almost 25% more electrons and positrons and positrons and then the agreement fluctuates up to $\sim 15\%$ for distances of 1 m to 700 m. This would suggest that the produced radio signal is more coherent for CORSIKA 8 and could potentially explain the excess of fluence that is observed.

The fluence maps in both frequency ranges reveal an interesting observation in the $\vec{v}$ (or vertical) polarization. The CoREAS formalism (both in CORSIKA 8 and CORSIKA 7) and ZHS (both in CORSIKA 8 and ZHAireS) behave differently close to the shower core. ZHS predicts radio emission near the center of the shower core as seen in the form of a "blip", whereas CoREAS predicts no significant signals in the same area. To further look into this, the pulses and frequency spectra are plotted for an antenna very close to the shower core at 25 m (fig. 46). It is evident that ZHS predicts predicts $\sim 80\%$ more amplitude with respect to CoREAS especially in frequencies below 500 MHz. In practical terms though, the $\vec{v}$ polarization is negligible when compared to the dominant $\vec{v} \times \vec{B}$ polarization. This is clearly a difference in the predictions of the 2 formalisms. In the section 9 a detailed CoREAS versus ZHS comparison of the two formalisms is going to be discussed.

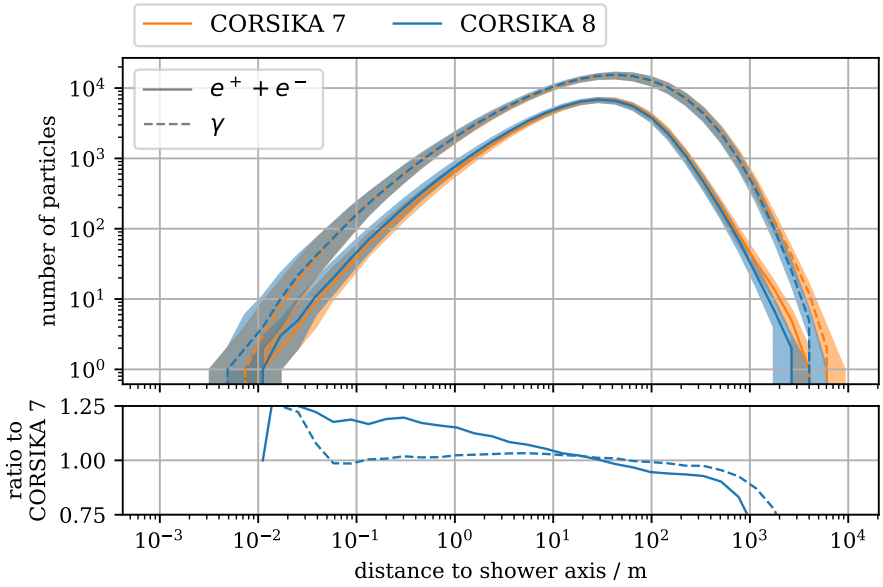


Figure 47: Median values of lateral profiles of electrons, positrons and photons for 100 TeV photon induced showers. Observation height for the evaluation is set to 5800 m, corresponding to $X_{\max} = 497\,\mathrm{g\,cm^{-2}}$. This plot was provided by Jean-Marco Alameddine [76].

8.2 A systematic comparison between CORSIKA 8 and CORSIKA 7

For a more systematic study a library of 200 CORSIKA 8 and 200 CORSIKA 7 simulations is produced. This library consists of 1 PeV iron induced vertical air showers. Energy cuts were set for both electromagnetic and hadronic components of the shower. For the electromagnetic cuts the value was set to 0.5 MeV, and for the hadrons and muons the cuts were set to 0.3 GeV. Regarding hadronic interaction models, Sibyll 2.3d [14] was used for high energies and for the low energy particles FLUKA [18] was used. Thinning was not applied. The environment was configured in all simulations to be the "US Standard Atmosphere", with a refractive index at sea level of $n = 1.000\,327$ and for higher altitudes scaling following the Gladstone-Dale law. A constant geomagnetic field of 50 µT aligned in the x direction was set. A star-shaped pattern of 80 time domain antennas for the antenna collection was set. The sampling period of each antenna is 0.1 ns. They are located at sea level, in 10 concentric rings spaced equally from 25 m to 250 m from the shower axis with 8 antennas distributed azimuthally in each ring. In all simulations the point of first interaction is fixed in order to minimize shower fluctuations. With this library the agreement of CORSIKA 8 and CORSIKA 7 is studied for different track lengths. This is achieved in organizing the simulations library in the following way. 100 of the CORSIKA 8 showers were simulated with setting the maximally allowed magnetic deflection angle (maxRad) equal to 0.2 rad and the other 100 with maxRad equal to 0.001 rad. The same way the CORSIKA 7 library was produced. The strategy for this study is to compare the showers produced with the maxRad value of 0.2 rad between themselves for CORSIKA 8 and CORSIKA 7 and evalute their

agreement in terms of their longitudinal profiles, fluence and radiation energy. Then, repeat the same comparison for the maxRad values of 0.001 rad.

To calculate and plot the mean total fluence per antenna i ($\overline{f^i_{\text{total}}}$), for CORSIKA 8 and CORSIKA 7 for all showers the following formula is used:

$$\overline{f^i_{\text{total}}} = \epsilon_0 c \Delta t \sum_{j=1}^{j=100} \sum_{t_{start}}^{t_{end}} E_{ji}(t)^2/100 \tag{28}$$

where j represents the shower number, in this case a total of 100. Using the same notation, to calculate and plot the mean values of the geomagnetic contribution and the charge excess on shower plane coordinates, the following formulas were used:

$$\overline{f^i_{geo}} = \sum_{j=1}^{j=100} \left(\sqrt{f^{ji}_{\vec{v}\times\vec{B}}} - \frac{cos\phi}{|sin\phi|} \cdot \sqrt{f^{ji}_{\vec{v}\times(\vec{v}\times\vec{B})}} \right)^2 /100 \tag{29}$$

$$\overline{f^i_{ce}} = \sum_{j=1}^{j=100} \left(\frac{1}{sin^2\phi} \cdot f^{ji}_{\vec{v}\times(\vec{v}\times\vec{B})} \right) /100 \tag{30}$$

ϕ is the polar angle between the antenna position and the positive $\vec{v}\times\vec{B}$ axis. For $\phi = 0$ or $\phi = \pi$ the $1/sin\phi$ term diverges. To tackle this, an interpolation takes place where these values are calculated as the mean value from their nearest neighbors on the same radial distance from the shower core.

Ratio maps for all showers in terms of percentage are calculated over the mean values of fluence of CORSIKA 8 and CORSIKA 7. For example, the percentage difference per antenna i (D_i) in total fluence, between CORSIKA 8 and CORSIKA 7 is taking the CORSIKA 7 value as reference and is calculated as such:

$$D_i = \frac{\overline{f^i_{\text{C7}}} - \overline{f^i_{\text{C8}}}}{\overline{f^i_{\text{C7}}}} \tag{31}$$

The same formula is applied for the geomagnetic contribution and the charge excess, in order to calculate the difference in percentage between CORSIKA 8 and CORSIKA 7. If the difference turns out positive that indicates that CORSIKA 7 produces more fluence than CORSIKA 8. If it is negative, then CORSIKA 8 produces more fluence than CORSIKA 7. Another important remark is how the total mean value of the difference of all antennas ($\overline{D}$) is calculated throughout the whole footprint. A weighted mean average scheme is utilized where every antenna location comes with a weight (w_i). The weight is defined as the fraction of the fluence of that antenna (f_i) over the maximum fluence (f_{max}) over the whole footprint. This

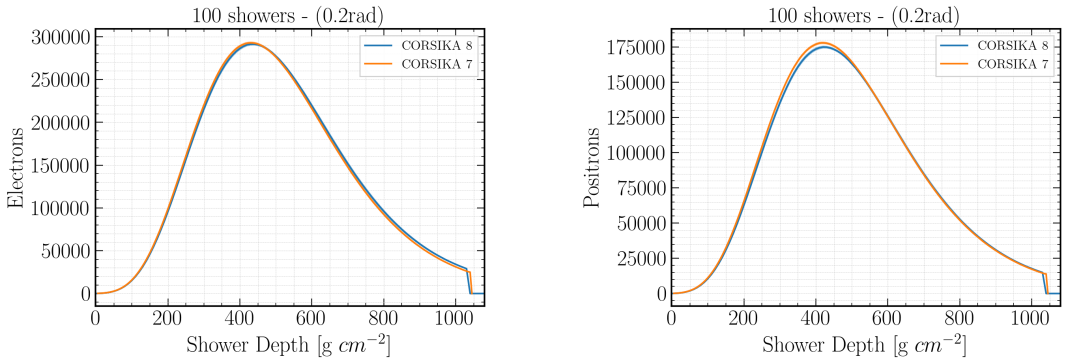


Figure 48: The longitudinal profiles over 100 showers for CORSIKA 7 and CORSIKA 8 of electrons (left) and positrons (right). The maximally allowed deflection angle is set to 0.2 rad.

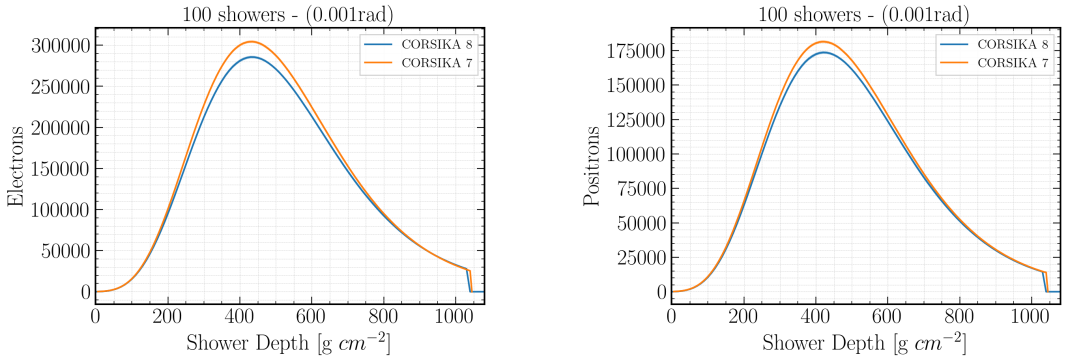


Figure 49: The longitudinal profiles over 100 showers for CORSIKA 7 and CORSIKA 8 of electrons (left) and positrons (right). The maximally allowed deflection angle is set to 0.001 rad.

way, it is ensured that to narrow down the total ratio percentage between CORSIKA 8 and CORSIKA 7, the antennas where the signal is strong matter more than the ones where the signal is weak. The quantity $\overline{D}$ can be interpreted as the quantity that describes how well the simulations compared (in this case CORSIKA 8 and CORSIKA 7 simulations), agree in terms of fluence over the whole footprint. A very similar scheme is going to be used throughout all the comparisons that take place in this section as well as in section 9 and section 10. The exact definition and usage of this scheme is going to be introduced in every different comparison. The weighting scheme can be summarized in the following formulas:

$$\overline{D} = \frac{\sum_i w_i D_i}{\sum_i w_i} \tag{32}$$

$$w_i = \frac{f_i}{f_{\max}} \tag{33}$$

Having defined the tools, the comparison can take place. Starting from the 0.2 rad case the longitudinal profiles (fig. 48) agree quite well with CORSIKA 7 producing $\sim 0.78\%$ more electrons and $\sim 0.57\%$ more positrons at the shower maximum. For the 0.001 rad case (fig. 49) the deviation gets larger with CORSIKA 7 producing $\sim 5.95\%$ more electrons and $\sim 1.21\%$ more positrons at the shower maximum.

Proceeding to the fluence maps, for the 0.2 rad case (fig. 50) CORSIKA 8 simulates consistently more fluence. In the 30 MHz to 80 MHz band, the smaller deviations are observed close to the shower core up to the Cherenkov ring and further where the signal grows weaker the deviations become larger. In the 50 MHz to 350 MHz, the picture is reversed with the deviations being large close to the shower core up to the Cherenkov ring and getting smaller as the signal gets weaker for distances larger than the Cherenkov radius. The overall agreement between the 2 simulation codes can be summarized by the radiation energy normalized over the electromagnetic energy of the shower squared. For 30 MHz to 80 MHz, CORSIKA 8 simulates $\sim 34.9\%$ more radiation energy, while for 50 MHz to 350 MHz CORSIKA 8 simulates $\sim 57.1\%$ more radiation energy. The differences for fluence and radiation energy are summarized in table 1 and table 2 respectively.

Looking into the 0.001 rad case (fig. 51) the comparison gets interesting. Both simulation codes now produce very small and fine tracks for the particle trajectories. Minus values mean that CORSIKA 8 predicts more fluence, while positive values mean that CORSIKA 7 predicts more fluence. In the 50 MHz to 350 MHz band, CORSIKA 7 consistently simulates more fluence with larger deviations being within the Cherenkov ring and them becoming smaller moving away from it. In the 50 MHz to 350 MHz band close to the shower core up to the Cherenkov ring, CORSIKA 8 simulates more fluence, while moving to larger distances than the Cherenkov radius, CORSIKA 7 simulates more fluence. Overall, in the 30 MHz to 80 MHz band CORSIKA 7 simulates $\sim 8.3\%$ more radiation energy and for the 50 MHz to 350 MHz band, both simulation codes almost converge with CORSIKA 8 simulating $\sim 1.7\%$ more radiation energy. The differences for fluence and radiation energy are summarized in table 3 and table 4 respectively.

This is a very interesting finding as it indicates that the differences in fluence that are observed are possibly a result of how the particles are distributed laterally. More particles close to the shower core spreading less in radial distances can lead to radio signals being more coherent and hence attributing to more fluence. According to fig. 47, that seems to be the case for CORSIKA 8. Forcing both simulation codes to produce very fine and small tracks seems to mitigate this issue at least in terms of radiation energy. Further systematic studies on lateral profiles of electrons and positrons for different maxRad values might prove to be very beneficial. A study that could also shed some light to this for the future, would be to produce big libraries of CORSIKA 7 and CORSIKA 8 simulations for proton, iron and photon showers and set up many observation levels on different stages of the shower development, especially close to $X_{\max}$. The issue of the lateral profiles of particles (in the case of radio, electrons and positrons) being different should become apparent and would explain the differences that the radio simulations hint us. Such a fundamental deviation between the 2 simulation codes could be explained by differences in the

tracking algorithm used, or the multiple scattering implementations. Regarding the tracking algorithm as was shown in section 6.2, one could could study further whether the default value for maximally allowed magnetic deflection angle used (0.2 rad) is sufficient to describe accurately the particle trajectories. The effect of different multiple scattering approximations on radio simulations is discussed in section 10.2.

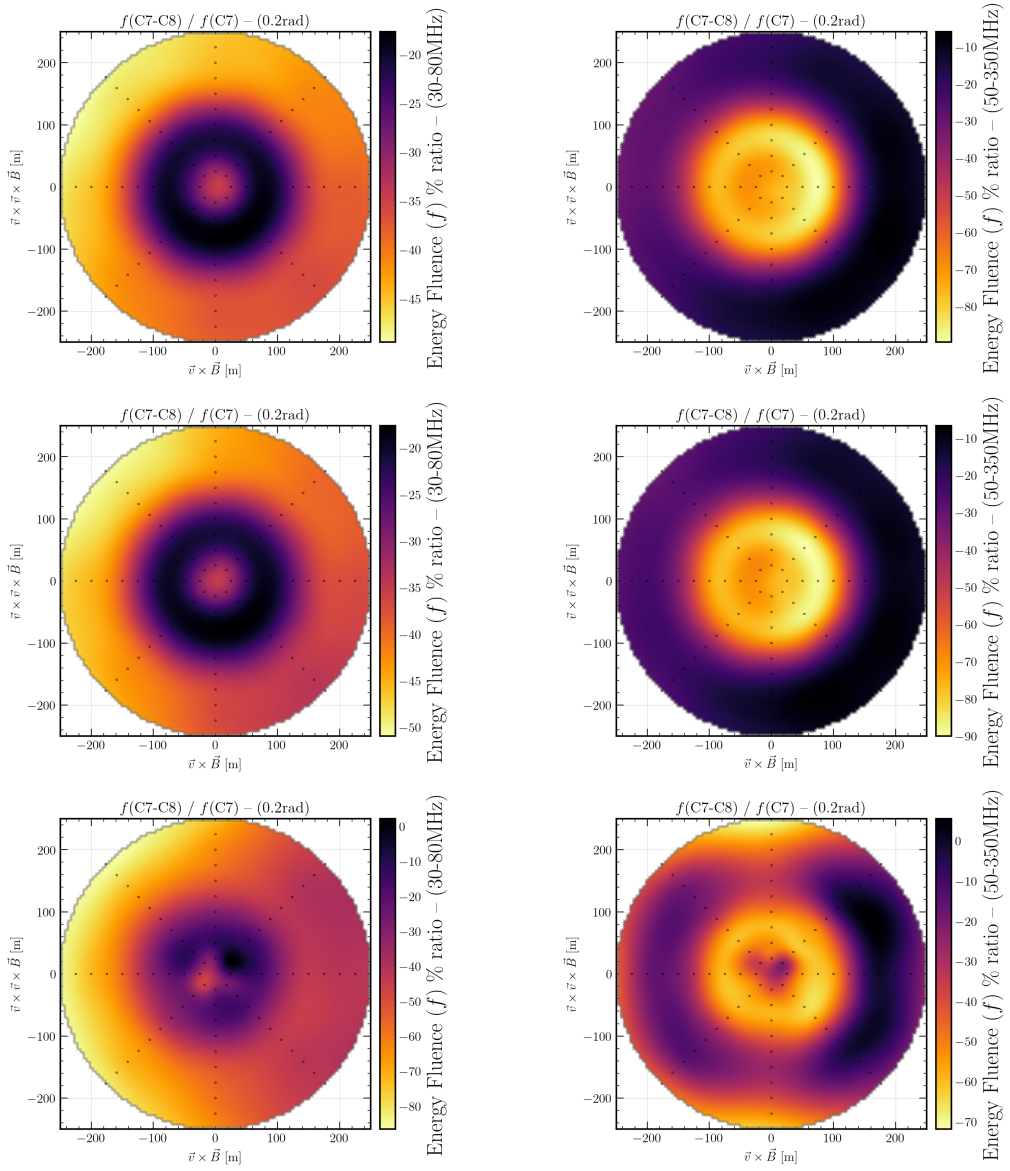


Figure 50: Ratio maps in fluence for 100, 1 PeV, iron induced vertical showers in the 30 MHz to 80 MHz (left column) and the 50 MHz to 350 MHz band (right column). In the first row, the ratio map in terms of total fluence is plotted, the second in terms of the geomagnetic contribution and the last row in terms of the charge excess. The maximally allowed deflection angle is set to 0.2 rad and CORSIKA 8 CoREAS is directly compared to CORSIKA 7 CoREAS.

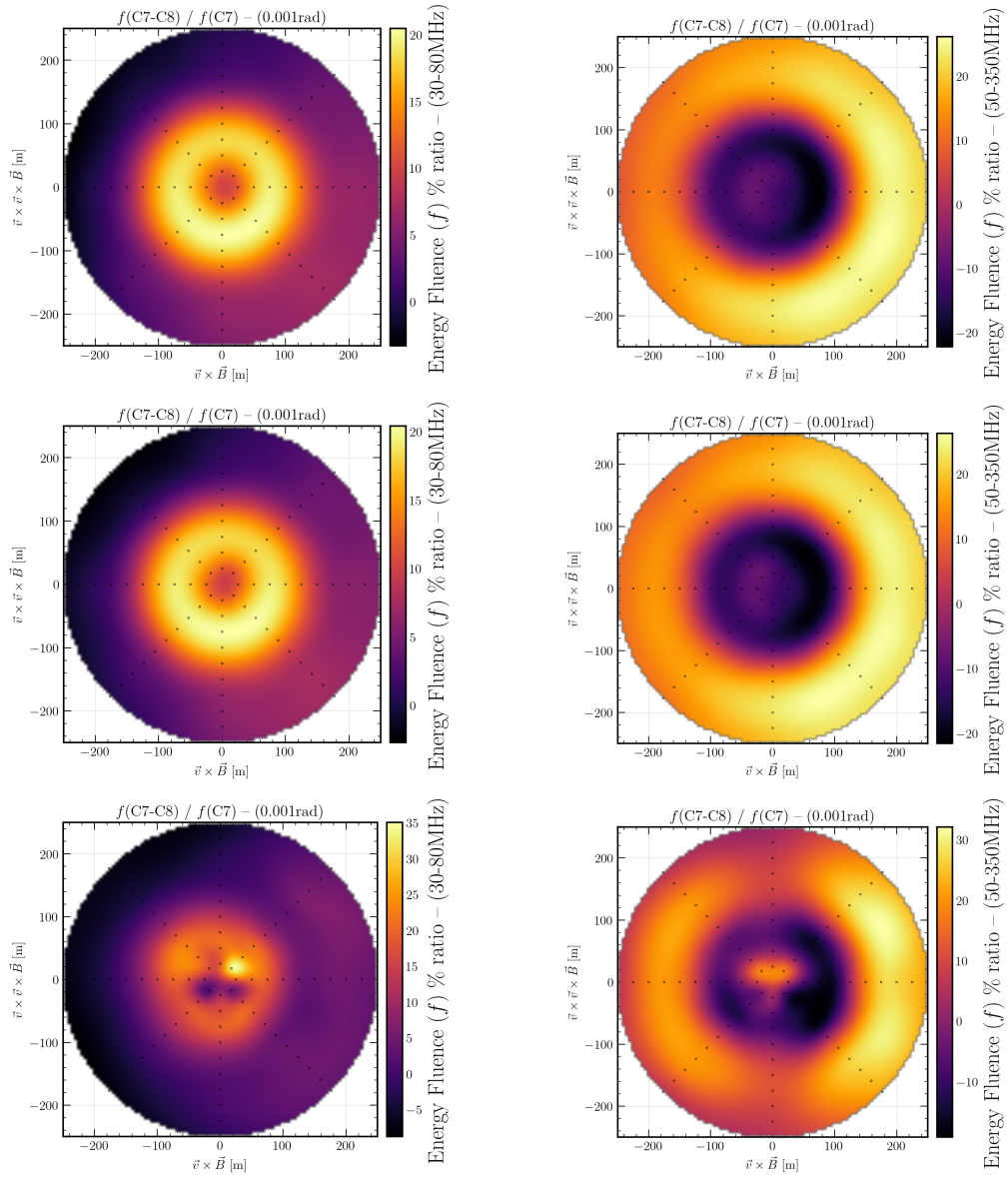


Figure 51: Ratio maps in fluence for 100, 1 PeV, iron induced vertical showers in the 30 MHz to 80 MHz (left column) and the 50 MHz to 350 MHz band (right column). In the first row, the ratio map in terms of total fluence is plotted, the second in terms of the geomagnetic contribution and the last row in terms of the charge excess. The maximally allowed deflection angle is set to 0.001 rad and CORSIKA 8 CoREAS is directly compared to CORSIKA 7 CoREAS.

CORSIKA 8 vs CORSIKA 7 – mean difference in terms of % (0.2rad)		
Fluence	30 MHz to 80 MHz	50 MHz to 350 MHz
Total Fluence	-28.2%	-62.4%
Geomagnetic Contribution	-28.4%	-62.7%
Charge Excess	-39.9%	-37.8%

Table 1: The difference in terms of % for fluence between CORSIKA 8 and CORSIKA 7 CoREAS for maxRad set to 0.2 rad. The minus sign indicates that more fluence is simulated with CORSIKA 8.

CORSIKA 8 vs CORSIKA 7 – mean difference in terms of % (0.2rad)		
Radiation Energy	30 MHz to 80 MHz	50 MHz to 350 MHz
Total Radiation Energy	-34.9%	-57.1%
Geomagnetic Contribution	-34.8%	-57.0%
Charge Excess	-49.3%	-34.1%

Table 2: The difference in terms of % for radiation energy between CORSIKA 8 and CORSIKA 7 CoREAS for maxRad set to 0.2 rad. The minus sign indicates that more radiation energy is simulated with CORSIKA 8.

CORSIKA 8 vs CORSIKA 7 – mean difference in terms of % (0.001rad)		
Fluence	30 MHz to 80 MHz	50 MHz to 350 MHz
Total Fluence	13.1%	-4.8%
Geomagnetic Contribution	12.9%	-4.7%
Charge Excess	8.1%	3.8%

Table 3: The difference in terms of % for fluence between CORSIKA 8 and CORSIKA 7 CoREAS for maxRad set to 0.001 rad. The minus sign indicates that more fluence is simulated with CORSIKA 8 and the plus sign that more fluence is simulated with CORSIKA 7.

CORSIKA 8 vs CORSIKA 7 – mean difference in terms of % (0.001rad)		
Radiation Energy	30 MHz to 80 MHz	50 MHz to 350 MHz
Total Radiation Energy	8.3%	-1.7%
Geomagnetic Contribution	8.2%	-1.8%
Charge Excess	2.6%	6.0%

Table 4: The difference in terms of % for radiation energy between CORSIKA 8 and CORSIKA 7 CoREAS for maxRad set to 0.001 rad. The minus sign indicates that more radiation energy is simulated with CORSIKA 8 and the plus sign that more radiation energy is simulated with CORSIKA 7.

8.3 Discussion

In this chapter the radio simulations with CORSIKA 8 were validated with well established software. For a similar shower as was initially compared with CORSIKA 7 and ZHAireS, CORSIKA 8 simulations predict consistently more fluence and stronger pulses than the other simulation codes up to 25%. To inspect this further, a library of CORSIKA 8 and CORSIKA 7 simulations were produced. It was found that for the standard setting that defines the track length of particle tracks, which is 0.2 rad, CORSIKA 8 predicts 34.9% more radiation energy in the 30 MHz to 80 MHz band and 57.1% more radiation energy in the 50 MHz to 350 MHz band. Such a large deviation when utilizing the default values is unacceptable. Although, when choosing a maxRad value of 0.001 rad which effectively makes the track length of simulated particle tracks shorter, the agreement gets much better. In the 30 MHz to 80 MHz band CORSIKA 7 predicts 8.3% more radiation energy and in the 50 MHz to 350 MHz band CORSIKA 8 predicts 1.7% more radiation energy. A roughly 10% difference like in the case of the 30 MHz to 80 MHz band, while not ideal, means a roughly 5% difference in energy scale. This is acceptable as it is not going to dominate over the systematic uncertainties of an energy scale measurement.

It is worth noting that CORSIKA 7 and CORSIKA 8 use different electromagnetic interaction models. However, such a disagreement could potentially be attributed to the difference in the tracking algorithms. Making smaller and finer tracks is computationally heavy, but one could argue that this is "closer to reality". The difference in tracking, especially for the 0.2 rad case could affect the lateral distribution of the showers. A future study on the dependency of the lateral shower profiles on maxRad could reveal the discrepancy, and thus initiate a revision of the CORSIKA 8 magnetic field tracking. It could be that in order for the CORSIKA 8 magnetic field tracking to match the accuracy of the CORSIKA 7 tracking algorithm, a lower maxRad value needs to be set. What we could potentially witness in the radio simulations is the effect of showers spreading "less", hence more particles are closer to the shower core, resulting to more coherent, stronger pulses for CORSIKA 8.

9 A systematic CoREAS versus ZHS comparison

The radio emission in air showers is a well understood process based on first principles [40], [41]. The radiation energy due to energy conservation, and the fact that no absorption is taking place in the atmosphere reflects the calorimetric energy in the electromagnetic cascade of an air shower. This means that the same shower can produce the same amount of radiation energy at different sites with altering factors being the magnetic field strength and the height at the different locations. As a result, the radiation energy represents a well defined quantity that is measured by the radio antennas of different experiments and can be used to compare them directly. To perform such a feat, the formalisms and their corresponding algorithms must agree among themselves to prove that the radio emission is well understood indeed. Technical parameters of the simulations should not affect the radio simulations ideally. This the main topic of this chapter. To compare and inspect technical parameters that might influence the two most widely used radio calculations formalisms, namely CoREAS and ZHS. A good agreement between the two, confirms that the radio emission is indeed well understood by the community even with different modelling approaches.

Having both the CoREAS and ZHS formalisms built into CORSIKA 8 allows a full-fledged comparison that takes out any uncertainty coming from the differences of their underlying simulation codes. This is the case when trying to compare the two using CORSIKA 7 and ZHAiRES. However, both algorithms in CORSIKA 8 are being fed with identical particle tracks, which allows for a "true" comparison. In this chapter CoREAS and ZHS are going to be compared for a variety of scenarios. These scenarios constitute of studies of their agreement in terms of different primary particle energy, different primary particle, different multiple scattering approaches and finally different simulated track lengths.

9.1 The track length as a factor of agreement between the CoREAS and ZHS formaslisms

For a first comparison, the simulation data produced in section 8 for the 100 PeV iron induced vertical shower with maxRad set to 0.2 rad is going to be used. The difference in total fluence for 2 frequency bands is being shown in fig. 52 and fig. 53 for the antennas where the signal is prominent, i.e. distances from the shower core up to 250 m. CoREAS is being compared with respect to ZHS, by calculating the ratio of the total fluence predicted with COREAS subtracted from the same quantity predicted with ZHS and then normalized over the total fluence calculated with ZHS. This is done for each antenna in the antenna array. The total fluence percentage ratio D_i is plotted in the form of a map, where the black dots indicate the antenna positions (i) and the in between values are interpolated (fig. 52, fig. 53 left). In fig. 52 and fig. 53 (right), every data point is the mean value of the difference of every concentric ring of antennas per 25 m. The error bars in fig. 52 and fig. 53 (right) indicate the standard deviation of the mean value of the difference of every concentric ring of antennas divided by the number of antennas in the

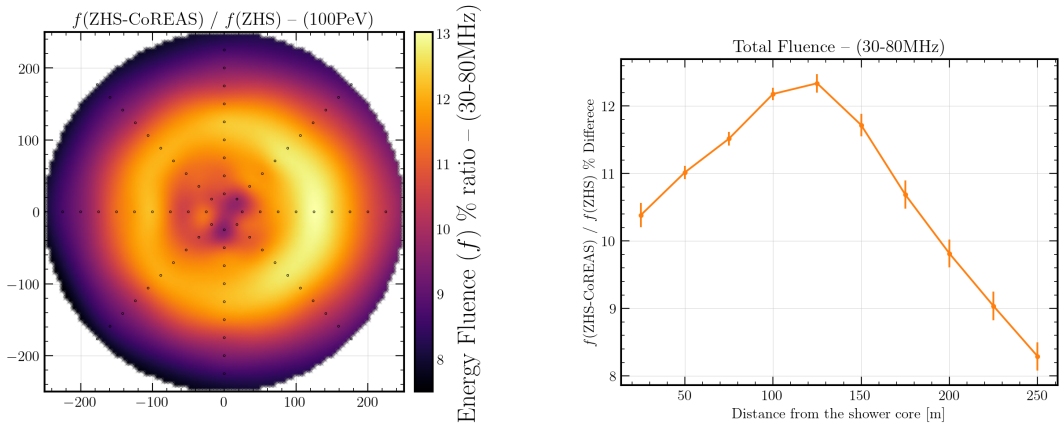


Figure 52: Difference in total fluence between CoREAS and ZHS in the 30 MHz to 80 MHz band for a 100 PeV iron induced vertical shower. ZHS produces consistently more fluence across the whole footprint.

concentric ring, which is 8.

Similarly to eq. (35), D_i is defined as the difference in total fluence between ZHS and CoREAS over the total fluence of ZHS per antenna i (eq. (34)), but in this case for one shower. The total difference over the whole footprint between the 2 algorithms is the weighted mean defined in the same was as eq. (32), where the weight w_i (eq. (33)) is defined as the fluence of the antenna i (f_i) over the maximum fluence of all antennas ($f_{\max}$) in the simulation.

$$D_i = \frac{f^i_{\text{ZHS}} - f^i_{\text{CoREAS}}}{f^i_{\text{ZHS}}} \tag{34}$$

In the 30 MHz to 80 MHz band ZHS produces roughly 11% more fluence than CoREAS, while for the 50 MHz to 350 MHz band ZHS produces roughly 4.8% more fluence. The picture is similar when one looks into the radiation energy released from the showers in both frequency bands. For the 30 MHz to 80 MHz band ZHS produces roughly 11% more energy and for the 50 MHz to 350 MHz band the energy released is almost 5% more for ZHS. An interesting feature is observed. For the lower frequencies (fig. 52), the worst disagreement happens close to the Cherenkov ring (100 m to 125 m) reaching 13%. On the other hand, for the higher frequencies (fig. 53) in the area close to the Cherenkov ring (100 m to 125 m) the agreement is at its best reaching almost 4%. For the higher frequencies the good agreement is to be expected, given the fact that CoREAS treats the particle tracks which are very close to the Cherenkov in a similar manner as ZHS, often called a "ZHS-like" approximation. It is surprising though to see, that this approximation does not seem to make the 2 algorithms agree more in the 30 MHz to 80 MHz band.

It is observed in this work that the level of agreement is severely influenced by the

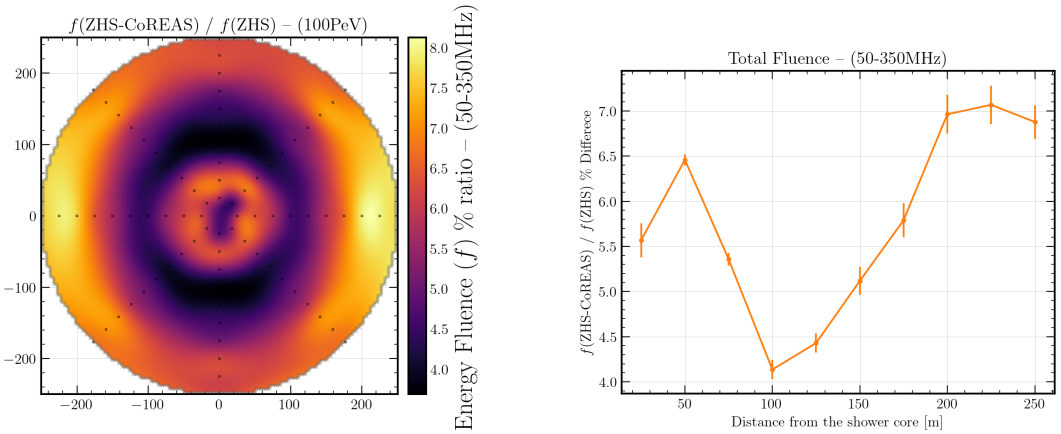


Figure 53: Difference in total fluence between CoREAS and ZHS in the 50 MHz to 350 MHz band for a 100 PeV iron induced vertical shower. ZHS produces consistently more fluence across the whole footprint.

track length of the particle tracks. The track length is indirectly controlled in the magnetic field tracking algorithm by a parameter that sets the maximally allowed deflection angle (also referred simply as maxRad in short), which is set to 0.2 rad by default. To study this, a library of shower simulations with different track lengths was produced. The settings are identical to the ones discussed in section 8.2 when it comes for the setup of the atmosphere, antenna array, energy cuts etc. The effect of different track lengths on fluence and radiation energy is being studied for 3 different primary energies. These are 10 TeV (fluence fig. 54, radiation energy fig. 55), 1 PeV (fluence fig. 56, radiation energy fig. 57) and 100 PeV (fluence fig. 58, radiation energy fig. 59). For the smaller energy showers, 10 TeV and 1 PeV, thinning is turned off. In fig. 54, fig. 56 and fig. 58, each data point represents the mean value of the difference in fluence of the 2 algorithms over 80 antennas for a variety of maxRad values. The error bars indicate the standard deviation divided by the number of antennas. It is worth noting, that the lower the value of the maxRad parameter is set, the slower the computing times get. This is normal, as CORSIKA 8 needs to produce and process a larger number of tracks.

In all 3 energies under consideration the trend is clear. Starting from the default value of 0.2 rad (which is also the default value of CORSIKA 7), ZHS consistently produces more fluence and radiation energy in both 30 MHz to 80 MHz (left) and 50 MHz to 350 MHz bands. As the maxRad value and hence the track length decrease, the agreement gets better and better, up to the point that for extremely small tracks (maxRad of the order of 0.001 rad and lower) CoREAS predicts more fluence and radiation energy than ZHS by a very small margin. In fig. 60 the level of agreement between the two formalisms in radiation energy is shown in two frequency bands with respect to primary particle energy for maxRad set to 0.2 rad. The agreement is not affected much as is also shown here, with the most striking deviation being the 10 TeV shower in the 50 MHz to 350 MHz. This can be probably explained by the fact that this is a very low energy iron induced shower where very low values of radiation energy are compared. For the higher and more relevant energies though, it is confirmed that the agreement is not affected. Taking into account the energy and track length dependence, a more thorough comparison of the 2 formalisms is presented next.

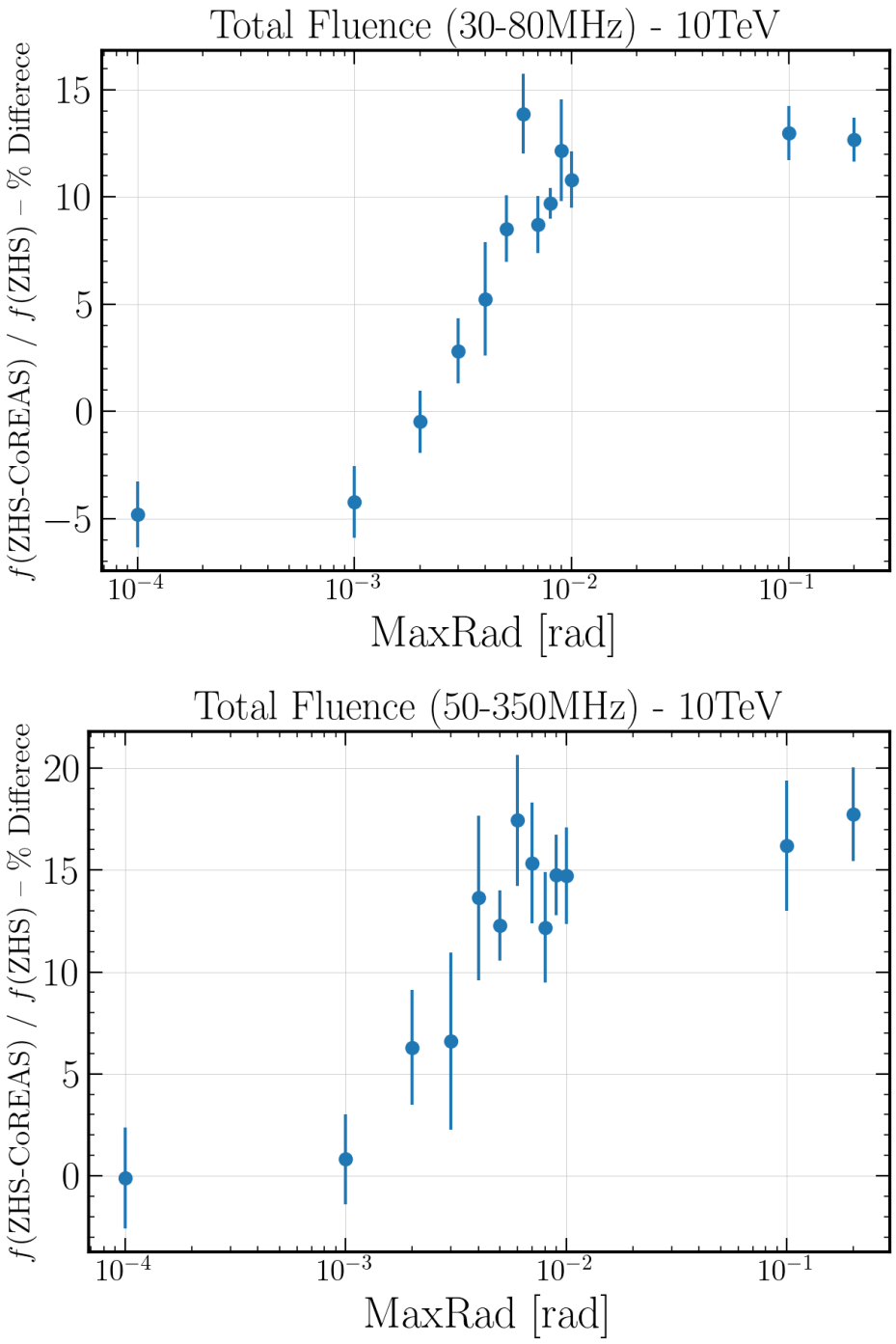


Figure 54: Difference in total fluence between CoREAS and ZHS in the 30 MHz to 80 MHz (top) and the 50 MHz to 350 MHz band (bottom) for a 10 TeV iron induced vertical shower. The change in the maximally allowed magnetic deflection angle influences the agreement between the 2 algorithms.

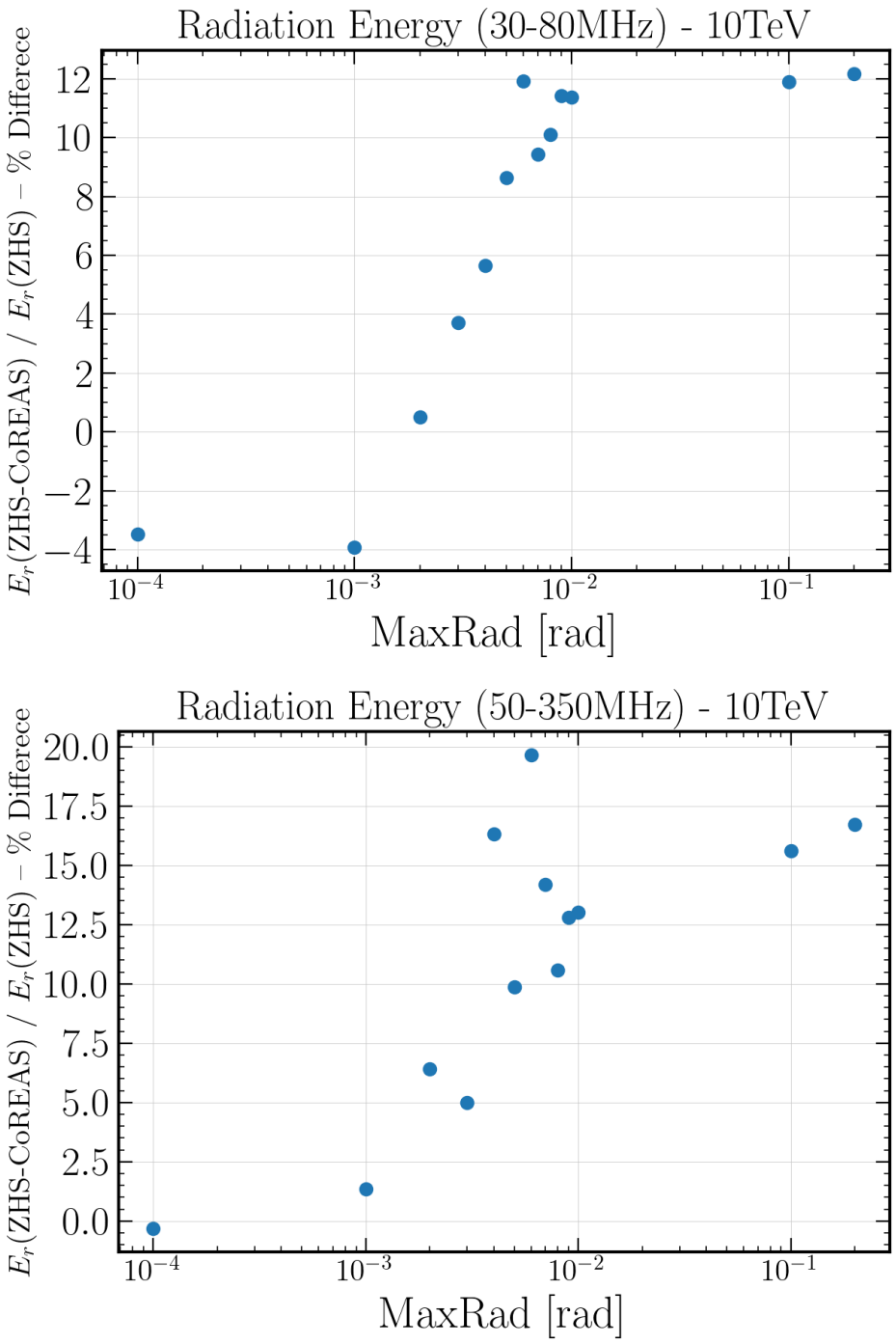


Figure 55: Difference in radiation energy between CoREAS and ZHS in the 30 MHz to 80 MHz (top) and the 50 MHz to 350 MHz band (bottom) for a 10 TeV iron induced vertical shower. The change in the maximally allowed magnetic deflection angle influences the agreement between the 2 algorithms.

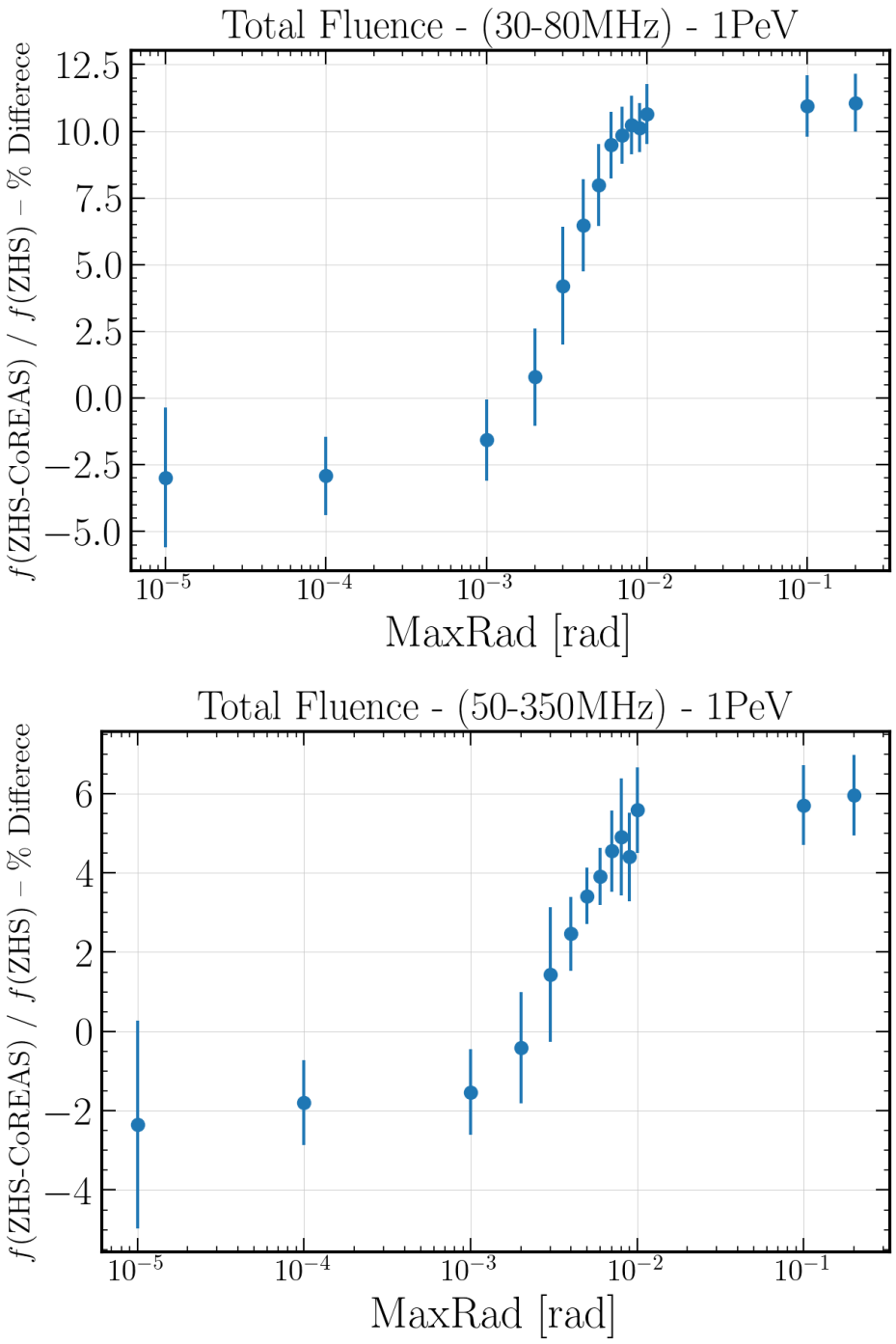


Figure 56: Difference in total fluence between CoREAS and ZHS in the 30 MHz to 80 MHz (top) and the 50 MHz to 350 MHz band (bottom) for a 1 PeV iron induced vertical shower. The change in the maximally allowed magnetic deflection angle influences the agreement between the 2 algorithms.

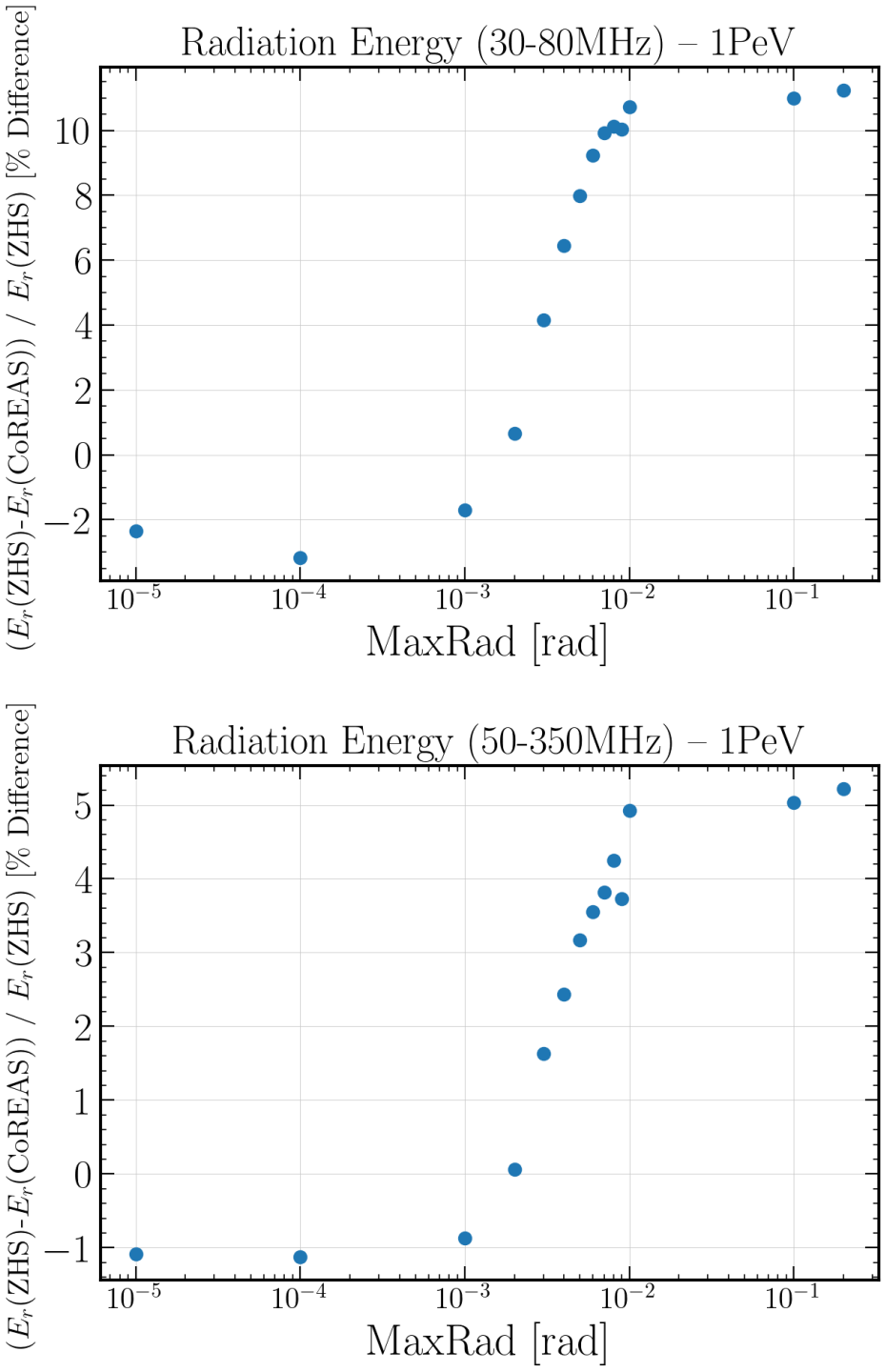


Figure 57: Difference in radiation energy between CoREAS and ZHS in the 30 MHz to 80 MHz (top) and the 50 MHz to 350 MHz band (bottom) for a 1 PeV iron induced vertical shower. The change in the maximally allowed magnetic deflection angle influences the agreement between the 2 algorithms.

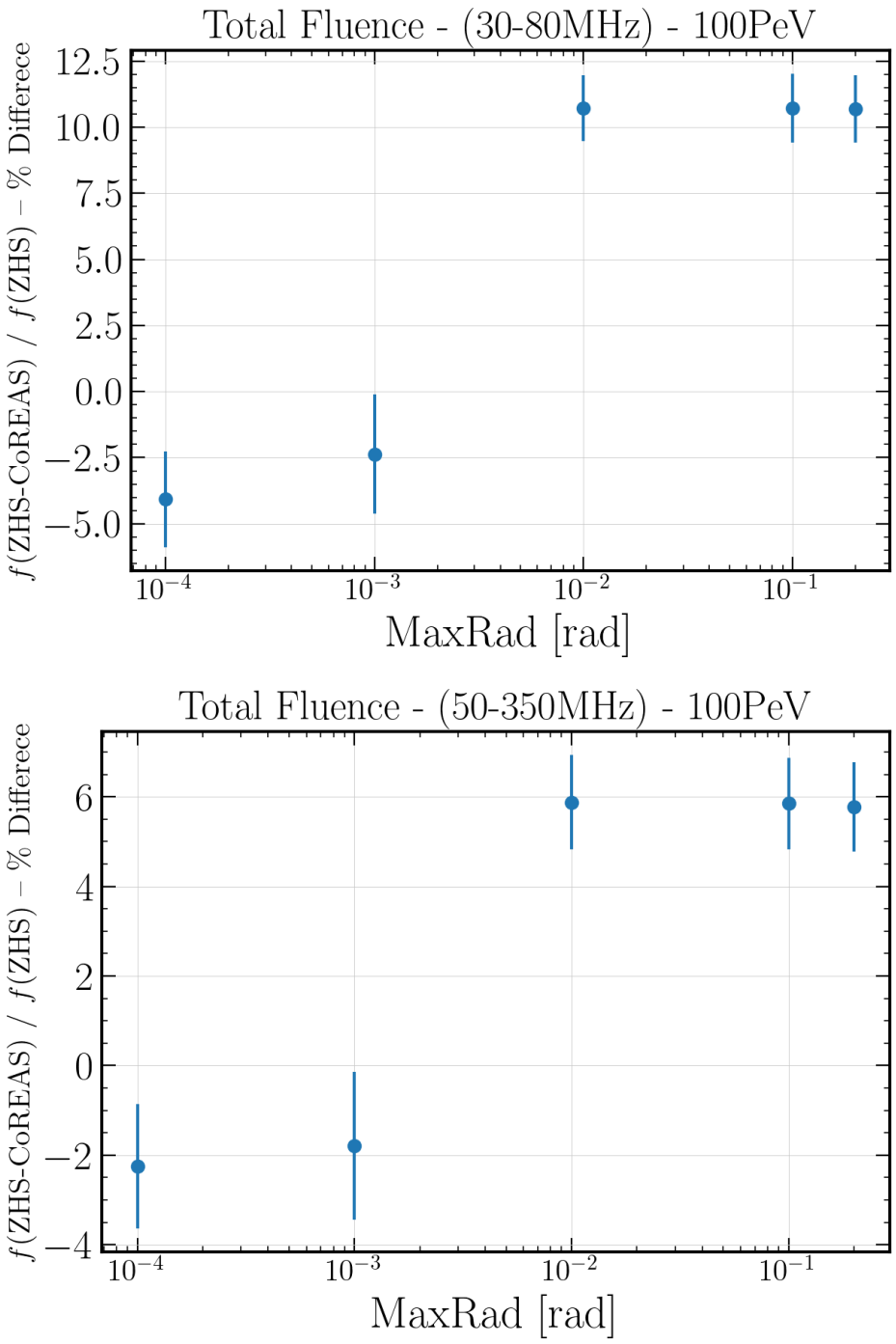


Figure 58: Difference in total fluence between CoREAS and ZHS in the 30 MHz to 80 MHz (top) and the 50 MHz to 350 MHz band (bottom) for a 100 PeV iron induced vertical shower. The change in the maximally allowed magnetic deflection angle influences the agreement between the 2 algorithms.

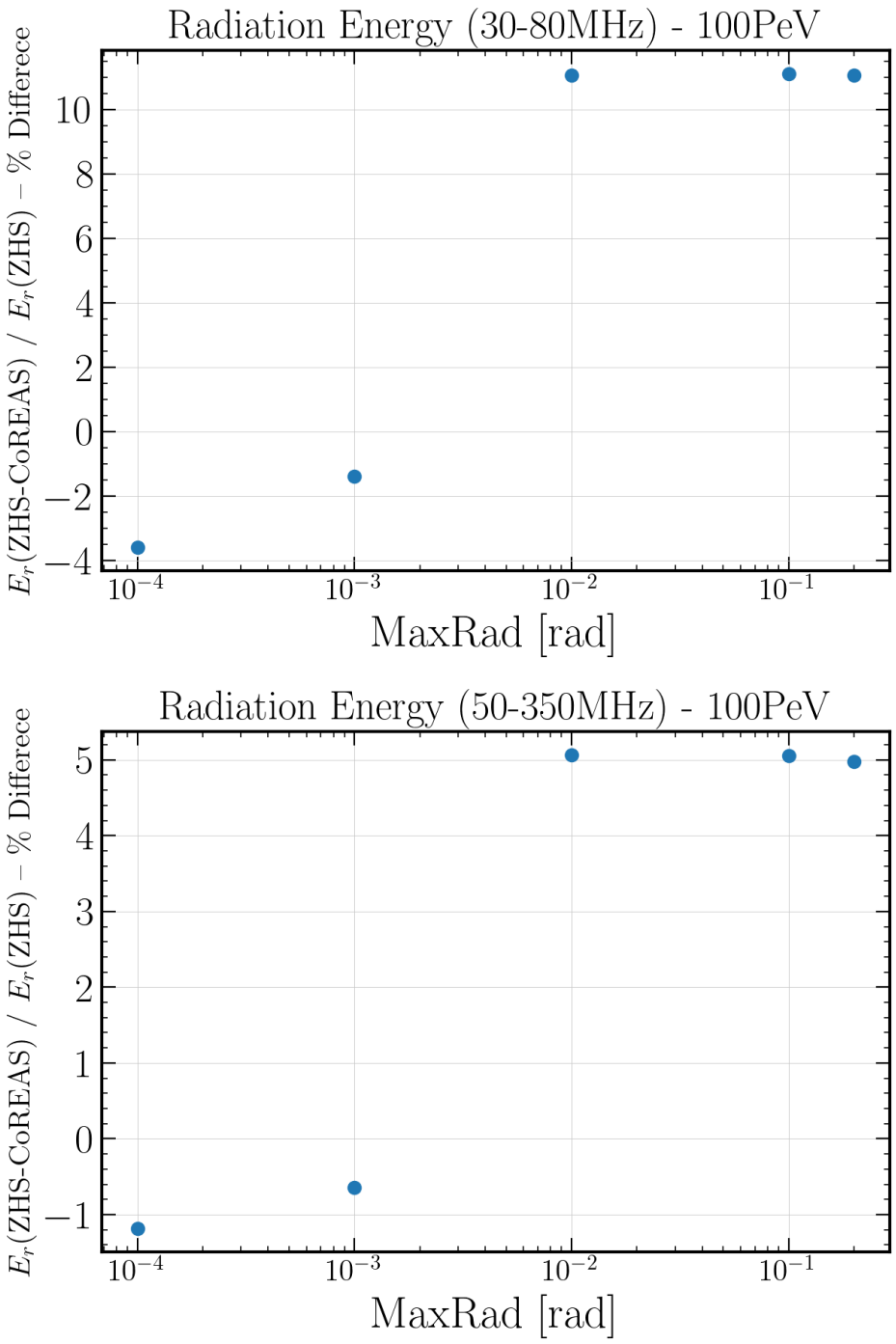


Figure 59: Difference in radiation energy between CoREAS and ZHS in the 30 MHz to 80 MHz (top) and the 50 MHz to 350 MHz band (bottom) for a 100 PeV iron induced vertical shower. The change in the maximally allowed magnetic deflection angle influences the agreement between the 2 algorithms.

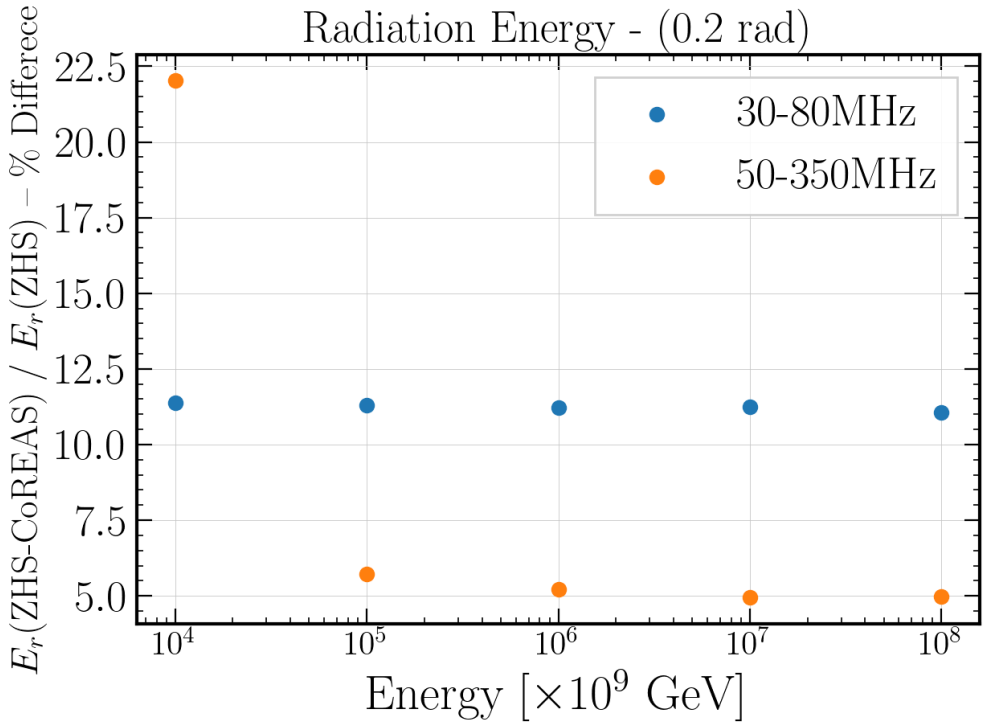


Figure 60: Level of agreement between the 2 formalisms in terms of radiation energy for iron induced air showers with different primary energies. The blue dots indicate the 30 MHz to 80 MHz frequency band and the orange ones the 50 MHz to 350 MHz. For all these shower simulations the maxRad parameter is set to 0.2 rad.

9.2 A systematic CoREAS versus ZHS comparison

9.2.1 Iron induced air showers

For a more systematic study a library of 100 iron induced vertical showers with a primary energy of 1 PeV with the maxRad value set to the default value of 0.2 rad was produced. Additionally, 100 showers that utilized the same random number seeds, with the maxRad value set to 0.001 rad were also produced. Once again, the settings of all 200 simulations are identical to the ones discussed in section 8.2 when it comes for the setup of the atmosphere, antenna array, energy cuts, interaction models etc. Thinning was not applied to the simulations of this section and the point of first interaction is fixed. In this way, the showers fluctuate less and the effect of the track length is being studied on the showers themselves, on the energy deposited to the ground and also on the agreement of CoREAS and ZHS. The mean value of the longitudinal profiles of the showers are shown for the different maxRad values (fig. 68) at later comparison.

This comparison is not restricted only in the total fluence; the fluence pattern produced from the geomagnetic emission and charge excess are decoupled and studied as well. Fluence ratio maps highlighting the agreement of CoREAS and ZHS across the whole footprint for different maxRad values and different frequency bands are shown (fig. 61, fig. 62). To calculate the mean total fluence of all showers per antenna i ($\overline{f^i_{\text{total}}}$), the mean value of the geomagnetic contribution ($\overline{f^i_{\text{geo}}}$) and the mean value of charge excess ($\overline{f^i_{\text{ce}}}$), eq. (35), eq. (29) and eq. (30) were used respectively. On fig. 63, 3 antennas at growing distances from the shower core on the arm facing "north" of the star-shapped array are chosen and the mean value together with its corresponding standard deviation of the CoREAS and ZHS frequency spectra are plotted in the 0 GHz to 1 GHz band. Finally, the difference in the total radiation energy and radiation energy due to geomagnetic and charge excess emission between the 2 formalisms is calculated.

For fig. 61 and fig. 62 the ratio maps in terms of percentage is calculated over the mean values of fluence of CoREAS and ZHS. For example, the percentage difference per antenna i (D_i) for all simulations in total fluence, between CoREAS and ZHS is taking the ZHS value as reference and is calculated as such:

$$D_i = \frac{\overline{f^i_{\text{ZHS}}} - \overline{f^i_{\text{CoREAS}}}}{\overline{f^i_{\text{ZHS}}}} \tag{35}$$

The same formula is applied for the geomagnetic contribution and the charge excess, in order to calculate the difference in percentage between the CoREAS and ZHS formalisms. If the difference turns out positive that indicates that ZHS produces more fluence than CoREAS. If it is negative, then CoREAS produces more fluence than ZHS. Another important remark is that the total mean value of the difference ($\overline{D}$) for all simulations is calculated throughout the whole footprint using eq. (32). Similarly, a weighted mean average scheme is utilized where every antenna location comes with a weight (w_i) as defined in eq. (33). This weighting scheme makes

antennas where the signal is strong matter more than the ones where the signal is weak. The quantity $\overline{D}$ expresses the level of agreement between the two formalisms for fluence over the whole footprint in this case.

Starting from fig. 61 CoREAS is compared over ZHS for the default maxRad value used in the magnetic field tracking algorithm, which is 0.2 rad. For both 30 MHz to 80 MHz and 50 MHz to 350 MHz frequency bands ZHS produces more fluence than CoREAS. Their difference in terms of % is summarized in table 5 for fluence and in table 6 for radiation energy. Total fluence, radiation energy, geomagnetic contribution and charge excess in fluence and radiation energy are inspected. Naturally, since the geomagnetic emission is the dominant mechanism of radio emission, its difference is very close to the one coming from the total contribution. For charge excess contribution, both formalisms tend to agree more. In the 30 MHz to 80 MHz band, the difference is at its highest for the total fluence and geomagnetic contribution in the area close to the Cherenkov ring, while for the charge excess the formalisms disagree more very close to the shower core. However, one needs to take into account that the charge excess fluences are very small. For the 50 MHz to 350 MHz band, the picture is similar for the charge excess. For the total fluence and hence the geomagnetic contribution though, the agreement is at its highest close to the Cherenkov ring and where the algorithms diverge more is the areas where the signal is weaker. To get an overall number of the level of agreement one can look into the differences of the radiation energy. For the 30 MHz to 80 MHz band, ZHS predicts $\sim 11.3\%$ more radiation energy, while for the 50 MHz to 350 MHz band, ZHS predicts $\sim 4.9\%$ more radiation energy. It is worth noting that a minus sign in these percentages simply means that CoREAS predicts more energy.

Proceeding to the 0.001 rad case the ratio maps are shown in fig. 62. The difference in terms of % is summarized in table 7 for fluence and in table 8 for radiation energy. In this case, the picture is reversed. CoREAS predicts more energy than ZHS although by a small margin, and in the areas where the signal is strong the 2 algorithms converge. The largest deviations are observed in the areas where the signal is weak. Overall, for the 0.001 rad and effectively very small tracks the 2 algorithms converge. As an indication, for the the 30 MHz to 80 MHz band CoREAS predicts $\sim 1.4\%$ more radiation energy and for the 50 MHz to 350 MHz band predicts $\sim 1\%$ more radiation energy than ZHS. This is a great result as it showcases that for detailed simulations where the track length is set to be very small and hence, more realistic, both formalisms practically predict the same outcome. This is also an indication that they radio emission is air showers is a well understood process.

The behavior observed in fluence and radiation energy can also be seen in the frequency spectra for both maxRad values (fig. 63) at different antenna locations. The average frequency spectra of 100 iron induced showers are shown for antennas at 25 m, 100 m and 200 m. It is evident that for 0.001 rad the 2 algorithms converge, and in general they diverge more for lower frequencies (< 50 MHz). Finally, the fluence maps of the average of 100 showers (per maxRad value) in terms of total fluence, fluence due to geomagnetic emission and due to charge excess for CoREAS and ZHS for both 30 MHz to 80 MHz and 50 MHz to 350 MHz bands are plotted in detail on the Appendix.

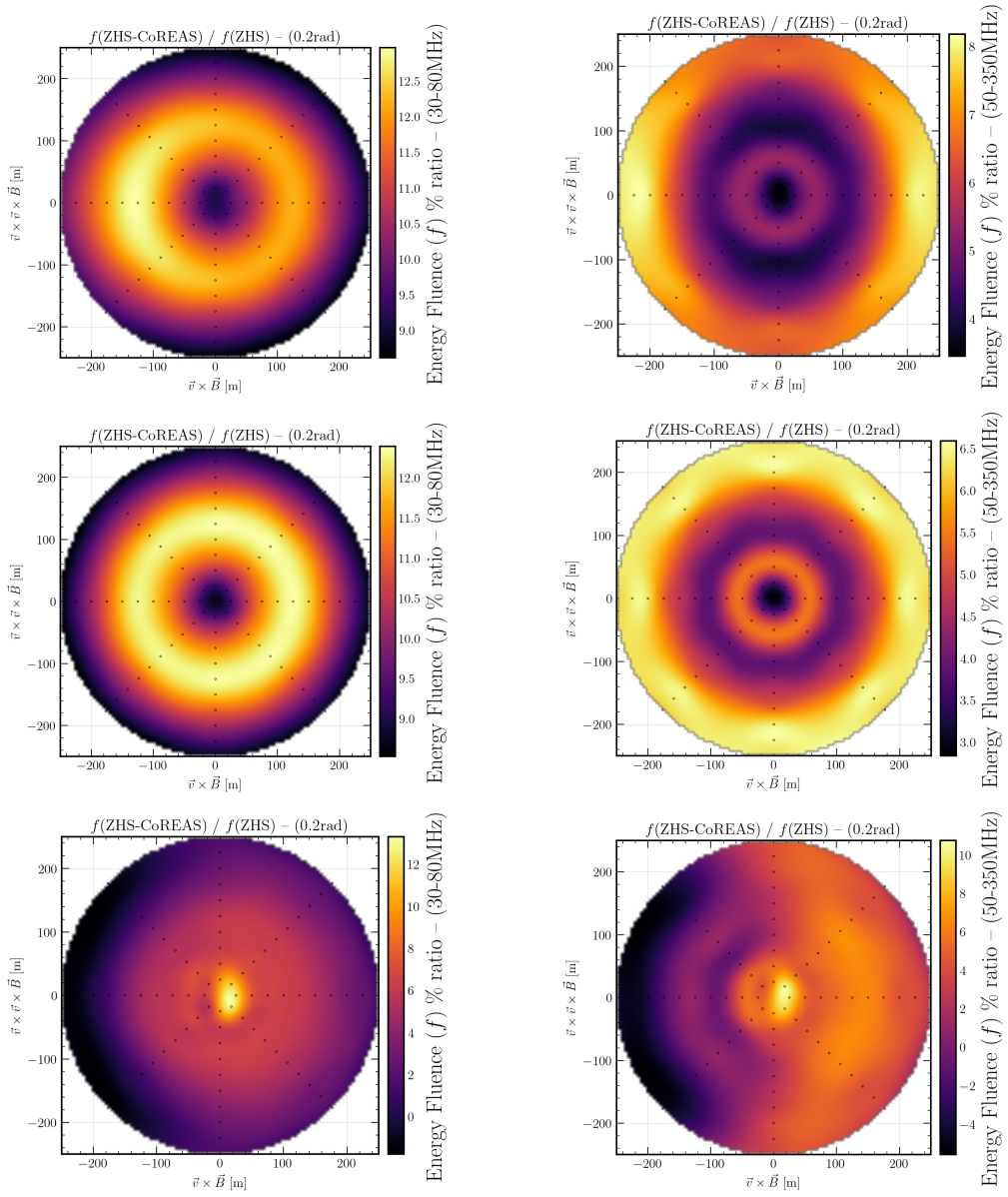


Figure 61: Ratio maps in fluence for 100, 1 PeV, iron induced vertical showers in the 30 MHz to 80 MHz (left column) and the 50 MHz to 350 MHz band (right column). In the first row, the ratio map in terms of total fluence is plotted, the second in terms of the geomagnetic contribution and the last row in terms of the charge excess. The maximally allowed deflection angle is set to 0.2 rad.

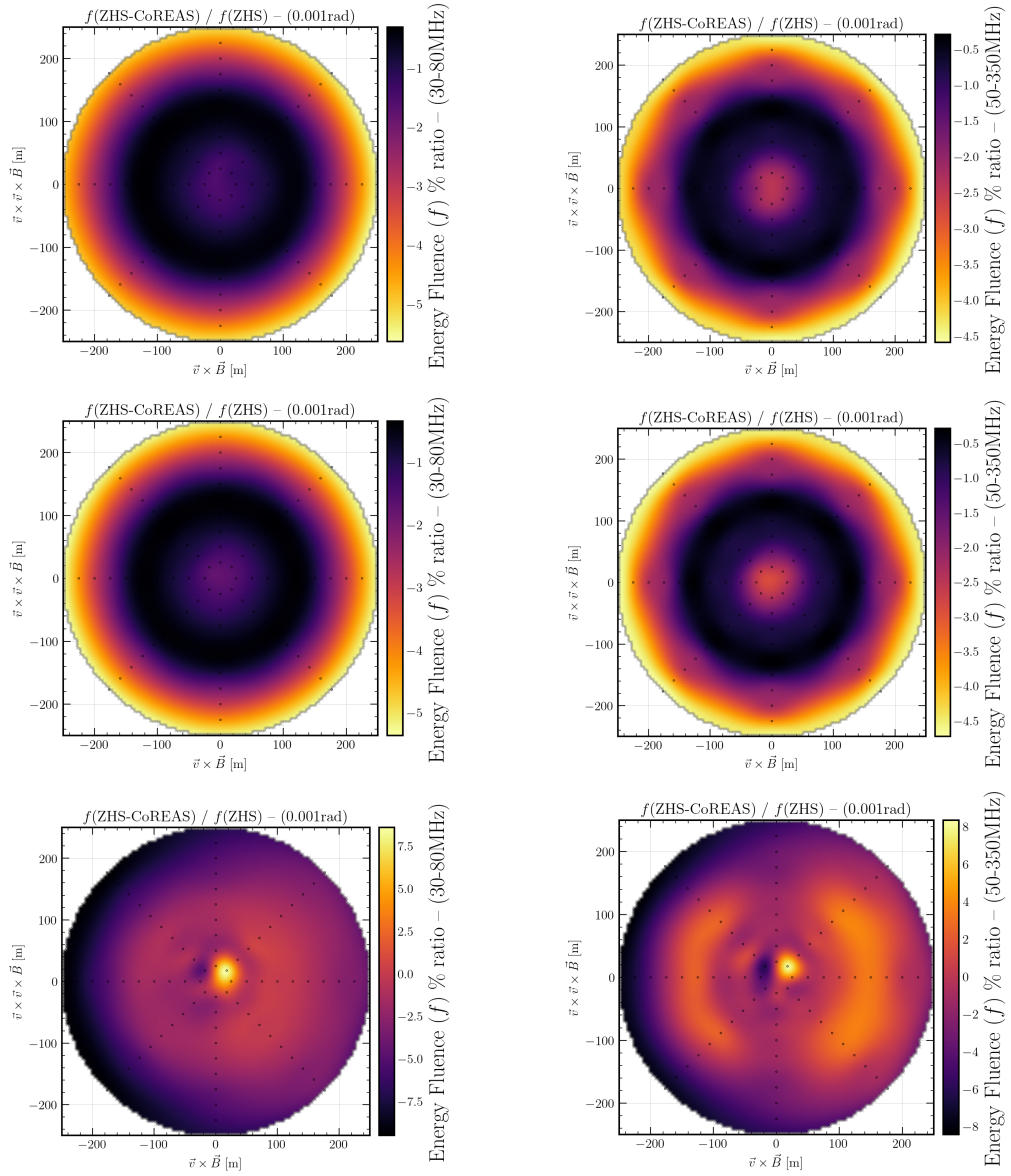


Figure 62: Ratio maps in fluence for 100, 1 PeV, iron induced vertical showers in the 30 MHz to 80 MHz (left column) and the 50 MHz to 350 MHz band (right column). In the first row, the ratio map in terms of total fluence is plotted, the second in terms of the geomagnetic contribution and the last row in terms of the charge excess. The maximally allowed deflection angle is set to 0.001 rad.

CoREAS vs ZHS – mean difference in terms of % (0.2 rad)		
Fluence	30 MHz to 80 MHz	50 MHz to 350 MHz
Total Fluence	11.1%	4.8%
Geomagnetic Contribution	11.1%	4.5%
Charge Excess	4.8%	3.2%

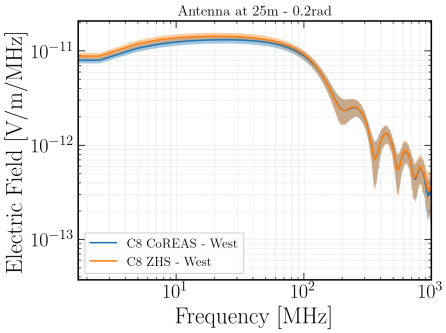
Table 5: The difference in in terms of % for fluence between CoREAS and ZHS for the case of 0.2 rad. ZHS predicts consistently more fluence with respect to CoREAS.

CoREAS vs ZHS – mean difference in terms of % (0.2 rad)		
Radiation Energy	30 MHz to 80 MHz	50 MHz to 350 MHz
Total Radiation Energy	11.3%	4.9%
Geomagnetic Contribution	11.2%	4.6%
Charge Excess	4.0%	3.1%

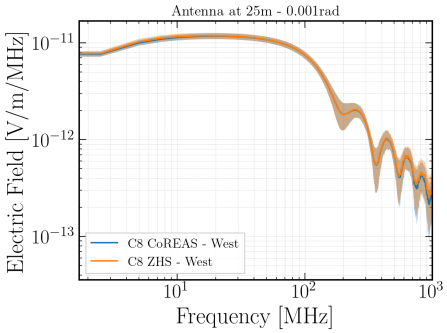
Table 6: The difference in in terms of % for radiation energy between CoREAS and ZHS for the case of 0.2 rad. ZHS predicts consistently more radiation energy with respect to CoREAS.

CoREAS vs ZHS – mean difference in terms of % (0.001 rad)		
Fluence	30 MHz to 80 MHz	50 MHz to 350 MHz
Total Fluence	-1.3%	-1.0%
Geomagnetic Contribution	-1.5%	-1.0%
Charge Excess	-2.3%	0.4%

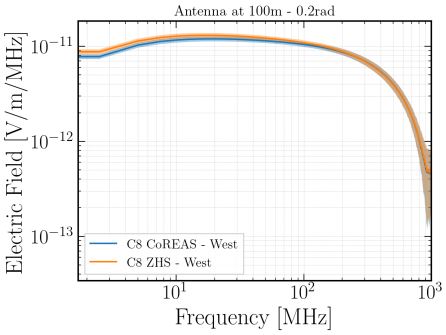
Table 7: The difference in terms of % for fluence between CoREAS and ZHS for the case of 0.001 rad. The minus signs before the % values indicate that CoREAS predicts more fluence with respect to ZHS. For 0.001rad both algorithms seem to converge.



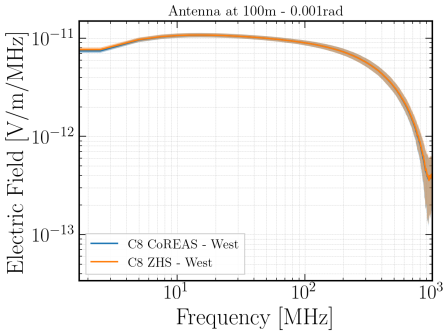
(a) Antenna at 25 m.



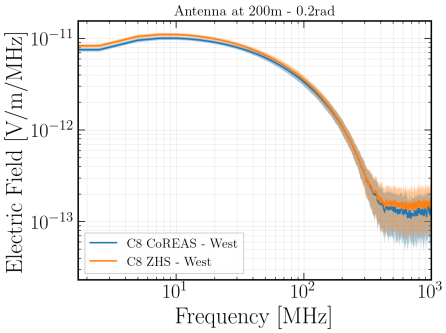
(b) Antenna at 25 m.



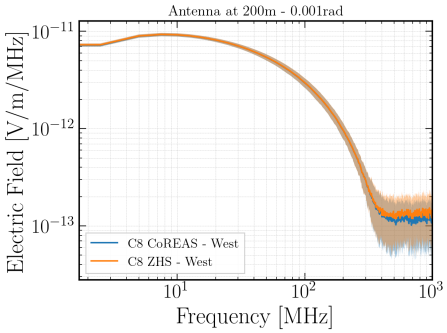
(c) Antenna at 100 m.



(d) Antenna at 100 m.



(e) Antenna at 200 m.



(f) Antenna at 200 m.

Figure 63: The mean value of 100 frequency spectra (dominant polarization) for different locations for both CoREAS and ZHS are plotted in the 0 GHz to 1 GHz band. The maximally allowed deflection angle is set to 0.2 rad for the left column plots and 0.001 rad for the right column plots.

CoREAS vs ZHS – mean difference in terms of % (0.001 rad)		
Radiation Energy	30 MHz to 80 MHz	50 MHz to 350 MHz
Total Radiation Energy	1.4%	-1.0%
Geomagnetic Contribution	-1.5%	-1.0%
Charge Excess	-2.8%	0.6%

Table 8: The difference in terms of % for radiation energy between CoREAS and ZHS for the case of 0.001 rad. The minus signs before the % values indicate that CoREAS predicts more radiation energy with respect to ZHS. For 0.001rad both algorithms seem to converge.

9.2.2 Photon induced air showers

In this section the agreement between the two algorithms is going to be inspected for photon induced vertical showers and different multiple scattering approximations. A library of a total of 200 CORSIKA 8 air shower simulations was produced. For all 200 showers, the primary particle chosen is a photon with an energy of 1 PeV impinging the atmospere vertically. Energy cuts were set for electromagnetic, muonic and hadronic components of the shower to 0.5 MeV for electromagnetic particles, and to 0.3 GeV for the rest. For high energies Sibyll 2.3d [14] was used, and for the low energy particles FLUKA [18] was chosen. No thinning was applied and the maxRad value of the tracking algorithm was set to the default value of 0.2 rad. The environment was configured to be the "US Standard Atmosphere", with a refractive index at sea level to $n = 1.000\,327$ and for higher altitudes scaling following the Gladstone-Dale law [75]. A constant geomagnetic field of 50 µT aligned in the x direction was set. For the antenna array a total of 96 time domain antennas were configured in a standard star-shaped pattern in 12 concentric rings spaced equally from 25 m to 300 m from the shower axis with 8 antennas distributed azimuthally in each ring. The sampling period of each antenna is set to 0.1 ns. The first point of interaction was fixed for all showers in order to avoid fluctuations as much as possible. The simulations were divided in 2 groups of 100 showers each. One group was configured with the Molière multiple scattering mechanism and the other with Highland. The same library is also going to be used in section 10.2 to inspect how the multiple scattering approach affects the fluence and radiation energy. The details of multiple scattering are going to be presented at that point.

The fluence ratio maps in 2 frequency bands for Highland are shown in fig. 64 and for Molière in fig. 65. Their mean differences per scattering mechanism for fluence are summarized in table 9 and table 11 respectively. Similarly for radiation energy in table 10 and table 12 The results of this comparison are extremely similar to table 5 and table 6. This means that the agreement between CoREAS and ZHS, for showers in the PeV range is not affected much neither by the primary particle, nor by the multiple scattering approximation utilized. The mean fluence maps of these showers can be found in the Appendix.

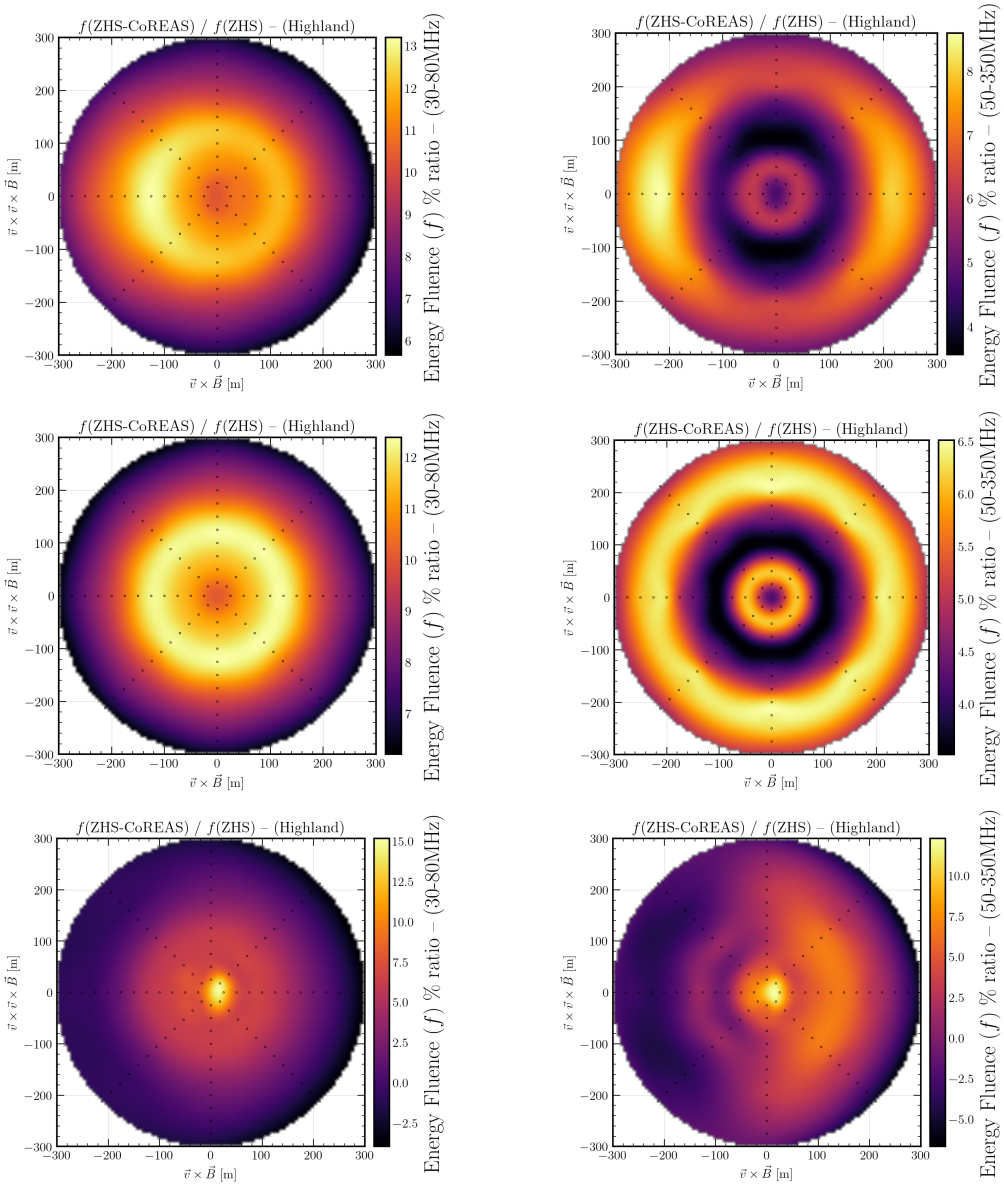


Figure 64: Ratio maps in fluence for 100, 1 PeV, photon induced vertical showers in the 30 MHz to 80 MHz (left column) and the 50 MHz to 350 MHz band (right column). In the first row, the ratio map in terms of total fluence is plotted, the second in terms of the geomagnetic contribution and the last row in terms of the charge excess. ZHS is directly compared with CoREAS for the Highland multiple scattering mechanism.

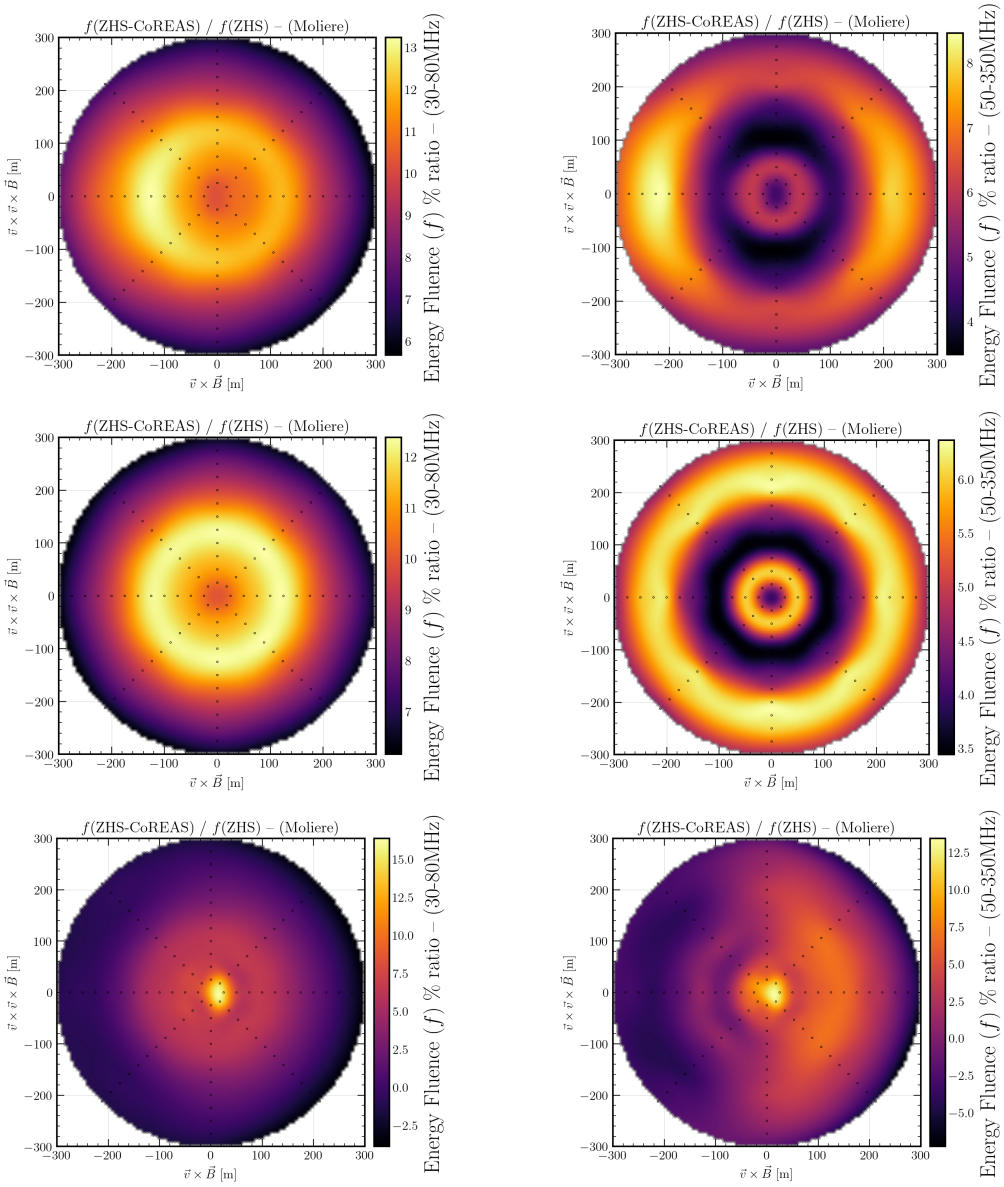


Figure 65: Ratio maps in fluence for 100, 1 PeV, photon induced vertical showers in the 30 MHz to 80 MHz (left column) and the 50 MHz to 350 MHz band (right column). In the first row, the ratio map in terms of total fluence is plotted, the second in terms of the geomagnetic contribution and the last row in terms of the charge excess. ZHS is directly compared with CoREAS for the Molière multiple scattering mechanism.

CoREAS vs ZHS – mean difference in terms of % (Highland)		
Fluence	30 MHz to 80 MHz	50 MHz to 350 MHz
Total Fluence	11.2%	5.0%
Geomagnetic Contribution	11.2%	4.6%
Charge Excess	4.9%	3.4%

Table 9: The difference in terms of % for fluence between CoREAS and ZHS for 100, 1 PeV photon induced vertical showers using the Highland multiple scattering approach. More fluence is simulated with ZHS.

CoREAS vs ZHS – mean difference in terms of % (Highland)		
Radiation Energy	30 MHz to 80 MHz	50 MHz to 350 MHz
Total Radiation Energy	11.1%	4.9%
Geomagnetic Contribution	11.2%	4.5%
Charge Excess	4.2%	3.2%

Table 10: The difference in terms of % for radiation energy between CoREAS and ZHS for 100, 1 PeV photon induced vertical showers using the Highland multiple scattering approach. More radiation energy is simulated with ZHS.

CoREAS vs ZHS – mean difference in terms of % (Molière)		
Fluence	30 MHz to 80 MHz	50 MHz to 350 MHz
Total Fluence	11.2%	4.9%
Geomagnetic Contribution	11.2%	4.6%
Charge Excess	4.0%	3.4%

Table 11: The difference in terms of % for fluence between CoREAS and ZHS for 100, 1 PeV photon induced vertical showers using the Molière multiple scattering approach. More fluence is simulated with ZHS.

CoREAS vs ZHS – mean difference in terms of % (Molière)		
Radiation Energy	30 MHz to 80 MHz	50 MHz to 350 MHz
Total Radiation Energy	11.1%	4.8%
Geomagnetic Contribution	11.2%	4.5%
Charge Excess	4.2%	3.1%

Table 12: The difference in terms of % for radiation energy between CoREAS and ZHS for 100, 1 PeV photon induced vertical showers using the Molière multiple scattering approach. More radiation energy is simulated with ZHS.

9.3 A "ZHS-like" approximation for CoREAS

The CoREAS algorithm can successfully calculate the radiation emitted in the form of radio waves from air showers utilizing the "Endpoints formalism". The contributions of the start and end of the particle track are added together to simulate the final radio pulse for each antenna. This algorithm falls short though for angles at the Cherenkov angle (θ_c) where the electric field calculation formula blows up because of the $(1 - n\beta cos\theta)^{-1}$ term. To avoid this pitfall, CoREAS introduces a threshold for the $(1 - n\beta cos\theta)^{-1}$ term. If for a particle track the term is evaluated above the threshold, the CoREAS formalism calculates the electric field using the middle point of the particle track, and not its start and endpoint. The emission is also scaled proportional to the track length. This is the way that ZHS calculates the vector potential, via the middle point of a track and hence considering the track as a whole. The threshold value for the "ZHS-like" approximation for air showers is set in CoREAS to 1000. If the medium that the radio waves are propagating through changes, the threshold value needs to be adjusted [77]. In this section, 8 identical 1 PeV iron induced vertical air showers for the 0.2 rad case are simulated using the same random number seed. This way it is ensured that these 8 showers are exactly the same. The only difference in these simulations, is that every time I set the ZHS-like fallback threshold to a different value. The purpose of this test, is to study the level of agreement between CoREAS and ZHS for different threshold values and whether the CoREAS formalism is stable. The same test is repeated with 8 more identical simulations for the 0.001 rad case.

In fig. 66 and fig. 67 the agreement between CoREAS and ZHS for different fallback thresholds is shown as a function of distance from the shower core. Each data point is the mean value of 8 antennas placed on the specified radial distance distributed azimuthally in steps of 45^o. The error bars per data point defined as the standard deviation of the mean value of 8 antennas divided by 8, are also shown. Smaller threshold values translate to larger angles around the Cherenkov angle where the "ZHS-like" approximation takes place, while a larger threshold means the opposite. For the 0.2 rad case (fig. 66), there are some interesting observations. First, it can be seen that for larger values of the threshold the "ZHS-like" approximation is being used less often because it is used for smaller angles around θ_c, which is most likely why the 2 algorithms diverge more. For smaller threshold values the 2 algorithms coincide further. The second observation is the fact that close to the Cherenkov ring (100 m to 150 m), where the "ZHS-like" approximation takes place for 30 MHz to 80 MHz the 2 algorithms present their higher deviation, which is surprising. For 50 MHz to 350 MHz though, this is not the case, and the 2 algorithms converge around the Cherenkov ring.

For the 0.001 rad case (fig. 67) the observations are different. For larger threshold values the 2 algorithms converge, which is the opposite behavior to the 0.2 rad case. Also, on both frequency bands the highest level of agreement is observed on the Cherenkov ring and the region close to it. At the time of writing, and given the youthful state of CORSIKA 8, I cannot give a definitive answer whether this contrast between the 0.2 and the 0.001 rad case is physics or some shortcoming of the current tracking algorithm in use in CORSIKA 8. As a final remark for CoREAS, the variation of the threshold value between $0.5 * 1000 \leq (1 - n\beta cos\theta)^{-1} \leq 2 * 1000$ is

less than 1%. For the $0.25 * 1000$ case, the deviation can vary up to $\sim 4\%$. Hence, the CoREAS formalism does not diverge much when altering the fallback threshold value.

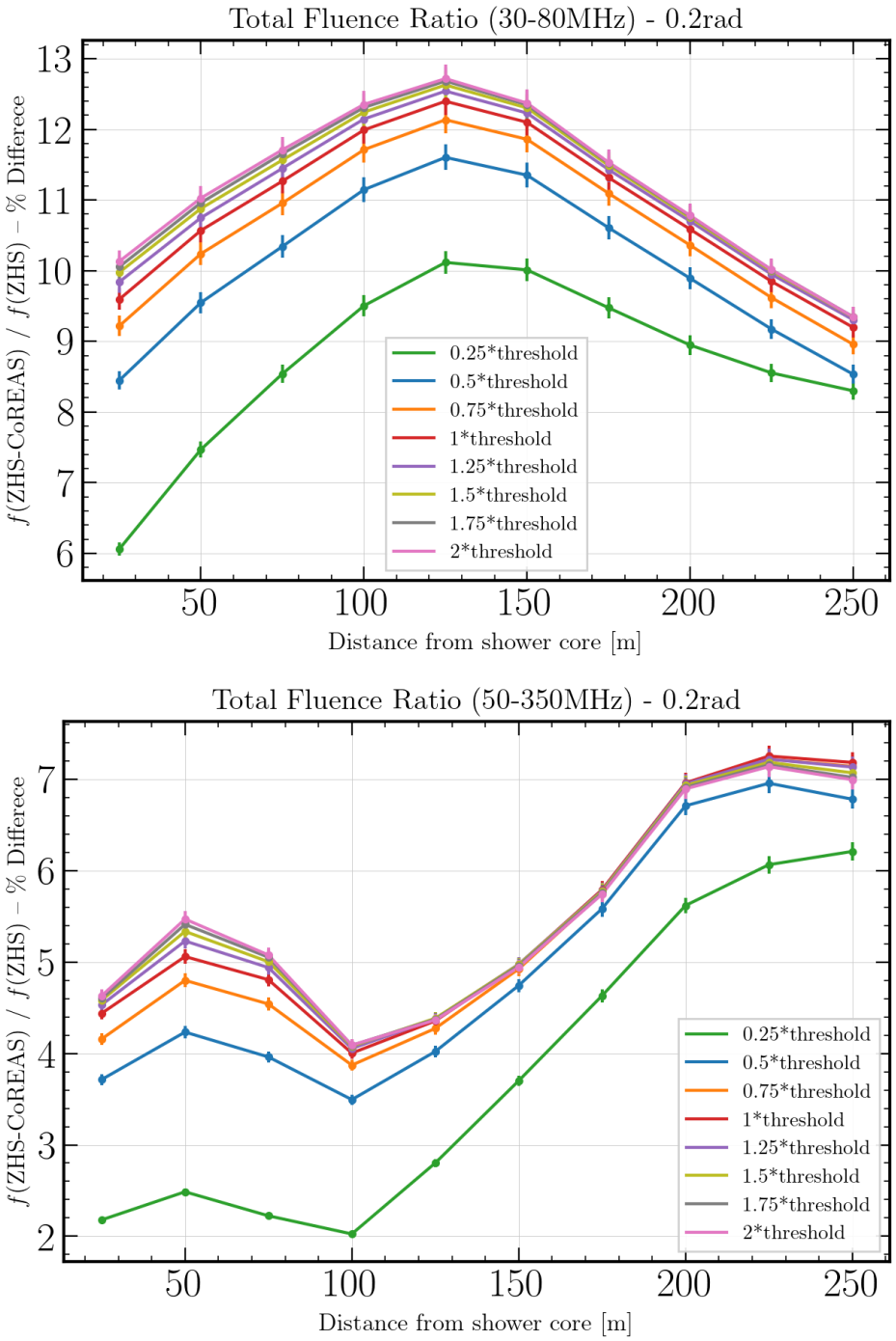


Figure 66: CoREAS versus ZHS for different threshold values of the CoREAS "ZHS-like" approximation for a maxRad value of 0.2 rad. 2 frequency bands are studied, the 30 MHz to 80 MHz (top) and the 50 MHz to 350 MHz (bottom). $1 * \text{threshold}$ refers to the value of 1000.

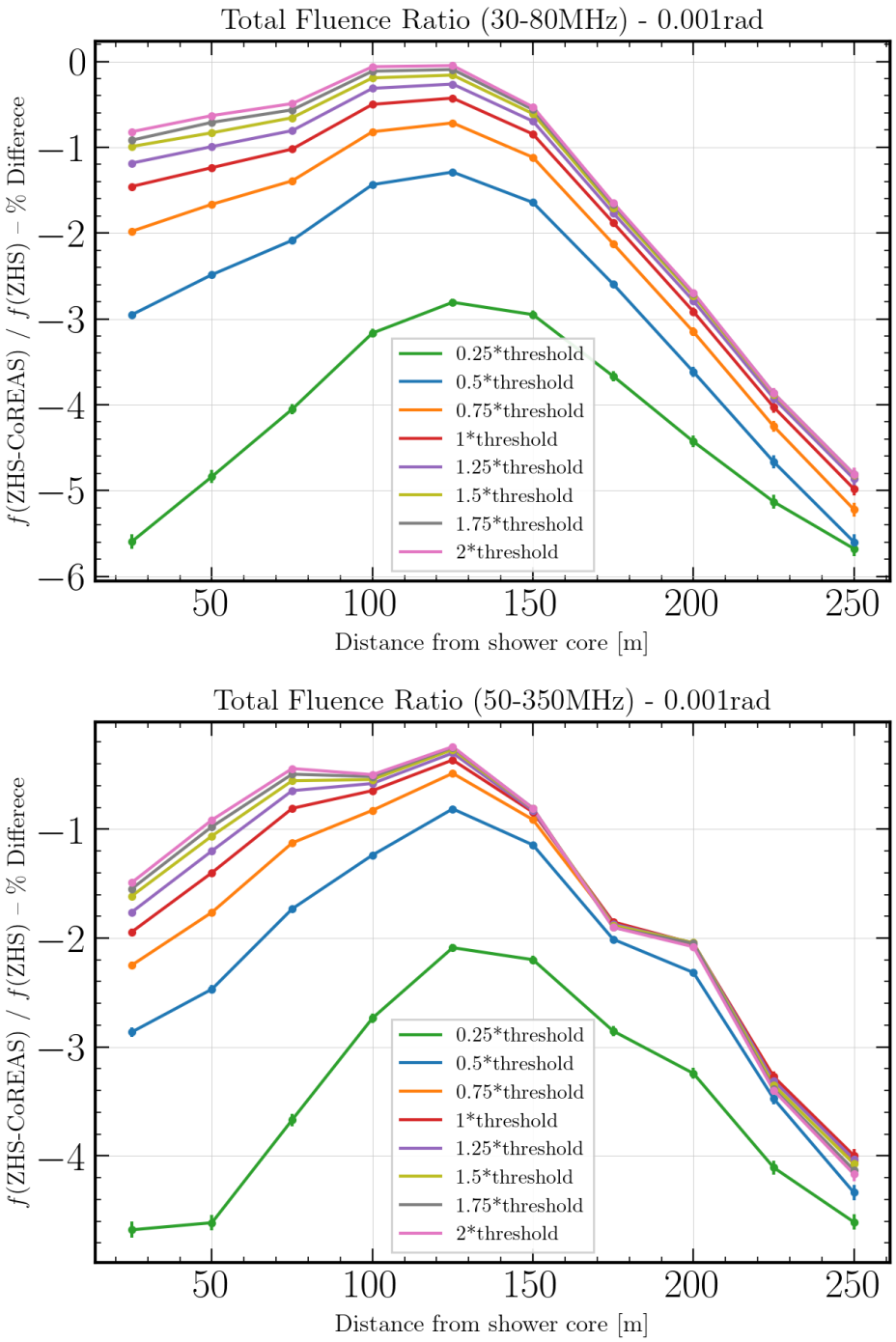


Figure 67: CoREAS versus ZHS for different threshold values of the CoREAS "ZHS-like" approximation for a maxRad value of 0.001 rad. 2 frequency bands are studied, the 30 MHz to 80 MHz (top) and the 50 MHz to 350 MHz (bottom). $1 * \text{threshold}$ refers to the value of 1000.

9.4 Discussion

In this section an extensive comparison of the CoREAS versus ZHS formalisms for air showers was presented. It was found that their level of agreement is not dependent on the primary particle, nor the multiple scattering approximation. It was also found that for very low energy showers the level of agreement gets slightly worse, but for higher energies (1 PeV and above) the level of agreement in the two frequency bands that were studied is not affected. Additionally, the effect of the track length of simulated particle tracks with CORSIKA 8 was inspected. It was found that for larger tracks (default tracking algorithm value) in the 30 MHz to 80 MHz band, ZHS predicts 11.3% more radiation energy than CoREAS. In the 50 MHz to 350 MHz band, ZHS predicts 4.9% more radiation energy as well. However, when one uses very small, fine tracks the two formalisms practically converge. CoREAS now predicts 1.4% more radiation energy than ZHS in the 30 MHz to 80 MHz band, and in the 50 MHz to 350 MHz band CoREAS predicts less than 1% more radiation energy than ZHS. Such a result gives the radio community some useful insights. First, it most likely indicates that to properly predict the radio emission one needs to have a very fine track length on the simulated particle tracks. This is justified by the fact that the two formalisms practically converge at this limit. The second and most important insight, is the fact that the two formalisms agree to a very large extent indeed. They have been both derived from first principles, under different assumptions, but the fact that they agree to that extent solidifies the argument that the radio emission from air showers is a well understood process that can be used to set the energy scale of different cosmic ray experiments, mass composition analysis and energy reconstruction.

10 Relevance of simulation details on energy and mass reconstruction

10.1 The effect of different track lengths on radio simulations

Comparing the 2 algorithms has revealed a dependency on the maximally allowed magnetic deflection angle (and effectively the track length). For very small maxRad values, and hence very small tracks, the 2 algorithms converge. The question arises on what is the impact of the track length on the shower simulations in general, not only as a factor of agreement between CoREAS and ZHS. Namely, when setting maxRad to a very small value like 0.001 rad for example, is more radiation energy being predicted with respect the default maxRad value? How is the fluence predicted for both maxRad cases? These are very important questions as their answers are tightly connected to energy and $X_{\max}$ reconstruction. Using the fluence footprint information one can reconstruct $X_{\max}$ as discussed in section 3.7 which is an indicator of the mass of the primary particle. Total radiation energy is also a crucial quantity to determine the energy of the electromagnetic cascade of particles, to which radiation energy scales quadratically.

Reflecting on section 9.1, two maxRad values are picked to answer the questions above, namely 0.2 rad and 0.001 rad. In section 9.2 a library of 200, 1 PeV iron induced vertical showers was produced for the systematic CoREAS vs ZHS comparison. 100 of them were set with maxRad equal to 0.2 rad and 100 with maxRad set to 0.001 rad. In this section a comparison is going to be made in a different way. Instead of comparing CoREAS with ZHS, the showers produced with 0.2 rad are going to be directly compared with the ones that were produced with the 0.001 rad setting. Changes in fluence, longitudinal profiles, electromagnetic and radiation energy are going to be inspected. To ensure that such a deviation in radio simulations for different maxRad values is not only a feature of CORSIKA 8, a library of simulations using CORSIKA 7 is also produced. The setting for the CORSIKA 7 simulations is identical with CORSIKA 8 ones. 100, 1 PeV iron induced vertical showers for 0.2 rad and 100 more for 0.001 rad are simulated. For CORSIKA 8 the effect of different track lengths is going to be studied for CoREAS and ZHS, while for CORSIKA 7 it is going to be studied for CoREAS.

The strategy to compare the 0.2 rad simulations with the 0.001 rad ones is the same as before. The same formula to compare the fluence as eq. (35) is used, but in this case it is expressed like so:

$$D_i = \frac{\overline{f^i_{0.2\mathrm{rad}}} - \overline{f^i_{0.001\mathrm{rad}}}}{\overline{f^i_{0.2\mathrm{rad}}}} \tag{36}$$

The total mean value of the difference over the whole footprint ($\overline{D}$) and weights (w_i) are defined in an identical manner as eq. (32) and eq. (33) respectively. The radio emission is highly dependent on the electromagnetic part of the shower. For this reason, it is important to understand whether the deviations observed in the radio simulations are because of changes on

the electromagnetic component of the shower not. It could be that when changing the maxRad values the electromagnetic particle content is influenced and the radio calculation simply follows. To count this possibility out, for both maxRad cases the electromagnetic energy released in the atmosphere and on the ground is calculated. For all the showers per maxRad, the radiation energy is normalized over the square of the electromagnetic energy deposit. So, in this section, radiation energy refers to the normalized radiation energy over the electromagnetic energy deposit.

10.1.1 CORSIKA 8

In fig. 68 the longitudinal profiles of electrons and positrons for the 0.2 rad and the 0.001 rad cases are plotted. For the 0.2 rad case there are $\sim 2.7\%$ more electrons, 1.6% more positrons and in total $\sim 2.3\%$ more electrons and positrons at the shower maximum. Since more particles are simulated for 0.2 rad, it is expected that the radio emission will follow this trend and radiate more power in that case. $\sim 0.4\%$ more electromagnetic energy is being deposited to the atmosphere for the 0.2 rad simulations. To get the energy deposit of electromagnetic particles for the CORSIKA 8 simulations, Merge Request 544 was utilized, since at the time of writing the existing `EnergyLossWriter` on master branch was only tracking down the energy deposit of all particles together. The fluence ratio maps for CoREAS are shown in fig. 69 and for ZHS in fig. 70 and the differences are summarized in table 13 and table 15 respectively.

For CORSIKA 8 CoREAS the overall deviation can be indicated by the difference in the normalized radiation energy, which is $\sim 23.1\%$ more for the 30 MHz to 80 MHz band and $\sim 28.2\%$ for the 50 MHz to 350 MHz band in case of the 0.2 rad maxRad. It is shown in fig. 69 that the difference is relatively consistent throughout the whole footprint, with the largest deviations occurring in areas where the signal is stronger. CORSIKA 8 ZHS behaves in a similar manner as can be seen in fig. 70. In this case the deviations are larger than CoREAS, but still favor the 0.2 rad case. Normalized radiation energy is $\sim 32.8\%$ more for 30 MHz to 80 MHz and $\sim 32.4\%$ more for 50 MHz to 350 MHz for 0.2 rad. The differences of CORSIKA 8 CoREAS in fluence are summarized in table 13 and in radiation energy in table 14. For CORSIKA 8 ZHS, the differences in fluence are summarized in table 15 and in radiation energy in table 16. For both algorithms these are rather large deviations that if true could in principle introduce a large systematic error on radio analyses, since they all rely heavily on simulations. Deviations in radiation energy directly affect the energy reconstruction accuracy, since a difference of e.g. $\sim 20\%$ in radiation energy would mean a difference of $\sim 10\%$ on energy reconstruction. Since CORSIKA 8 is still a software package under development, and such a large deviation could be potentially attributed to some bug or inconsistency of the magnetic field tracking algorithm, it is imperative to perform the same test with the well established CORSIKA 7.

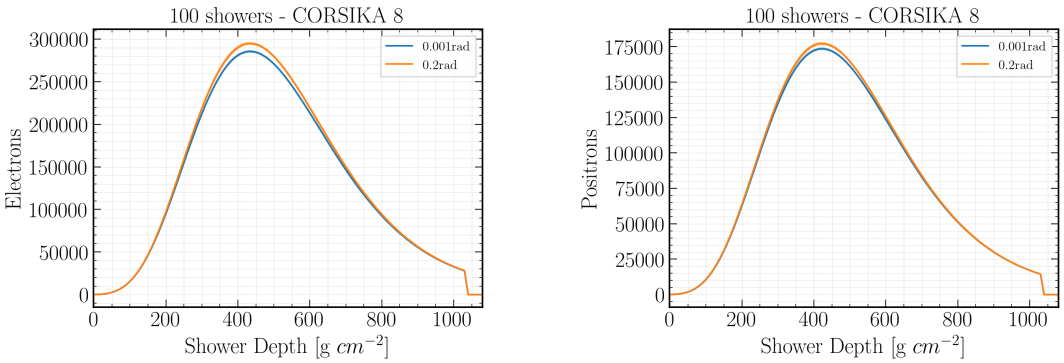


Figure 68: The longitudinal profiles over 100 showers per maxRad value of electrons (left) and positrons (right).

0.2rad vs 0.001rad – mean difference in terms of % (C8 CoREAS)		
Fluence	30 MHz to 80 MHz	50 MHz to 350 MHz
Total Fluence	23.6%	28.3%
Geomagnetic Contribution	23.2%	28.4%
Charge Excess	26.6%	21.3%

Table 13: The difference in terms of % for fluence between 0.2 and 0.001 rad for CORSIKA 8 CoREAS. More fluence is emitted for the 0.2 rad case.

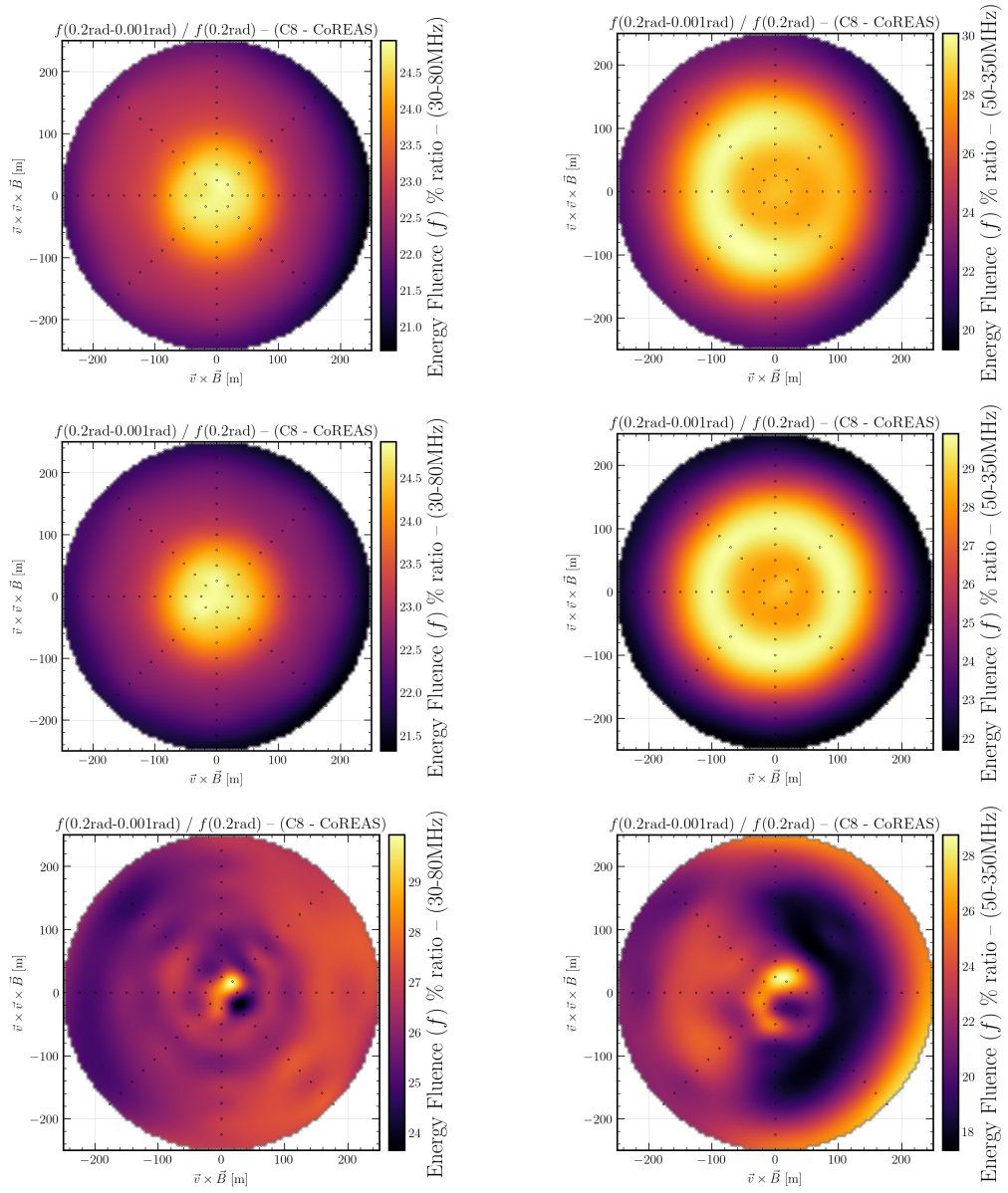


Figure 69: Ratio maps in fluence for 100, 1 PeV, iron induced vertical showers in the 30 MHz to 80 MHz (left column) and the 50 MHz to 350 MHz band (right column). In the first row, the ratio map in terms of total fluence is plotted, the second in terms of the geomagnetic contribution and the last row in terms of the charge excess. The effect of the maximally allowed deflection angle is studied for 0.2 over 0.001 rad for CoREAS.

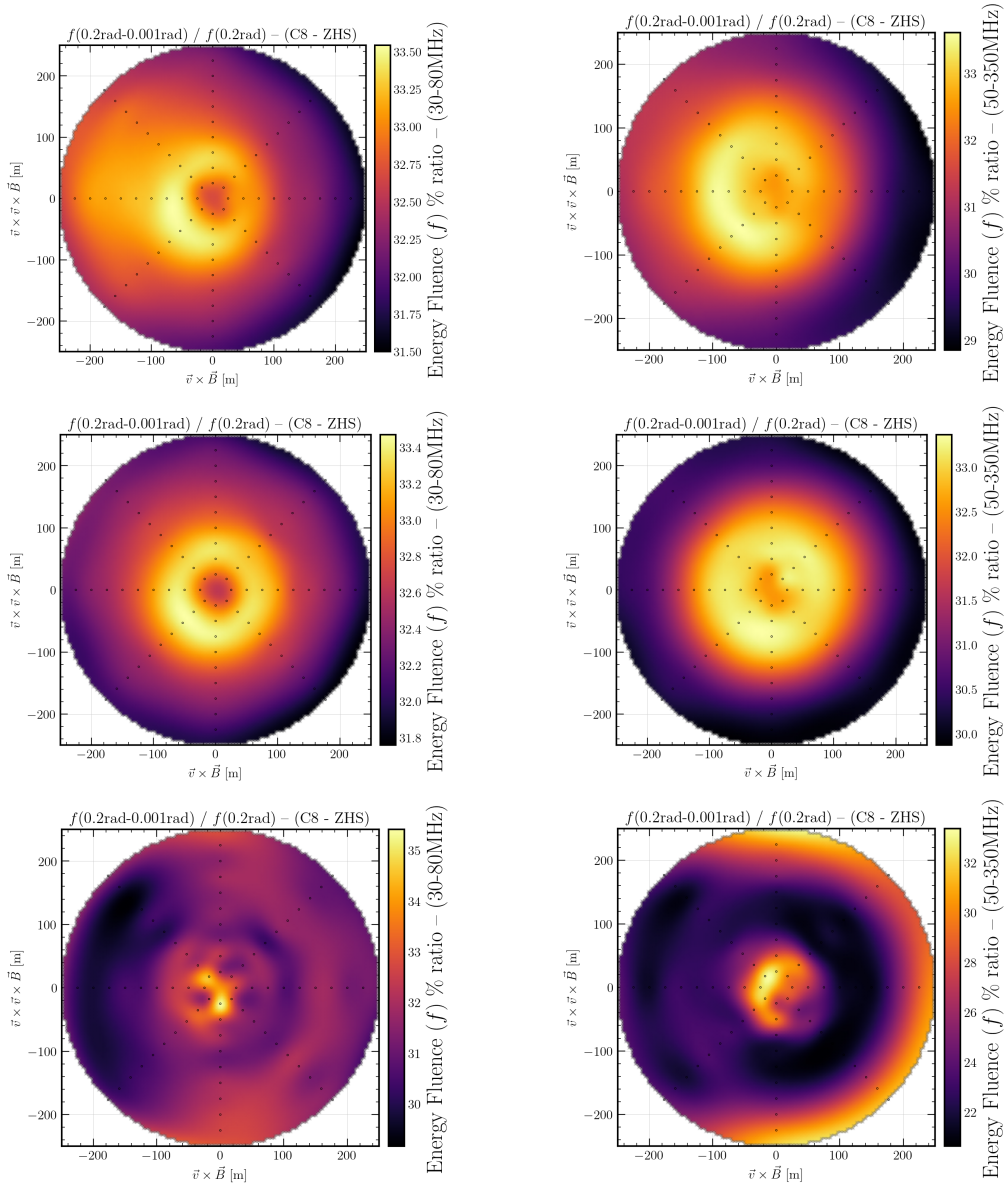


Figure 70: Ratio maps in fluence for 100, 1 PeV, iron induced vertical showers in the 30 MHz to 80 MHz (left column) and the 50 MHz to 350 MHz band (right column). In the first row, the ratio map in terms of total fluence is plotted, the second in terms of the geomagnetic contribution and the last row in terms of the charge excess. The effect of the maximally allowed deflection angle is studied for 0.2 over 0.001 rad for ZHS.

0.2rad vs 0.001rad – mean difference in terms of % (C8 CoREAS)		
Radiation Energy	30 MHz to 80 MHz	50 MHz to 350 MHz
Total Radiation Energy	22.7%	27.7%
Geomagnetic Contribution	22.6%	27.8%
Charge Excess	25.8%	20.6%

Table 14: The difference in terms of % for radiation energy between 0.2 and 0.001 rad for CORSIKA 8 CoREAS. More radiation energy is emitted for the 0.2 rad case.

0.2rad vs 0.001rad – mean difference in terms of % (C8 ZHS)		
Fluence	30 MHz to 80 MHz	50 MHz to 350 MHz
Total Fluence	32.9%	32.5%
Geomagnetic Contribution	32.7%	32.4%
Charge Excess	31.4%	23.5%

Table 15: The difference in terms of % for fluence between 0.2 and 0.001 rad for CORSIKA 8 ZHS. More fluence is emitted for the 0.2 rad case.

0.2rad vs 0.001rad – mean difference in terms of % (C8 ZHS)		
Radiation Energy	30 MHz to 80 MHz	50 MHz to 350 MHz
Total Radiation Energy	32.4%	32.0%
Geomagnetic Contribution	32.2%	31.9%
Charge Excess	30.7%	22.5%

Table 16: The difference in terms of % for radiation energy between 0.2 and 0.001 rad for CORSIKA 8 ZHS. More radiation energy is emitted for the 0.2 rad case.

10.1.2 CORSIKA 7

An identical setup with the CORSIKA 8 one is simulated using CORSIKA 7. Starting right away with the comparison, one can already notice a strange behavior. The longitudinal profiles of electrons and positrons (fig. 71) behave the opposite way than CORSIKA 8. For CORSIKA 7 the 0.001 rad case simulates consistently more particles than the 0.2 rad one. In more detail, for 0.001 rad $\sim 3.3\%$ more electrons and $\sim 1.4\%$ more positrons are simulated at the shower maximum. Summing them up, $\sim 2.61\%$ more electrons and positrons are predicted for the 0.001 rad case. The energy deposit of the electromagnetic part of the shower is accessed through the `DATXXX.long` files by summing up the gammas, em cut and em ioniz columns. For 0.001 rad, $\sim 0.5\%$ more electromagnetic energy is being deposited to the atmosphere and on the ground than the 0.2 rad case.

The minus sign in the ratio maps (fig. 72) and the summary tables table 17 and table 18 indicates that that in all scenarios 0.001 rad produces more fluence and more radiation energy than 0.2 rad. Because of the minus sign, notice that in the ratio maps the colormap is inverted, meaning that in the darker areas is where the deviation is less. The pattern is obvious, in the regions where the signal is strong the deviation is smaller, while for the weak signal regions, the deviation gets larger. The fluence maps in both frequency bands and both maxRad values under inspection, for total fluence, geomagnetic emission and charge excess can be found on the Appendix. Looking into the normalized radiation energy difference, for 30 MHz to 80 MHz $\sim 11.5\%$, and for 50 MHz to 350 MHz $\sim 9.1\%$ more radiation energy is simulated with the 0.001 rad setting. The 55 MHz to 65 MHz was also inspected given that the LOFAR antennas operate in that band. The same behavior was found. This result is not surprising and is actually consistent to previous work [52]. There the dependency of the radiation energy on the "STEPFC" CORSIKA7 parameter was studied. This parameter is responsible of altering the electron multiple scattering length factor, and hence, when set to smaller values, effectively producing finer and smaller tracks. The maxRad setting that it is studied here has the same effect, so one can conclude that radio simulations are indeed affected by the track length of the simulated particle tracks. This is observed in both CORSIKA 7 and CORSIKA 8, with the CORSIKA 8 case producing larger deviations though. It is worth pointing out once more that CORSIKA 7 and CORSIKA 8 do not use neither the same magnetic field tracking algorithm nor the same electromagnetic interaction models. Furthermore, one could argue that since for very small tracks the CoREAS and ZHS formalisms agree to up to $\sim 2\%$ (section 9.2) and the CORSIKA 7 comparison presented here is consistent with [52], simulating the radio emission utilizing very small, fine tracks represents the ground truth. It is known that there is a difference between the $X_{\max}$ measurements of LOFAR and AERA. For their analyses, the default value of 0.2 rad has been used. Although, computationally heavy to redo the simulations used for their analysis (up to a factor of 5) it might prove fruitful to do so.

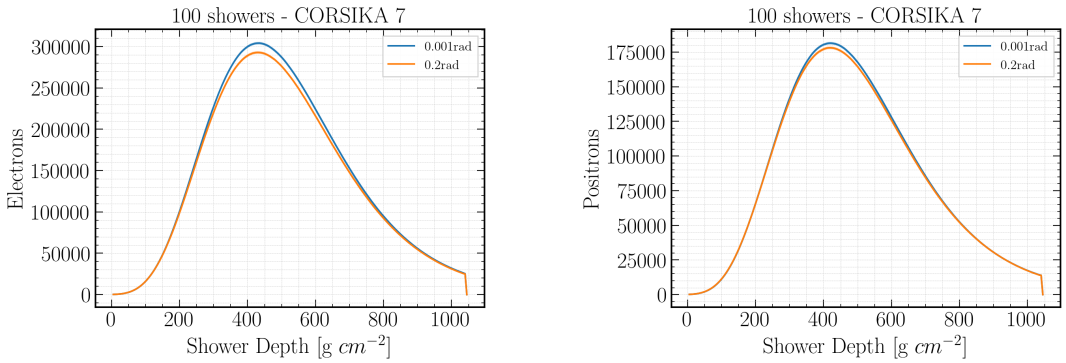


Figure 71: The longitudinal profiles over 100 showers per maxRad value of electrons (left) and positrons (right).

0.2rad vs 0.001rad – mean difference in terms of % (C7 CoREAS)		
Fluence	30 MHz to 80 MHz	50 MHz to 350 MHz
Total Fluence	-12.4%	-10.5%
Geomagnetic Contribution	-12.4%	-10.3%
Charge Excess	-11.5%	-12.3%

Table 17: The difference in terms of % for fluence between 0.2 and 0.001 rad for CORSIKA 8 ZHS. The minus sign indicates that more fluence is emitted for the 0.001 rad case.

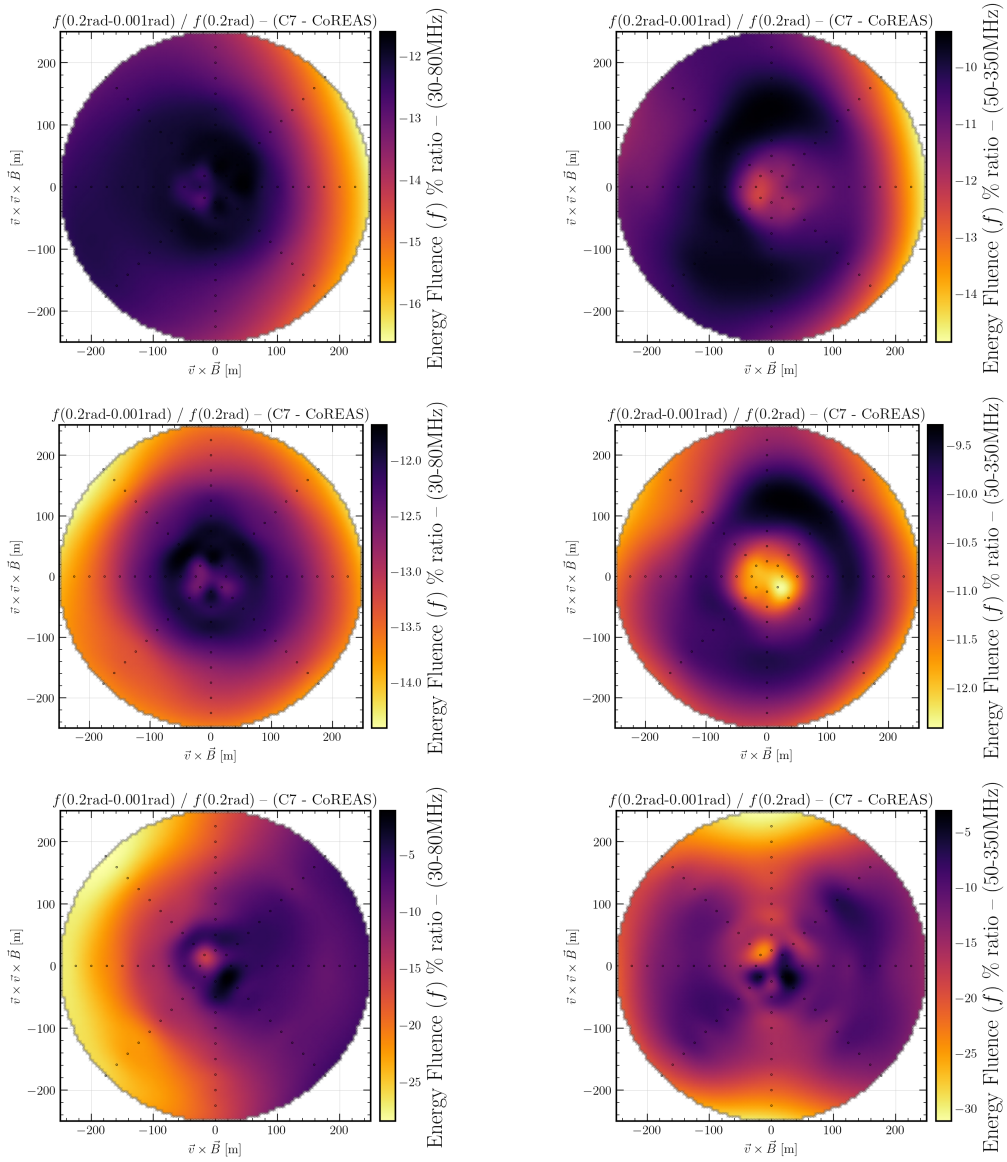


Figure 72: Ratio maps in fluence for 100, 1 PeV, iron induced vertical showers in the 30 MHz to 80 MHz (left column) and the 50 MHz to 350 MHz band (right column). In the first row, the ratio map in terms of total fluence is plotted, the second in terms of the geomagnetic contribution and the last row in terms of the charge excess. The effect of the maximally allowed deflection angle is studied for 0.2 over 0.001 rad for CORSIKA 7 CoREAS.

0.2rad vs 0.001rad – mean difference in terms of % (C7 CoREAS)		
Radiation Energy	30 MHz to 80 MHz	50 MHz to 350 MHz
Total Radiation Energy	-11.5%	-9.1%
Geomagnetic Contribution	-13.7%	-11.3%
Charge Excess	-13.8%	-13.3%

Table 18: The difference in terms of % for radiation energy between 0.2 and 0.001 rad for CORSIKA 8 ZHS. The minus sign indicates that more radiation energy is emitted for the 0.001 rad case.

10.2 The effect of different multiple scattering approaches in radio simulations

As charged particles in the shower move along the atmosphere they interact with atomic nuclei elastically with minimal energy losses. However, the momentum direction of these charged particles is affected and can severely influence the lateral spread of the shower. The effect of these elastic interactions is often referred to as multiple scattering. Different multiple scattering calculations are integrated in CORSIKA 8 within the electromagnetic interaction model that is used, PROPOSAL. Although, at the time of writing there is no user friendly interface to change the default multiple scattering calculation, it is quite easy to do so in the corresponding class within `ContinuousProcess.inl` of PROPOSAL. The available multiple scattering approximations have been compared for electrons in air with density of $1.205\,\mathrm{g\,cm^{-3}}$ and can be seen in fig. 73 [76]. It is shown that Highland (and Highland integral) are accurate descriptions for small scattering angles, while Molière (and Molière interpolation) extends to larger angles. It has also been shown in [76] that this has an effect on the lateral development of the showers, where a library of photon showers simulated with CORSIKA 8 was studied. The showers generated with the Highland multiple scattering approximation show a shift towards the shower axis, meaning that the charged particles scatter less and hence the shower spreads less, laterally. The purpose of this section is to inspect how these differences affect the radio simulations. To achieve this, the library of a total of 200 CORSIKA 8 air shower simulations presented in section 9.2.2 was used. The effect of different multiple scattering approximations is studied for the fluence footprint and the radiation energy.

Comparing Highland with Molière and their effect on radio simulations is being done the same way as all the comparisons in the latest sections. For the fluence of all showers ($\overline{f}$) comparison, eq. (35) is altered in terms of notation accordingly:

$$D_i = \frac{\overline{f^i_{\mathrm{Highland}}} - \overline{f^i_{\mathrm{Molière}}}}{\overline{f^i_{\mathrm{Highland}}}} \tag{37}$$

A positive value signifies that Highland scattering emits more fluence, while negative

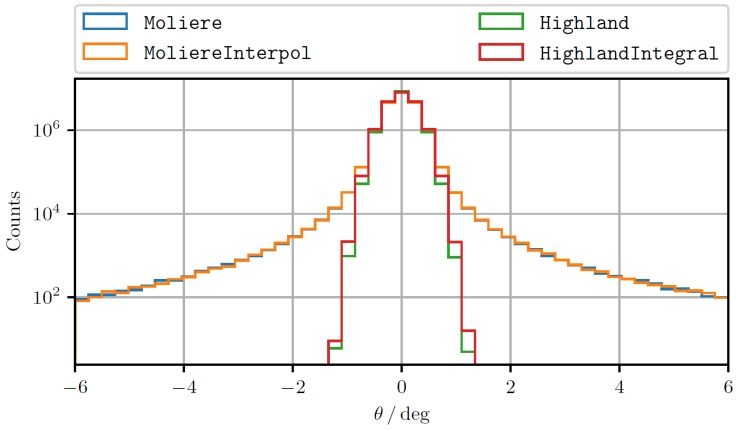


Figure 73: Molière, MolièreInterpol, Highland and HighlandIntegral approximations which are available through PROPOSAL are used to sample scattering angles θ of electrons in air at standard density. An initial energy of $E_i = 1\,\text{GeV}$ and a final energy of $E_f = 0.9\,\text{GeV}$, corresponds to a grammage of $X \simeq 3.52\,\text{g}\,\text{cm}^{-2}$, or a distance in air of $x \simeq 39.22\,\text{m}$. The Molière (and Molière interpolation) distribution extends to larger angles and can accurately describe them when compared to Highland (and HighlandIntegral). Figure taken from [76].

values indicate that Molière emits more fluence. The mean value of the difference ($\overline{D}$) and weights (w_i) are defined in the same way as eq. (32) and eq. (33) respectively. Radiation energy in this section refers to radiation energy normalized over the square of the of electromagnetic energy deposit in the atmosphere and ground. For the radio simulations comparisons both formalisms, CoREAS and ZHS are studied.

Starting the comparison from the longitudinal profiles in fig. 74, the showers simulated with Highland scattering produce $\sim 2.9\%$ more electrons and $\sim 1.9\%$ more positrons at shower maximum. In total, $\sim 2.5\%$ more electrons and positrons are produced with Highland scattering at shower maximum. When inspecting the fluence ratio maps, CoREAS (fig. 75) and ZHS (fig. 76) reveal a very similar trend. For both frequency bands, total fluence and geomagnetic emission deviate more close to the shower core and in the areas where the signal is strong. For charge excess, the deviation is less when compared to the geomagnetic emission, but again the larger deviations are observed in the area where the signal is stronger. All the differences in terms of percentage in fluence are summarized for CoREAS in table 19 and for ZHS in table 21. Similarly, the differences in total radiation energy are summarized for CoREAS in table 20 and for ZHS in table 22. As an overall indication of the deviations, it is useful to highlight the deviations in normalized radiation energy. For both algorithms, at 30 MHz to 80 MHz, Highland scattering predicts roughly 22.8% more radiation energy and roughly 21.7% at 50 MHz to 350 MHz. For CORSIKA 8 the default multiple scattering approximation used is Molière and this choice produces results closer to CORSIKA 7 simulations. Since it is established in [76] that showers with Highland scattering do spread less on the lateral dimension, it is safe to observe that these showers produce more coherent radio pulses and hence the difference in energy between the 2 multiple scattering mechanisms.

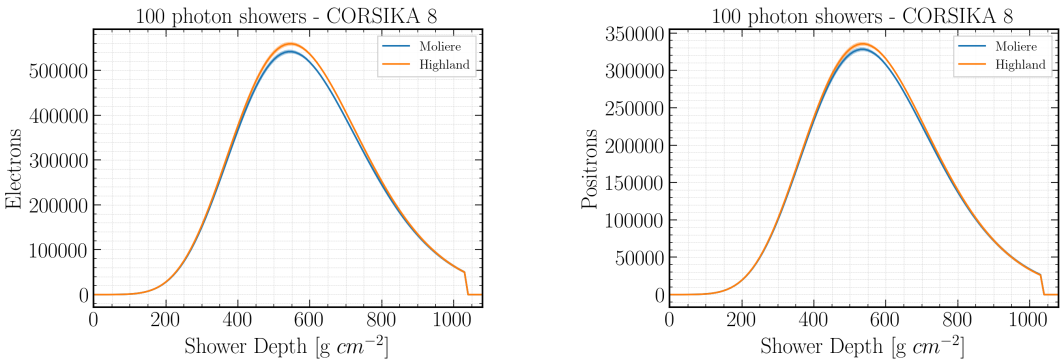


Figure 74: The longitudinal profiles over 100 photon induced vertical air showers per multiple scattering mechanism of electrons (left) and positrons (right).

The fluence maps of both formalisms, both multiple scattering mechanisms in 2 frequency bands (30 MHz to 80 MHz and 50 MHz to 350 MHz) can be found on the Appendix.

Highland vs Molière – mean difference in terms of % (C8 CoREAS)		
Fluence	30 MHz to 80 MHz	50 MHz to 350 MHz
Total Fluence	23.3%	22.3%
Geomagnetic Contribution	22.9%	21.9%
Charge Excess	19.9%	21.9%

Table 19: The difference in terms of % for fluence between Highland and Molière multiple scattering mechanisms for CORSIKA 8 CoREAS. More fluence is emitted for the case of Highland scattering.

Highland vs Molière – mean difference in terms of % (C8 CoREAS)		
Radiation Energy	30 MHz to 80 MHz	50 MHz to 350 MHz
Total Radiation Energy	22.7%	21.6%
Geomagnetic Contribution	22.9%	21.7%
Charge Excess	19.9%	21.8%

Table 20: The difference in terms of % for radiation energy between Highland and Molière multiple scattering mechanisms for CORSIKA 8 CoREAS. More radiation energy is emitted for the case of Highland scattering.

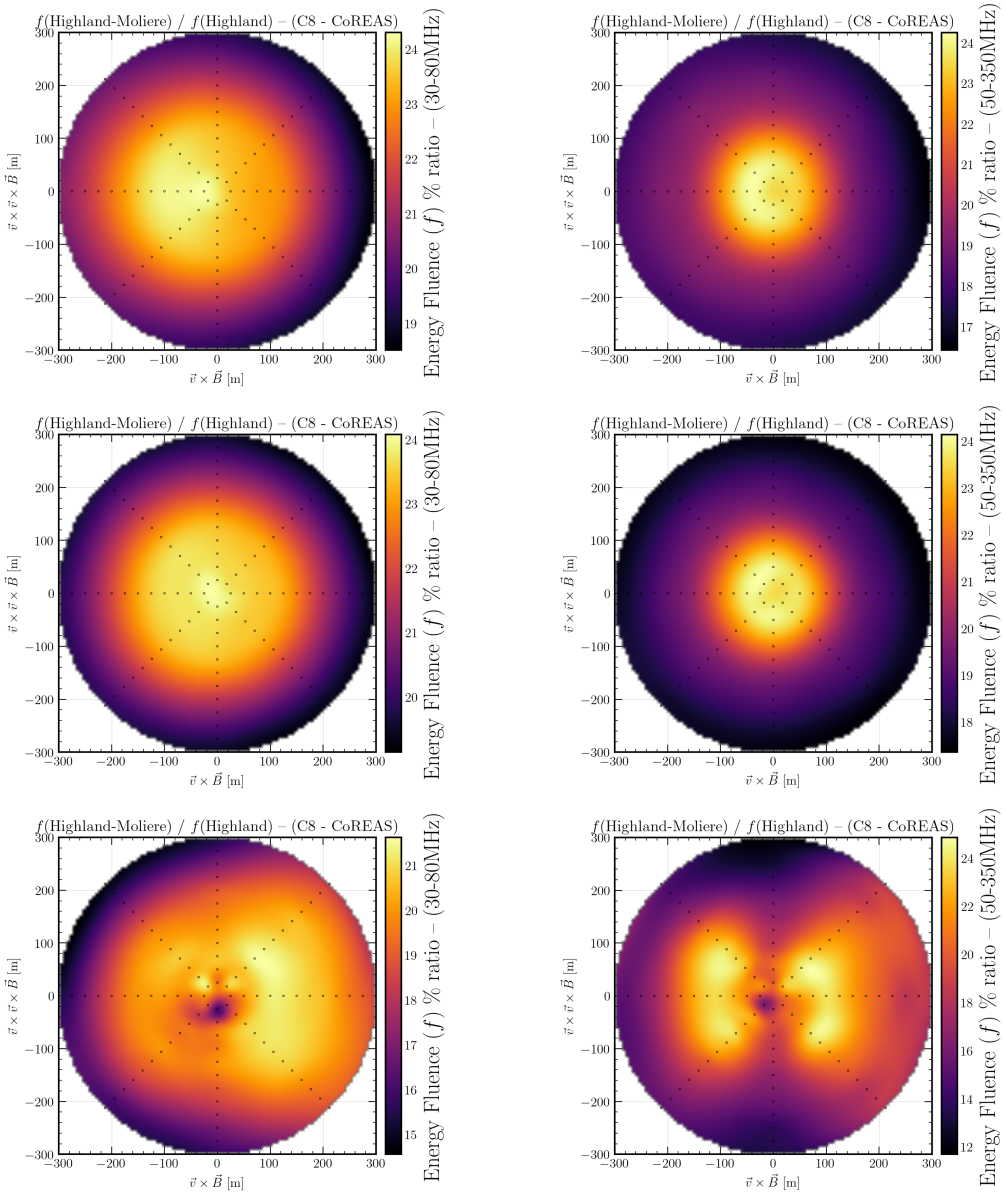


Figure 75: Ratio maps in fluence for 100, 1 PeV, photon induced vertical showers in the 30 MHz to 80 MHz (left column) and the 50 MHz to 350 MHz band (right column). In the first row, the ratio map in terms of total fluence is plotted, the second in terms of the geomagnetic contribution and the last row in terms of the charge excess. The radio emission algorithm that was used is CoREAS.

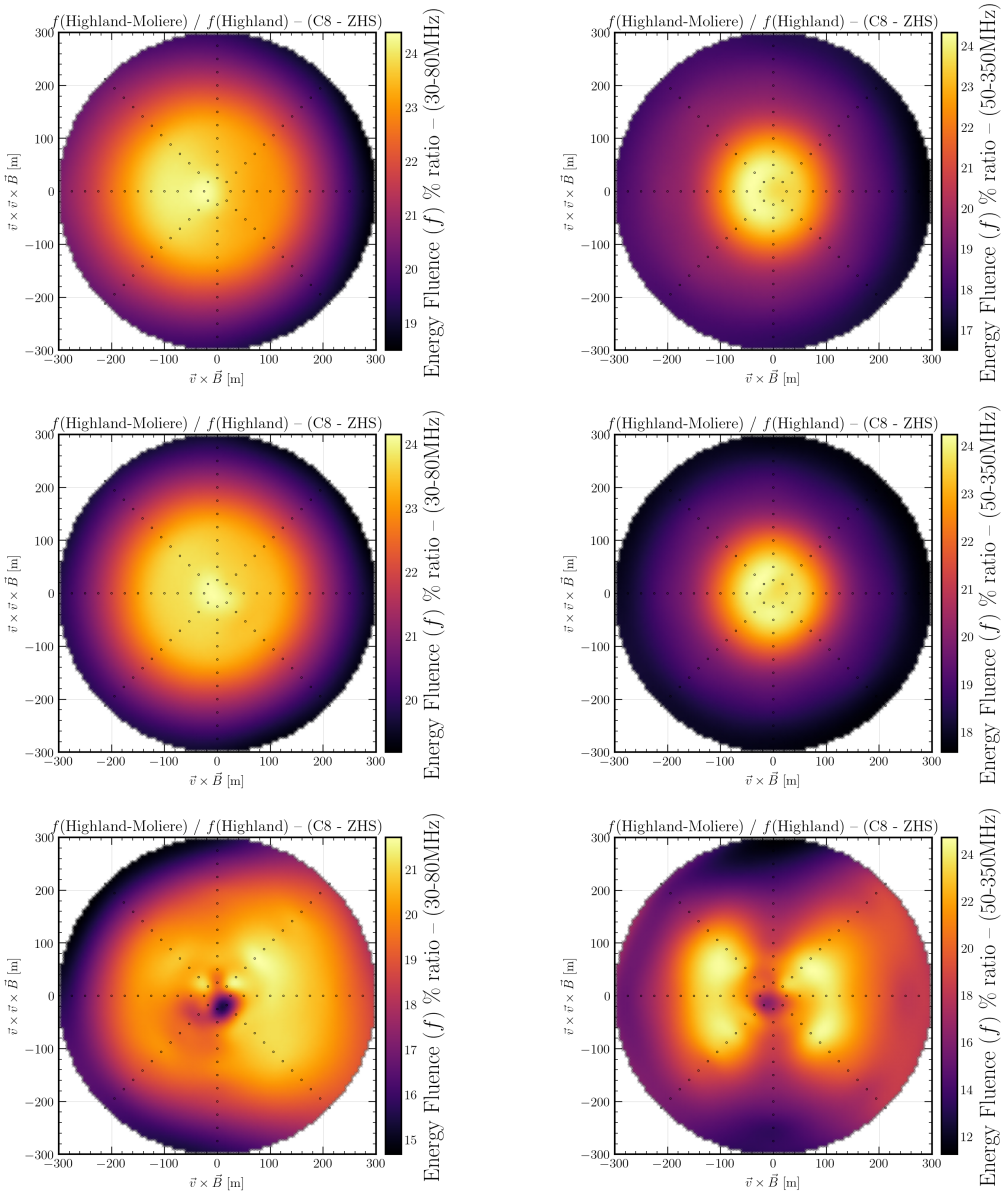


Figure 76: Ratio maps in fluence for 100, 1 PeV, photon induced vertical showers in the 30 MHz to 80 MHz (left column) and the 50 MHz to 350 MHz band (right column). In the first row, the ratio map in terms of total fluence is plotted, the second in terms of the geomagnetic contribution and the last row in terms of the charge excess. The radio emission algorithm that was used is ZHS.

Highland vs Molière – mean difference in terms of % (C8 ZHS)		
Fluence	30 MHz to 80 MHz	50 MHz to 350 MHz
Total Fluence	23.3%	22.4%
Geomagnetic Contribution	22.9%	22.0%
Charge Excess	19.9%	21.9%

Table 21: The difference in terms of % for fluence between Highland and Molière multiple scattering mechanisms for CORSIKA 8 ZHS. More fluence is emitted for the case of Highland scattering.

Highland vs Molière – mean difference in terms of % (C8 ZHS)		
Radiation Energy	30 MHz to 80 MHz	50 MHz to 350 MHz
Total Radiation Energy	22.8%	21.7%
Geomagnetic Contribution	22.9%	21.8%
Charge Excess	19.9%	21.8%

Table 22: The difference in terms of % for radiation energy between Highland and Molière multiple scattering mechanisms for CORSIKA 8 ZHS. More radiation energy is emitted for the case of Highland scattering.

10.3 Discussion

Simulations are crucial for a successful energy and $X_{\max}$ reconstruction. An approach used by LOFAR as explained in section 3.7 where one fits $X_{\max}$ with respect to the fluence of the best fitted simulation compared to measurements is a common technique also used in AERA [78], [27], [79]. One and can then extract the electromagnetic energy of the shower either by its quadratic dependency on radiation energy or by the longitudinal energy release of the electromagnetic component of the best fitted shower simulation. For energy reconstruction, one can use Lateral Distribution Functions (LDFs) that describe the lateral distribution of the fluence footprint. These LDFs though are a result of numerous simulation studies [80], [81]. Once again, this highlights the importance of the radio emission process to be well understood in order for a robust $X_{\max}$ and energy reconstruction. In this part of this thesis, the following insights have been showcased. The radio simulations do indeed depend on the track length of the simulated particle tracks. Smaller track lengths predict more radiation energy of up to 11.5% for the 30 MHz to 80 MHz band and up to 9.1% for the 50 MHz to 350 MHz band. Such an effect is already known and is being taken into account on the reconstruction procedures. The change in track length is more pronounced in CORSIKA 8, which indicates that a thorough revision of the tracking algorithm and its precision might be very fruitful. The fact though, that the CoREAS and ZHS formalisms practically converge for very small tracks indicates that simulating the radio emission for very small is the "ground truth". Furthermore, the effect of different multiple scattering approaches on the radio simulations was also studied. It was shown that the Highland approximation can predict more than 20% more radiation energy when compared to the Molière approach. This is because the Highland approach predicts more particles close to shower core which leads to the pulses being more coherent and produce more radiation energy. The Molière approach can more accurately describe the scattering angles of the charged particles, and it is chosen as the default in CORSIKA 8 and CORSIKA 7.

Part III

Using pulse-shape information for reconstructing Cosmic Ray Air showers with LOFAR and SKA

In the previous part, the prediction of radiation energy and the fluence footprint was studied for both CORSIKA 8 and CORSIKA 7. It was shown that in the case of very small, fine tracks the predictions are reliable and that the two main formalisms used by the community, namely "Endpoints" (CoREAS) and ZHS as implemented in CORSIKA 8, agree to a great extent. In this part of the thesis, the topic is going to be shifted to potential improvements on $X_{\max}$ reconstruction procedures. The focus will be the introduction and benchmarking of an $X_{\max}$ reconstruction technique which is based on the information hidden in the radio pulse aside its signal strength. This technique was developed in the context of LOFAR and SKA, in order to complement and potentially improve the resolution reached with the standard $X_{\max}$ reconstruction procedure used for LOFAR [49]. The standard procedure relies on the energy per area, or fluence. For the fluence calculation, as highlighted in the previous sections, only the signal strength and duration is being utilized. The rest of the information hidden in the pulse is thrown away and could potentially improve the $X_{\max}$ reconstruction procedure resolution. However, as can be seen in fig. 77 there are more characteristics to the radio pulse regarding its width and shape. It can be wider or narrower depending on $X_{\max}$ [82] and the distance of the antenna from the shower core. Close to the shower core the pulse appears wider and as one moves closer to the Cherenkov ring it reaches its maximum amplitude and gets narrower. Accordingly, the frequency spectra close to the shower core fall steeply and close to the Cherenkov ring become flat. Moving away from the Cherenkov ring the signal starts to get weaker, the pulse becomes wider and the frequency spectra start to fall steeply again. These observations are the main motivation of this study. The goal of this part of the thesis is to associate the pulse shape with $X_{\max}$ by devising a suitable reconstruction procedure.

The rest of this part is structured as follows: The method that associates the pulse shape with $X_{\max}$ is going to be presented, along with an example case study. Then the addition of noise to simulations is going to be introduced, along with necessary changes to the reconstruction procedure. The scheme is going to be benchmarked for the LOFAR (30 MHz to 80 MHz) and SKA (50 MHz to 350 MHz) frequency bands, using a large library of simulated LOFAR events. Finally, a few interesting ideas that were realized while developing the procedure will be discussed.

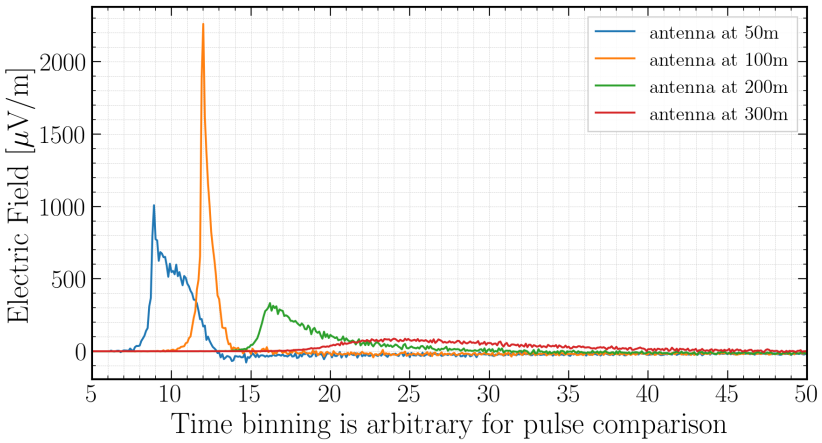


Figure 77: A CORSIKA 7 generated simulation of a $10^{17.2}$ eV iron induced air shower with the zenith angle set to 28.8°. The antennas are on the arm facing east of a standard star-shaped array pattern. The dominant polarization is plotted. The signal pulse shape varies for different distances from the shower core.

11 An $X_{\max}$ reconstruction scheme utilizing the pulse shape

For the development of the $X_{\max}$ reconstruction scheme that will be discussed here a large library of CORSIKA 7/CoREAS simulations is used. These are simulations of iron and proton showers with energies $\geq 10^{17.3}$ eV that have been previously used for reconstructing LOFAR measured events. There are a total of 148 events that were used in this study. All of them were produced with the default CORSIKA 7 values of maxRad equal to 0.2 rad and STEPFC equal to 1.0. Each event contains $\sim$ 30 simulations totalling roughly to a bit over 4500 simulations. The main idea of these simulations is that for each event, 30 simulations with identical settings are produced differing only in the seed used for the random number generator. Hence, only the shower development is affected resulting in a different $X_{\max}$ per simulation of that specific event. Regarding the antenna array, all simulations use a star-shaped pattern of 160 antennas (sampling period is set to 0.1 ns) located at the ground, in 20 concentric rings spaced equally from 25 m to 500 m from the shower axis with 8 antennas distributed azimuthally in each ring. With this setup one is able to obtain signal pulses from 160 antennas for showers whose only difference is $X_{\max}$. For instance, for a random antenna of the star-shaped grid there are $\sim$ 30 pulses for showers which differ only in $X_{\max}$. In order to find a quantifiable metric for $X_{\max}$ information, for every antenna location all these $\sim$ 30 pulses are compared among each other.

As an example, a reference simulation with an $X_{\max}$ value of 768 g cm^{-2} is being compared in the time and frequency domain with 3 different $X_{\max}$ values for 3 antenna locations. One value is significantly smaller, one is close to the reference value and the final one is significantly larger. For an antenna at 25 m from the shower core the comparison is shown in fig. 78, for an antenna at 100 m in fig. 79 and for an antenna at 200 m in fig. 80. The purpose of these plots is

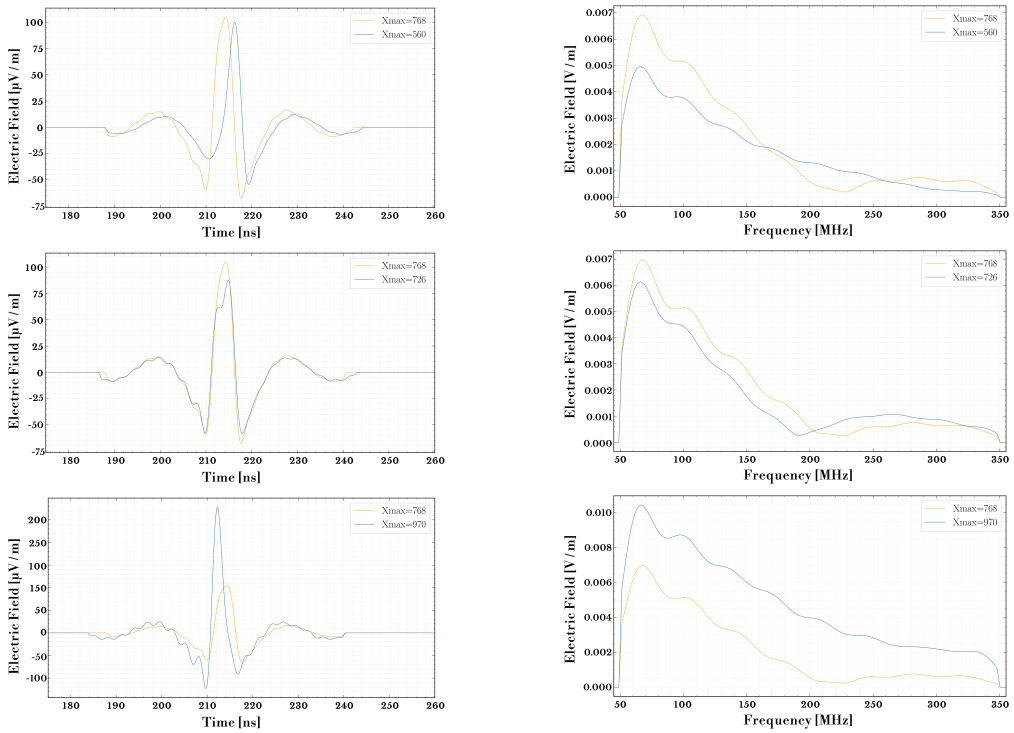


Figure 78: A radio signal from a shower with a reference $X_{\max}$ of 768 g cm^{-2} is directly compared with identical simulations differing only in $X_{\max}$ for an antenna located 25 m from the shower core. The radio pulses of the dominant polarization are shown on the left column and the corresponding frequency spectra on the right column. The top row illustrates the comparison for an $X_{\max}$ value of 560 g cm^{-2} versus the reference. The middle row illustrates the comparison for an $X_{\max}$ value of 726 g cm^{-2} versus the reference. The bottom row illustrates the comparison for an $X_{\max}$ value of 970 g cm^{-2} versus the reference. For all the pulses a band-pass filter in the 50 MHz to 350 MHz frequency band, along with a 55 ns window around the pulse peak is applied.

to simply illustrate quantitative differences in both time and frequency domains. By looking at the amplitudes of all the pulses one can easily confirm that showers with higher $X_{\max}$ spread less on the ground. For this analysis though, the information revealed from the amplitudes is going to be intentionally left out. The subtle differences in the pulse shape between identically set showers with different shower development is going to be utilized. It was quickly realized during this study that taking out the amplitude information and then correlate the pulses would always lead to correlation factor values very close to 1. This effort is going to be discussed on a later section as it provided me with some useful insights. However, it was also realized that the frequency domain allows easier identification of $X_{\max}$ dependent features, even when taking out the amplitude information. For this reason, the reconstruction scheme presented next was developed in the frequency domain.

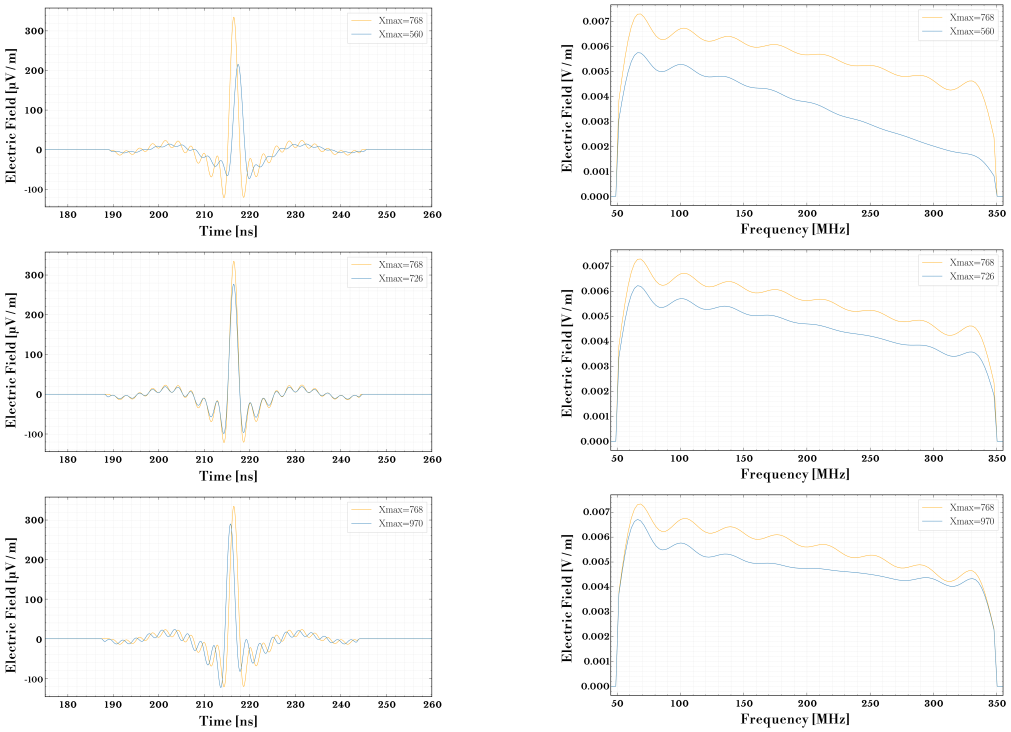


Figure 79: A radio signal from a shower with reference $X_{\max}$ of $768\,\mathrm{g\,cm^{-2}}$ is directly compared with identical simulations differing only in $X_{\max}$ for an antenna located 100 m from the shower core. The radio pulses of the dominant polarization are shown on the left column and the corresponding frequency spectra on the right column. The top row illustrates the comparison for an $X_{\max}$ value of $560\,\mathrm{g\,cm^{-2}}$ versus the reference. The middle row illustrates the comparison for an $X_{\max}$ value of $726\,\mathrm{g\,cm^{-2}}$ versus the reference. The bottom row illustrates the comparison for an $X_{\max}$ value of $970\,\mathrm{g\,cm^{-2}}$ versus the reference. For all the pulses a band-pass filter in the 50 MHz to 350 MHz frequency band, along with a 55 ns window around the pulse peak is applied.

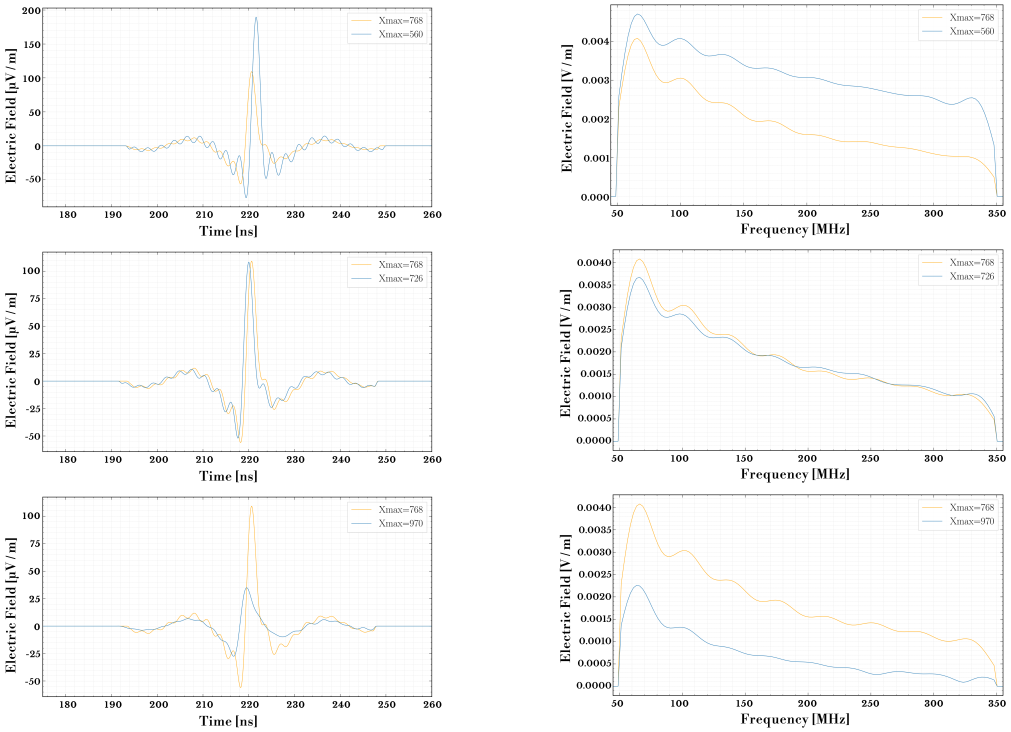


Figure 80: A radio signal from a shower with reference $X_{\max}$ of $768\,\mathrm{g\,cm^{-2}}$ is directly compared with identical simulations differing only in $X_{\max}$ for an antenna located 200 m from the shower core. The radio pulses of the dominant polarization are shown on the left column and the corresponding frequency spectra on the right column. The top row illustrates the comparison for an $X_{\max}$ value of $560\,\mathrm{g\,cm^{-2}}$ versus the reference. The middle row illustrates the comparison for an $X_{\max}$ value of $726\,\mathrm{g\,cm^{-2}}$ versus the reference. The bottom row illustrates the comparison for an $X_{\max}$ value of $970\,\mathrm{g\,cm^{-2}}$ versus the reference. For all the pulses a band-pass filter in the 50 MHz to 350 MHz frequency band, along with a 55 ns window around the pulse peak is applied.

11.1 Fluence percentage

The pulse shape information in the frequency domain is encapsulated in the slope of the frequency spectra, so one could argue that taking the ratio of 2 spectra with different $X_{\max}$ might provide access to $X_{\max}$. Such an idea, although interesting would be seriously affected by noise, so a more robust metric has to be devised. Starting from the frequency spectra of one polarization, the square of the amplitude spectrum of each frequency bin is calculated and then its cumulative sum. Finally, all the bins of the cumulative sum are normalized with respect to the total fluence of the polarization of that antenna. This way, this quantity encapsulates the rate at which the energy is deposited to the antenna independent of the amplitude. This is defined as *fluence percentage* (F), since at the start of the frequency band in which we filter our pulses, 0% of the total fluence is deposited to the antenna, while at the end of the frequency band 100% of the total fluence has been deposited to it. For a random bin j, the *fluence percentage* can be summarized with the following formula:

$$F_j = \frac{\sum_{i=1}^{j} A_i^2(v)}{\sum_{i=1}^{N} A_i^2(v)} \tag{38}$$

With A, the amplitude spectra and with v the frequency are denoted. It is worth noticing that the denominator of eq. (38) multiplied by the constant factors $\epsilon_0 c \Delta t$, where ϵ_0 is the vacuum permittivity, c is the speed of light in vacuum and Δt is the finite time binning is the *total fluence* with respect to which, every bin j is normalized. These constant factors appear also in the numerator of eq. (38) so they eventually cancel out.

The fluence percentage of the dominant polarization for different $X_{\max}$ values are plotted for an antenna at 25 m from the shower core in fig. 81, at 100 m in fig. 82 and at 200 m in fig. 83. A reference $X_{\max}$ value of $768\,\mathrm{g\,cm^{-2}}$ is compared with the fluence percentage from different showers that differ only in $X_{\max}$ as discussed earlier. Drawn with grey, is the area between the fluence percentage of the 2 showers as is shown in each of the antenna locations (fig. 81, fig. 82 and fig. 83). The hypothesis is that this *difference in area* (D_A) between the 2 showers is large when the showers differ strongly in $X_{\max}$, gets smaller for similar $X_{\max}$ values and is *zero* for identical $X_{\max}$ values.

To test this hypothesis a library of ~ 30 simulations that differ only in $X_{\max}$ is used as explained below. One of these simulations is picked and treated as *reference* and is compared with the rest. Each antenna of this *reference* is being compared with the same antenna with another simulation for all polarizations. As a result, each antenna is characterized by the metric presented earlier, i.e. the *difference in area* (D_A) of the fluence percentage for each polarization between the *reference* and the simulation it is compared with. The average of the *difference in area* (D_A) of all polarizations gives us the total *difference in area* (D_A^i) for that antenna i. Finally, the average of the *difference in area* (D_A^i) of each antenna in the antenna array are added together to provide one final number (D_A^{total}). This quantity acts as the final metric for each different comparison between the *reference* and the simulation it was compared with. Once again, the smaller the

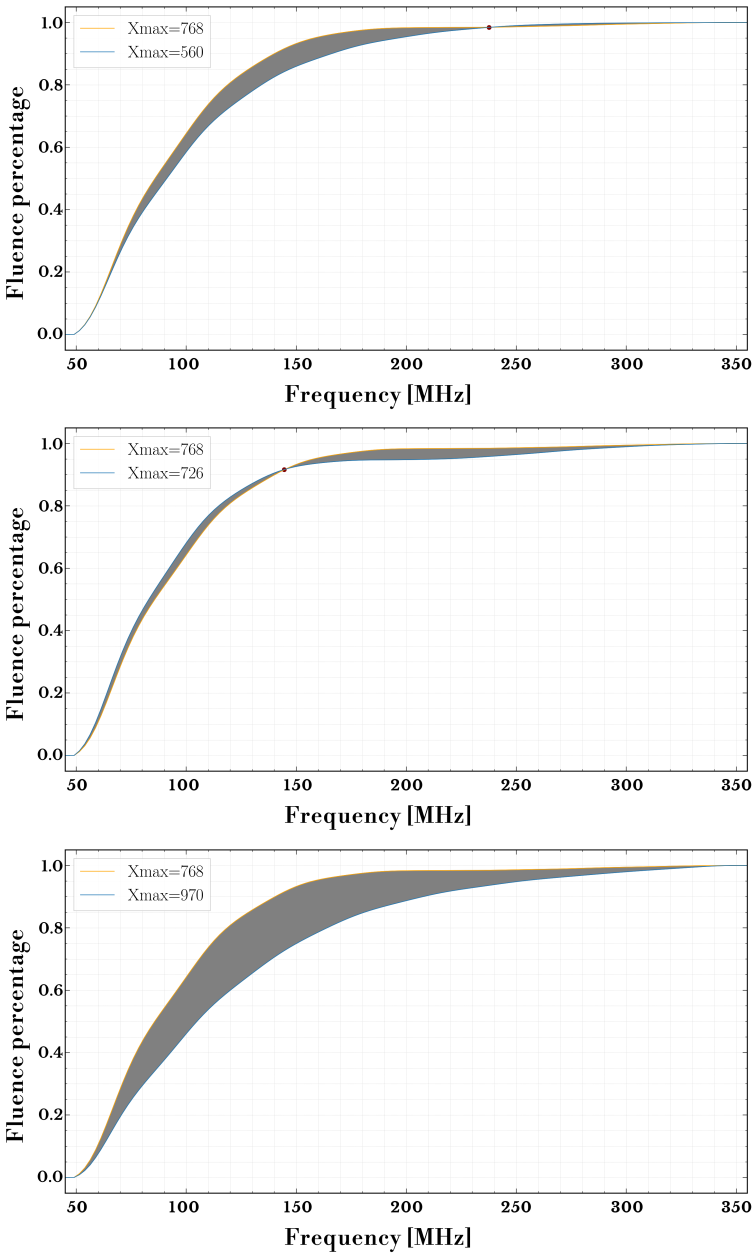


Figure 81: The fluence percentage for a reference $X_{\max}$ of $768\,\mathrm{g\,cm^{-2}}$ is directly compared with identical simulations differing only in $X_{\max}$ for an antenna located 25 m from the shower core. The top plot illustrates the comparison for an $X_{\max}$ value of $560\,\mathrm{g\,cm^{-2}}$ versus the reference. The middle plot for an $X_{\max}$ value of $726\,\mathrm{g\,cm^{-2}}$ versus the reference, and the bottom plot for an $X_{\max}$ value of $970\,\mathrm{g\,cm^{-2}}$ versus the reference. For all the pulses a band-pass filter in the 50 MHz to 350 MHz frequency band, along with a 55 ns window around the pulse peak is applied.

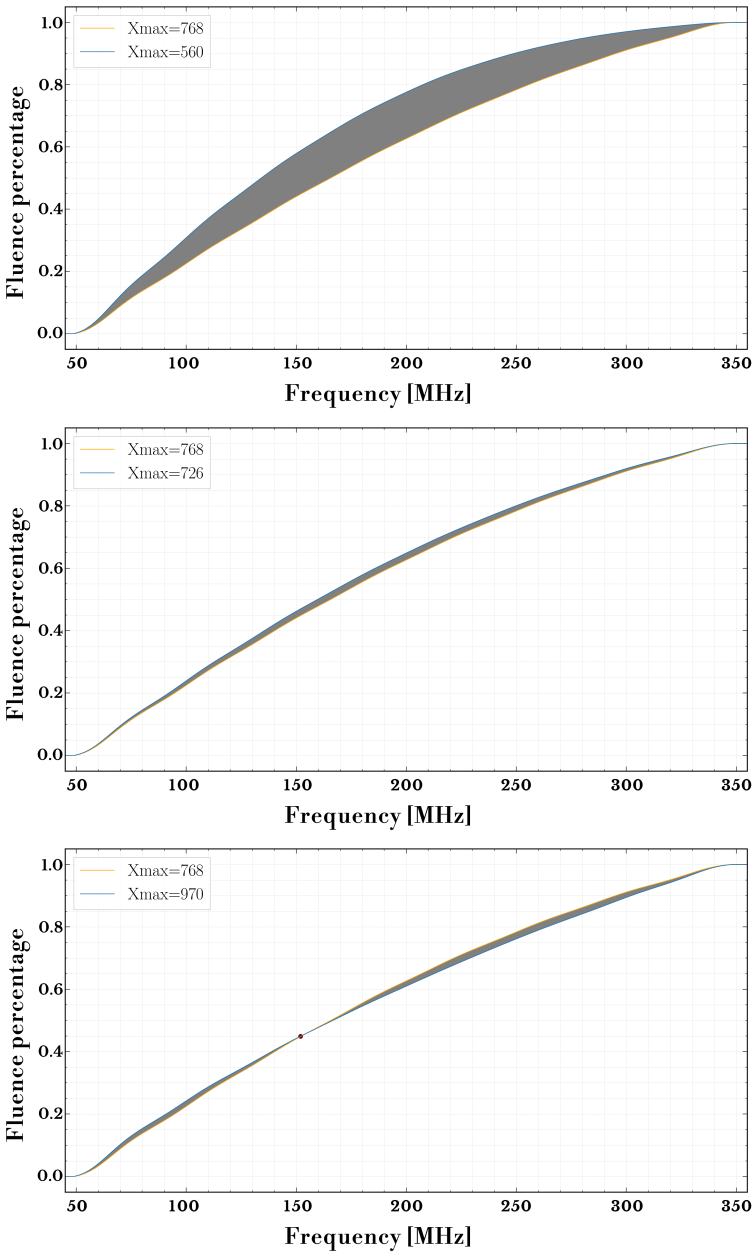


Figure 82: The fluence percentage of the dominant polarization for a reference $X_{\max}$ of 768 g cm^{-2} is directly compared with identical simulations differing only in $X_{\max}$ for an antenna located 100 m from the shower core. The top plot illustrates the comparison for an $X_{\max}$ value of 560 g cm^{-2} versus the reference. The middle plot for an $X_{\max}$ value of 726 g cm^{-2} versus the reference, and the bottom plot for an $X_{\max}$ value of 970 g cm^{-2} versus the reference. For all the pulses a band-pass filter in the 50 MHz to 350 MHz frequency band, along with a 55 ns window around the pulse peak is applied.

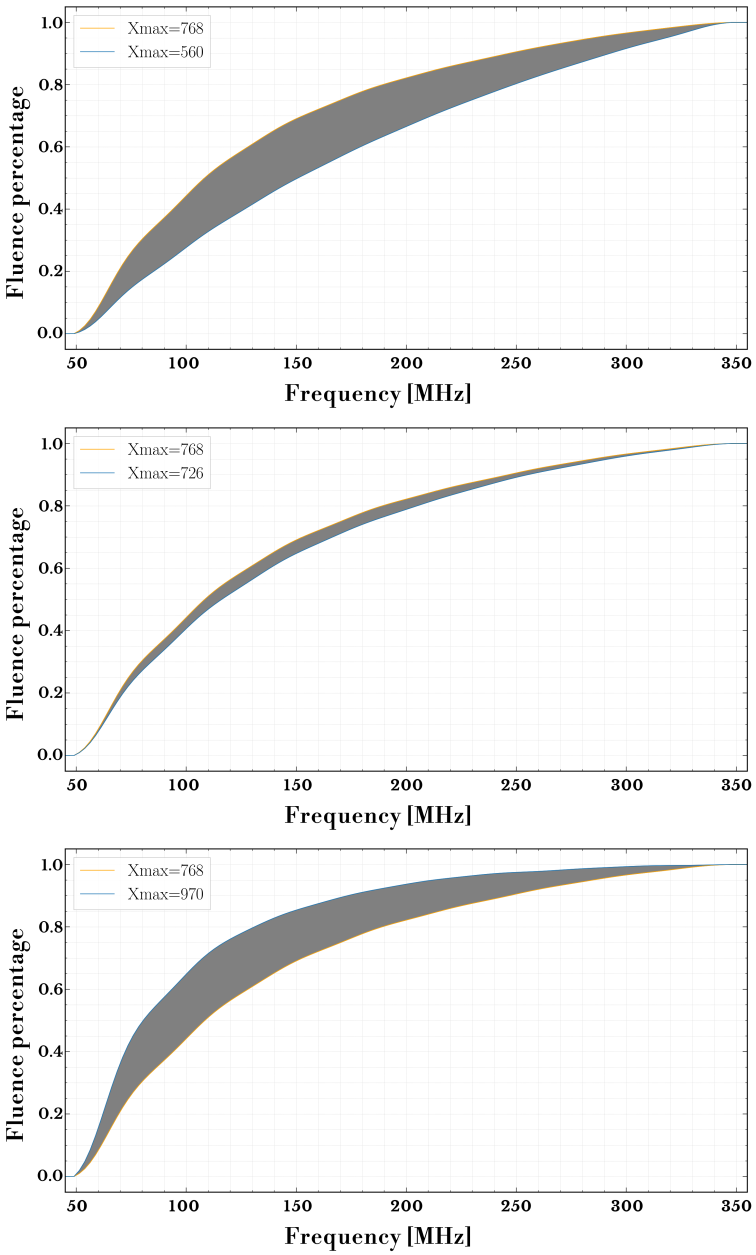


Figure 83: The fluence percentage of the dominant polarization for a reference $X_{\max}$ of 768 g cm^{-2} is directly compared with identical simulations differing only in $X_{\max}$ for an antenna located 200 m from the shower core. The top plot illustrates the comparison for an $X_{\max}$ value of 560 g cm^{-2} versus the reference. The middle plot for an $X_{\max}$ value of 726 g cm^{-2} versus the reference, and the bottom plot for an $X_{\max}$ value of 970 g cm^{-2} versus the reference. For all the pulses a band-pass filter in the 50 MHz to 350 MHz frequency band, along with a 55 ns window around the pulse peak is applied.

value of D_A^{total}, the closer the *reference* is with the simulation it is compared with in terms of $X_{\max}$. All the simulations in the library are compared according to the procedure explained with the *reference*. Hence, for each comparison of a simulation with the *reference* corresponds a D_A^{total} value (also referred simply as difference in area in the plots). It is worth noting at this point, that the antenna locations used for the comparisons that will be shown from here and on are the ones with distance from the shower core up to 250 m. The reason for this choice is simply to ensure that there is indeed at least some weak signal and not plain noise in our simulations.

11.2 Parabola fitting

The data-points that include the information in *difference in area* per simulation are shown in fig. 84. The *reference* simulation used for this plot has an $X_{\max}$ value of $619.9\,\mathrm{g\,cm^{-2}}$ and as expected, when compared with itself its D_A^{total} is zero. However, this data-point is excluded from fig. 84 so as not to interfere with the parabola fitting scheme that is going to presented next. As can be seen, simulations with near $X_{\max}$ values to the reference have a smaller D_A^{total}, while the simulations that differ more strongly from the reference have a larger D_A^{total}. Just by observing the data-points one can see that there is structure. Inspired by the LOFAR standard $X_{\max}$ reconstruction procedure [48] a very similar parabola fitting scheme is being performed here as well.

First, the D_A^{total} of *reference* simulation is excluded as this would shift the fitted parabola towards it and it would produce biased results. Having this data-point thrown away, the point with the minimum D_A^{total} is found and its $X_{\max}$ value is identified. Then a "grammage window" is created around this point to include the points that are close in terms of D_A^{total}. This window can be very important as it can alter the results as will be shown later. A check is being made for points left and right around the minimum point separately. If a point with a specific $X_{\max}$ has a greater value in D_A^{total} from both of its adjacent points then this point is excluded. In fig. 84 all points can be seen in green color, while the purple points are the ones that fulfill the condition that was described and are at the same time within the desired "grammage window". Finally, these points are the ones that are used to fit a simple parabola. The minimum of this parabola is considered the reconstructed shower maximum ($X_{\max-\mathrm{reco}}$). For simulations, in the absence of noise this reconstruction scheme seems to be working quite well, since the difference between $X_{\max-\mathrm{reco}}$ and reference $X_{\max}$ is $\sim 0.01\,\mathrm{g\,cm^{-2}}$.

Although, this result seems very satisfying there some parameters that can seriously affect the reconstructed $X_{\max-\mathrm{reco}}$. These are going to be discussed now. Maybe the most important parameter that can affect the outcome is how many simulations are available around the true $X_{\max}$ value. If there are not enough simulations around it, one can end up with a situation like the one shown in fig. 85. In the absence of a "cluster of simulations" around the *reference* simulation the parabola fit is not very accurate anymore and in the case of fig. 85 the reconstruction is short for $\sim 27\,\mathrm{g\,cm^{-2}}$. Another important technical factor that influences the fitting procedure is the size of the "grammage window" chosen for the parabola fitting. 2 cases of parabola fits for 2 cases with varying "grammage windows" are shown in fig. 86. On the left column the *reference* simulation has a "cluster" of simulations around it, while on the right column the points around the chosen *reference* simulation are sparse. The choice of window size affects the resolution of the reconstruction as shown in fig. 86 and for some cases, the fitting can also break down. Both parameters discussed are not considered a problem though, as one is always able to run more simulations and adjust the "grammage window" in order to improve the fitting procedure. As a final and more obvious remark, the parabola fitting scheme will break down if the *reference* simulation is chosen near the edge of the simulation set, i.e. there is no cluster of simulations around its $X_{\max}$ and rather only simulations with either greater or smaller $X_{\max}$.

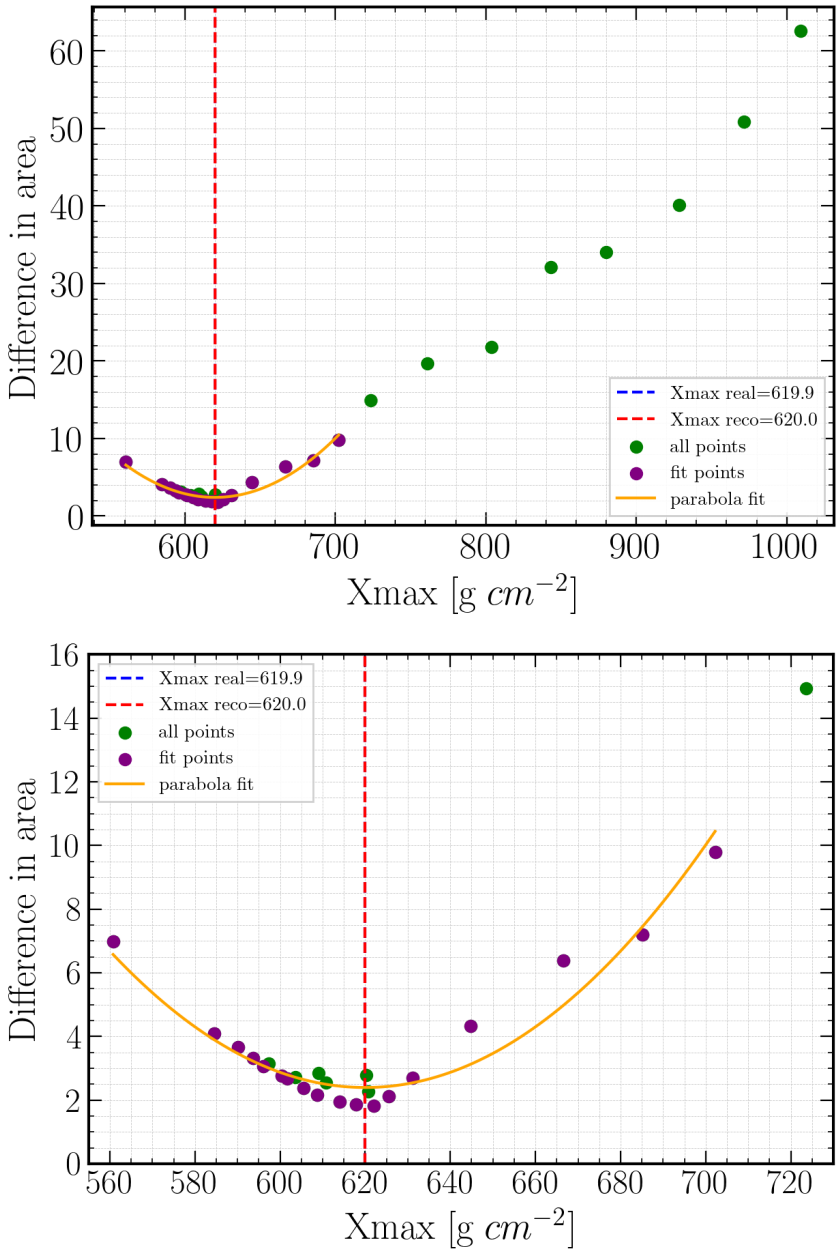


Figure 84: Parabola fit within a "cluster" of simulations around the *reference* $X_{\max}$. Having enough simulations around the $X_{\max}$ we seek, improves the resolution of this method by a considerable amount. On the bottom plot, the region of interest where the parabola fitting takes place is zoomed in.

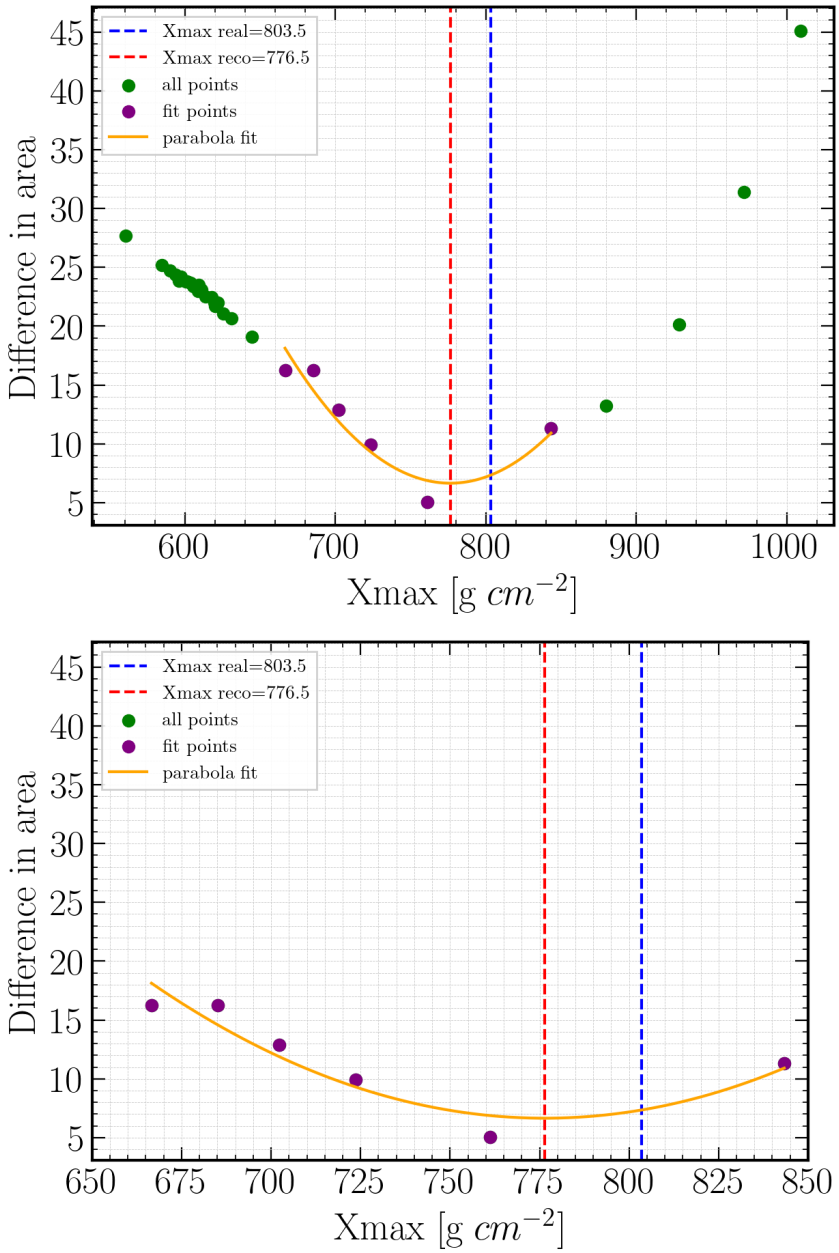


Figure 85: Parabola fit with very few of simulations around the *reference* $X_{\max}$. Not having enough simulations around the $X_{\max}$ we seek, throws off the resolution of this method. On the bottom plot, the region of interest where the parabola fitting takes place is zoomed in.

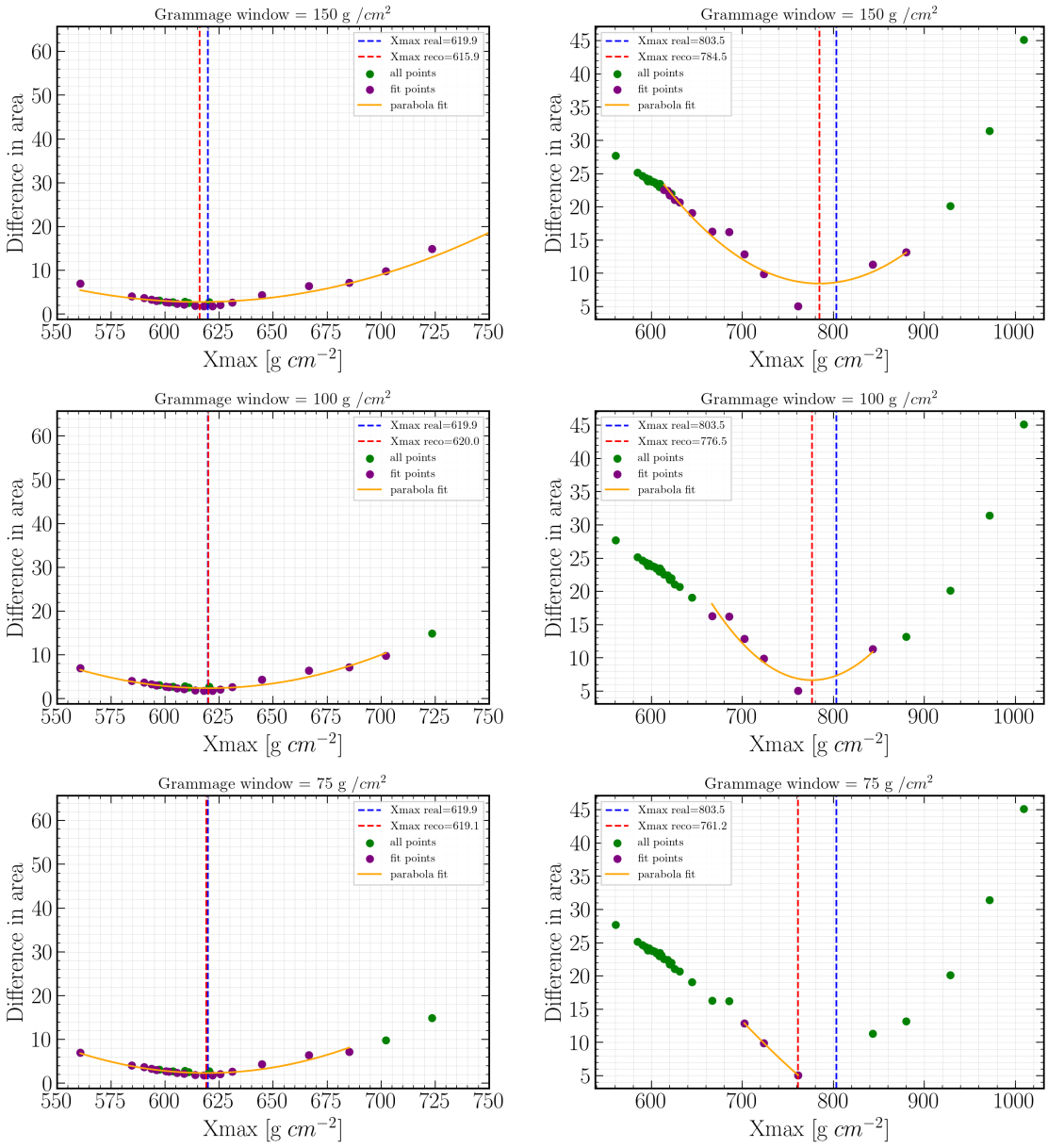


Figure 86: The effect of of the "grammage window" on reconstructing $X_{\max}$ is shown for 2 cases. On the left column the *reference* simulation has a "cluster" of simulations around it making the parabola fit work for different "grammage windows" of 150, 100 and 75 g cm^{-2} from top to bottom. The accuracy of the reconstruction is affected with smaller windows reaching a closer value to the *reference* $X_{\max}$. On the right column, the *reference* simulation has no "cluster" of simulations around it. As a result, the fit is poor and for decreasing values of the "grammage window" (top to bottom) and it can even break due to small amount of points being fitted.

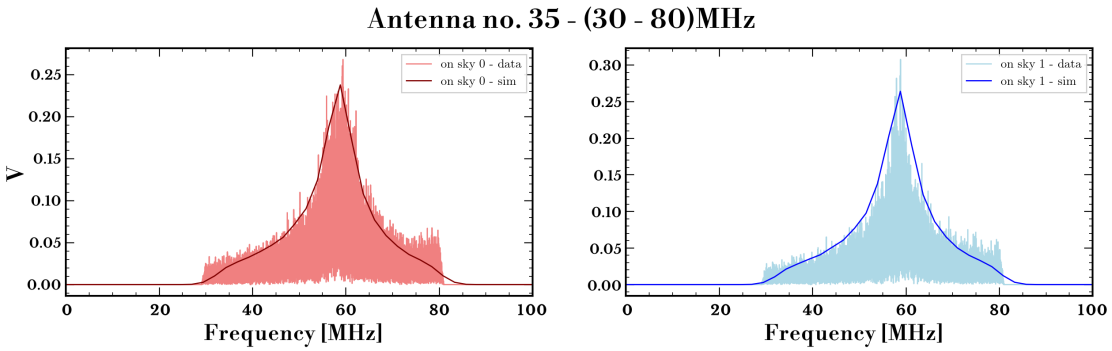


Figure 87: Frequency spectra of experimental data versus the corresponding CoREAS simulation folded with the LOFAR antenna model for a random antenna in a random LOFAR station. Both polarizations are shown in left and right. The LOFAR antennas are highly resonant around 60 MHz.

11.3 Generating noise for the simulations

The idea of developing an $X_{\max}$ reconstruction method was initially thought for LOFAR. Although, moving to the frequency domain implies that a wide range of the frequency spectra is covered by the antennas used in order for the reconstruction scheme presented here to work. Unfortunately, this is not the case for LOFAR as can be seen in fig. 87. In fig. 87 a real event measured with LOFAR in one of the antennas of the inner stations is plotted versus the corresponding simulations. The simulations are folded with the LOFAR antenna model and it is shown that data and simulations agree quite well. This indicates that the antenna model used in LOFAR is reasonably well understood. However, this also showcases that while LOFAR is nominally detecting radio signals from cosmic rays in the 30 MHz to 80 MHz band, in reality the antennas are very resonant around ~ 60 MHz. This was the main reason why it was decided to proceed with the development of the pulse-shape reconstruction method for SKA where the planned frequency band for detection is 50 MHz to 350 MHz. For SKA there is no noise model at the moment of developing this scheme, hence a simplistic way of generating noise for the simulations was developed in order to showcase how the scheme works in the presence of Gaussian noise.

Next, the way that noise was generated for application to the existing library of simulations is described. In real world analysis scenarios, the experimental data are noisy and CORSIKA 7/CoREAS simulations (which have no noise) are used for reconstruction. Hence, noise is added to test the scheme only in the *reference* simulation. In the *reference* simulation, the antenna with the strongest signal in its dominant polarization is identified. A fraction of the amplitude of this signal pulse (5%, 10% and 20%) is used to set the width of the Gaussian distribution used for noise generation. For every antenna separately, random Gaussian noise is generated in the time domain and the width of the Gaussian is a fraction of the amplitude mentioned above. The generated noise is then added bin by bin on top of the simulated antenna pulse. This way every antenna has a different noise signature, with the noise amplitude being

the same for all antennas. Ideally, the goal is to minimize the amount of noise in the frequency spectra. To achieve that, a Hann time window [83] of 55 ns around the pulse's peak is applied and the rest of the trace is set to 0 values. The advantage of applying this window is that the the edges of the pulse between the zero values and the non-zero values are becoming smooth, which reduces the noise in the frequency domain. The pulse is then transformed to the frequency domain and a simple band pass filter in the 50 MHz to 350 MHz range is applied.

An example of the procedure in the time domain is shown in fig. 88. The *reference* simulation (plotted in blue) is initially filtered in the 50 MHz to 350 MHz band. On top of that, random Gaussian noise with a width of 10% of the strongest pulse of all antennas within the *reference* simulation is added and in orange the reuslting pulse is plotted. Next, the pulse with the included noise is filtered in the 50 MHz to 350 MHz band and finally a 55 ns "Hann" window is applied around the pulse's peak. It is shown that this procedure alters the pulse to some extent depending on the width of the Gaussian distribution of the generated noise. The choice of a time window of 55 ns is inspired by the LOFAR standard $X_{\max}$ reconstruction method where a 55 ns window is used there as well for the analysis. Having constrained the pulse in this time window, one can now switch to the frequency domain. In fig. 89 the frequency spectra are shown along with the corresponding fluence percentage. It is worth noting once more that the pulses start out identical and after the noise generation and addition procedure, one can see the effect of noise in both the frequency spectra and the fluence percentage. The noise distribution generated is chosen to be different for each antenna in the simulation and is the same only in the width of the Gaussian. This way every antenna has the same noise magnitude but not identical noise distributions.

It is now instructive to inspect in a library of ~ 30 simulations how the presence of generated noise affects the parabola fitting procedure as was shown earlier. This is shown in fig. 90 for 3 cases of noise. From left to right, $5\%, 10\%$ and 20% of noise is added. It can be seen that the addition of noise starts to break the structure of the data-points especially the ones with similar $X_{\max}$. Naturally, the parabola fitting is not as efficient under these circumstances and especially with the addition of a large amount of noise. Hence, $X_{\max}$ is not reconstructed with a satisfying level of accuracy anymore.

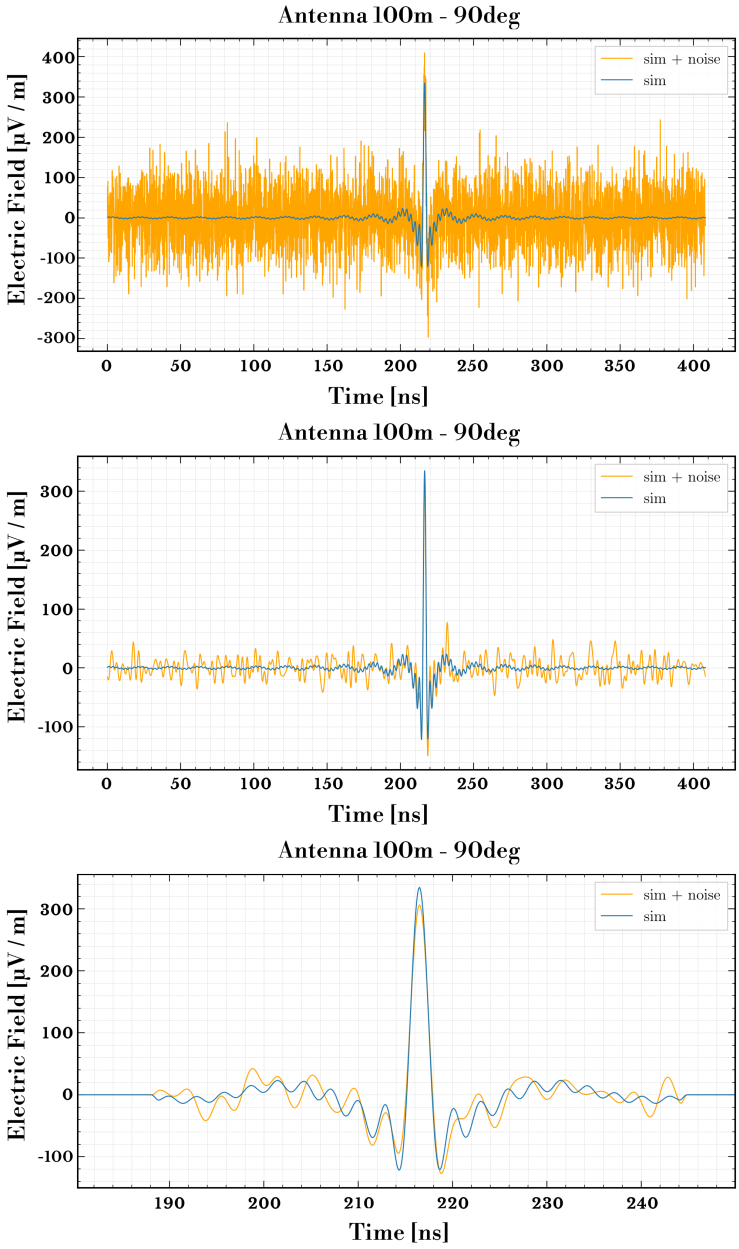


Figure 88: A simulation of a signal pulse filtered in the 50 MHz to 350 MHz band is shown with the blue line. The orange line indicates the same simulation with the addition of noise in steps. On top, unfiltered Gaussian noise with 10% of the amplitude is added to the whole trace. In the middle, the noise is filtered and on the bottom plot a "Hann" window of 55 ns is applied to both pulses.

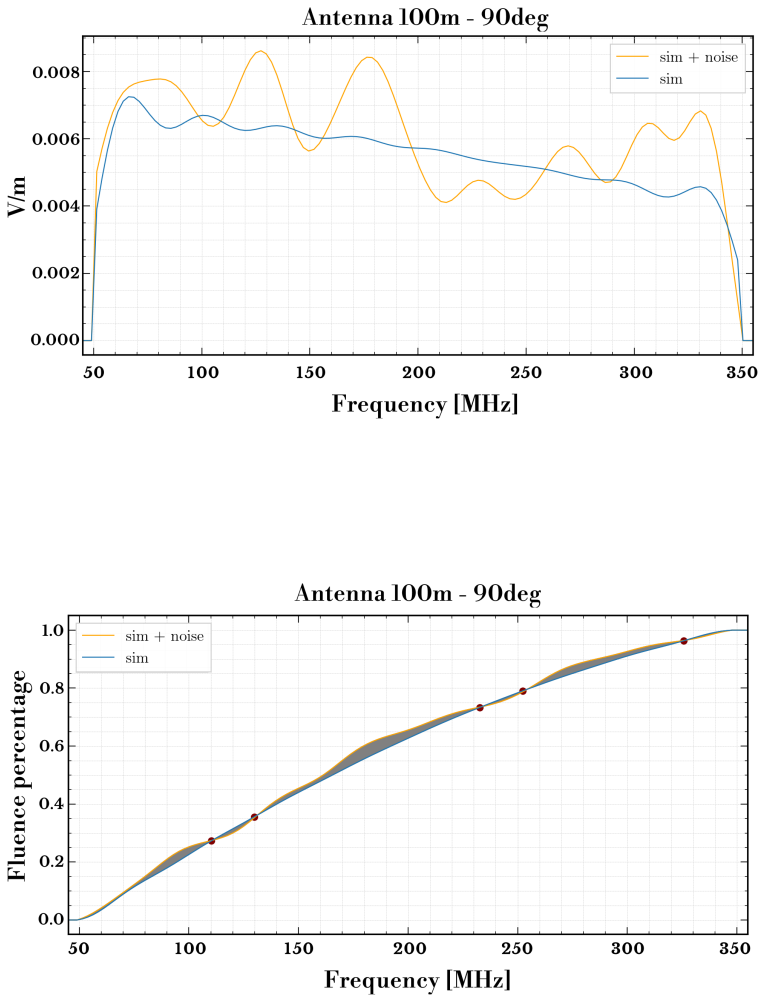


Figure 89: The corresponding frequency spectra of fig. 88 (bottom) is shown on the top. On the bottom, the fluence percentage is plotted and the effect of noise is shown.

This indicates that a way needs to be found in order for the contribution of the noise to be taken into account and to be corrected. The following scheme was implemented to do so. Before applying the 55 ns time window to the pulses that include the generated noise, one can see the whole trace with filtered noise in the 50 MHz to 350 MHz band for a random antenna i. At this stage, a same sized (55 ns) time window is applied and stored separately at the end of the trace as indicated with a red box in fig. 91. At those time bins the signal is very weak and the contribution to the fluence percentage comes mainly from the noise. The amplitude spectra of this part of the trace (red box) is calculated per antenna i and denoted as $cor_i(v)$, where v is the frequency. This is perceived as a "correction" due to noise and this quantity squared is subtracted from the contribution to the fluence percentage as shown in eq. (39). The quantity in eq. (39) is referred as the "corrected fluence percentage" (F_j^c) and its components are defined in the same way as eq. (38).

$$F_j^c = \frac{\sum_{i=1}^{j}\Big(A_i^2(v) - cor_i^2(v)\Big)}{\sum_{i=1}^{N}\Big(A_i^2(v) - cor_i^2(v)\Big)} \tag{39}$$

To test the correction due to noise that was just introduced another set of $\sim$ 30 simulations, where the *reference* simulation has a "cluster" of simulations around its $X_{\max}$ and 3 levels of added noise ($5\%, 10\%$ and 20%) are added as before. During the parabola fitting step of the reconstruction scheme, the data-points have been corrected for the noise. Therefore, some structure is more evident and the method returns a reconstructed $X_{\max-\mathrm{reco}}$ with improved accuracy. This is shown in fig. 92, where for 5% added noise $X_{\max-\mathrm{reco}}$ is off by $\sim 2.17\,\mathrm{g\,cm^{-2}}$, for 10% it is off by $\sim 10.41\,\mathrm{g\,cm^{-2}}$ and for 20% it is off by $\sim 35.51\,\mathrm{g\,cm^{-2}}$.

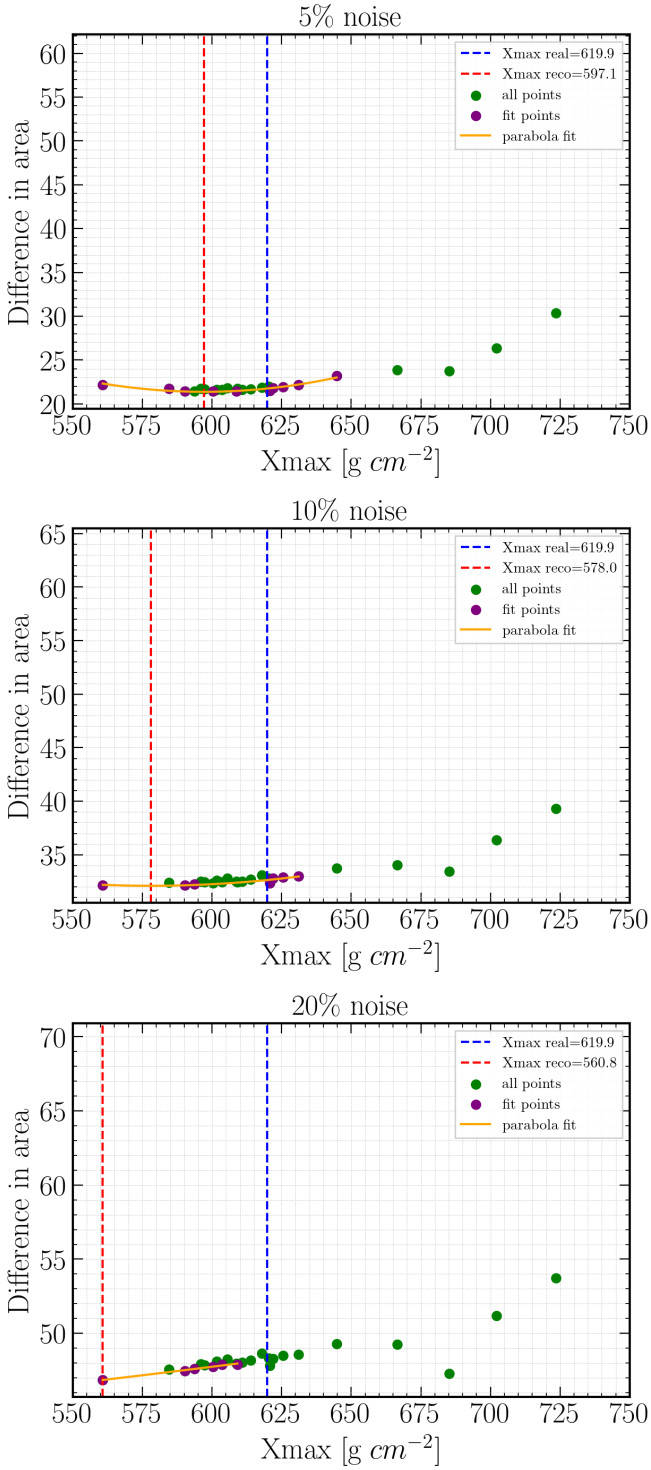


Figure 90: The parabola fitting procedure under the presence of different levels of generated noise. From top to bottom according to the noise addition procedure described in the text, $5\%, 10\%$ and 20% of noise is added to the simulations. Adding more and more noise seems to break the structure close to the *reference* simulation making the parabola fitting even fail on some occasions.

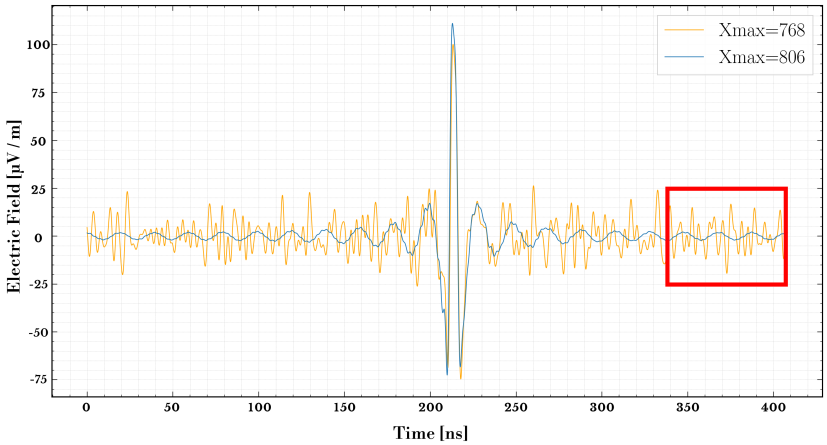


Figure 91: A simulated pulse with no noise is indicated with blue. With orange, a simulated pulse with added filtered noise in the 50 MHz to 350 MHz band is shown. The red box at the end of the trace indicates a 55 ns time window where the time bin contributions of the orange pulse come mainly from the noise, since the signal is very weak there. The noise contribution in this red box is used to "correct" the fluence percentage due to noise in eq. (39).

11.4 Benchmarking the $X_{\max}$ reconstruction scheme

In this section, the $X_{\max}$ reconstruction scheme that was presented is going to be benchmarked. For this, the whole library of simulations of 148 LOFAR events totalling to ≥ 4500 mentioned at the start of this chapter are used. All these simulations were generated with CORSIKA 7/CoREAS for a star-shaped antenna array on the ground. The primary energies of these proton and iron showers are $\geq 10^{17.3}$ eV and differ in zenith angle ($0° - 50°$). Each event consists of roughly 30 simulations that differ only in random number seed and thus shower evolution. For each event, an array of its simulations is sorted in ascending order with respect to $X_{\max}$. The fifth shower in that array is picked as a *reference* shower in order to have enough simulations with a smaller $X_{\max}$ than the *reference* one and not feed the parabola fitting scheme with an edge case that will break it. The $X_{\max}$ reconstruction method is then performed by comparing the chosen *reference* shower with the rest of showers in the array. As bias (b) the following quantity is defined:

$$b = X_{\max-\text{reco}} - X_{\max} \tag{40}$$

The bias of this comparison is stored and as a *reference* the next shower in the array of the same event is used. This is now compared with the rest of the showers of that event and its bias is stored. This procedure is repeated until the fifth shower from the end of the array is picked. A total bias of the event is defined as the mean value of the bias calculated for each

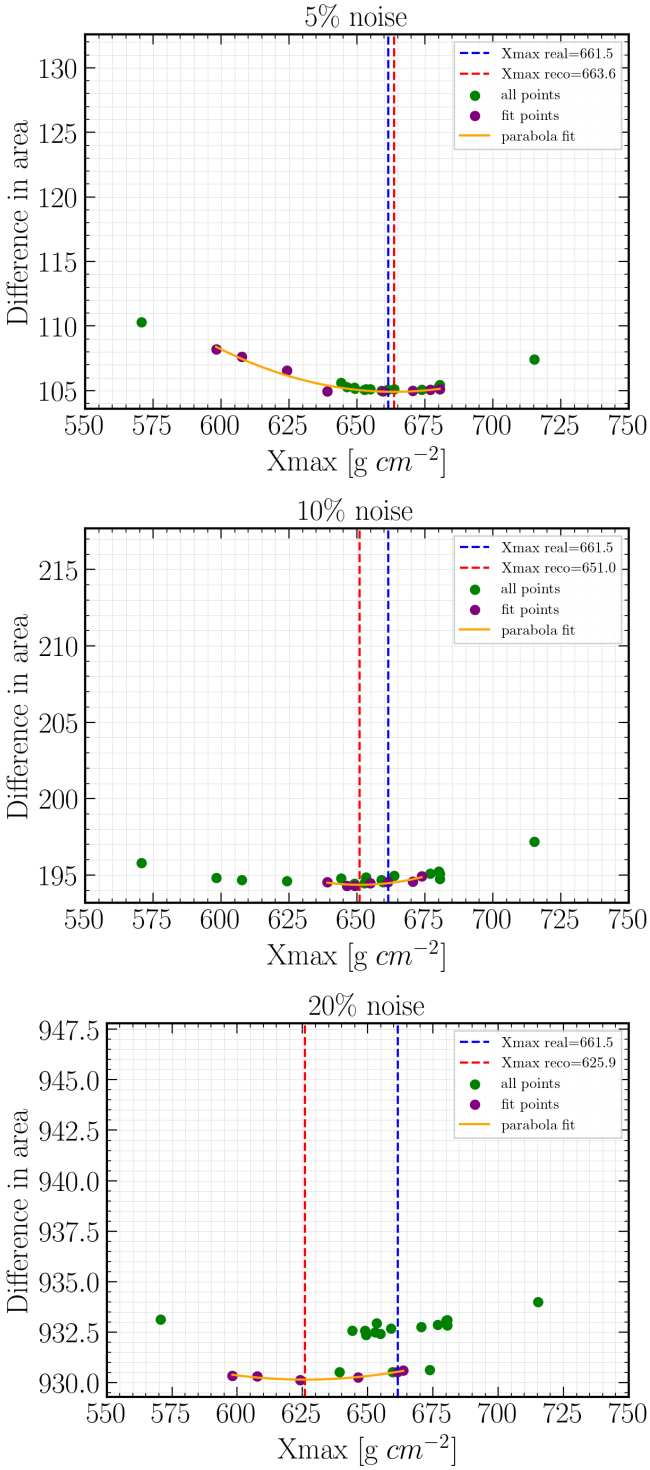


Figure 92: The parabola fitting procedure under the presence of different levels of generated noise. From top to bottom according to the noise addition procedure described in the text, $5\%, 10\%$ and 20% of noise is added to the simulations. The fitting procedure combined with the "corrected" fluence percentage formula (eq. (39)) yields improved accuracy for all levels of noise.

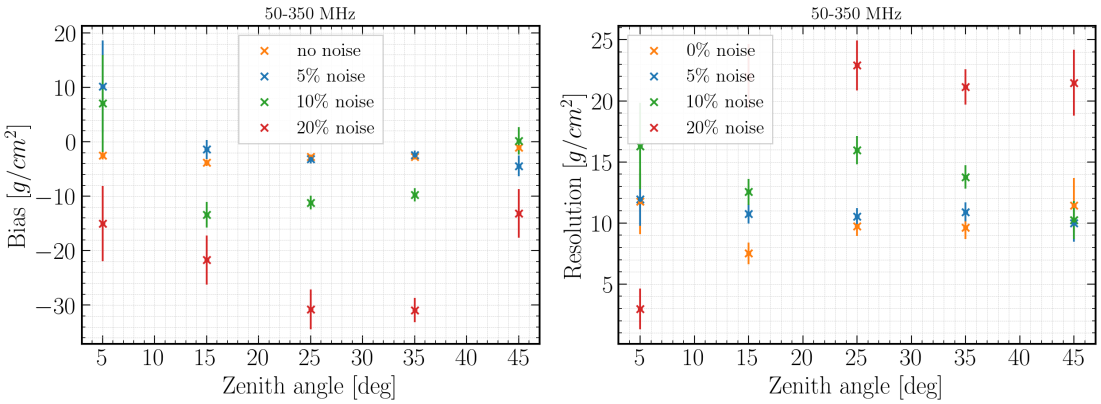


Figure 93: The simulated events are organized in zenith bins with a 10^o step for the 50 MHz to 350 MHz frequency band. The reconstruction scheme is tested for the case of no noise and for generated noise levels of 5%, 10% and 20%. On the left the bias is plotted and on the right the corresponding resolution is shown.

combination. As resolution, the standard deviation of the distribution of bias values is considered. The same procedure takes place for all 148 simulated events. Finally, all the events are grouped together with respect to their zenith angle in 10^o intervals.

Benchmark results (50 MHz to 350 MHz)		
Noise level	Bias ($\mathrm{g\,cm^{-2}}$)	Resolution ($\mathrm{g\,cm^{-2}}$)
No noise	-2.7	9.7
5% noise	-2.2	10.7
10% noise	-9.0	14.2
20% noise	-27.6	21.1

Table 23: The weighted average values of the bias and the resolution for the 50 MHz to 350 MHz frequency band are summarized in this table. The weight is dictated by the number of events in each zenith bin.

The reconstruction scheme was benchmarked for 2 frequency bands, namely 30 MHz to 80 MHz and 50 MHz to 350 MHz. The bias and corresponding resolution are shown in fig. 93 and in fig. 94 respectively. The error bars indicate the standard deviation over the number of simulations on that bin. It can be observed that the results are better for the zenith angles bins of 15^o, 25^o, 35^o and 45^o, while for the 5^o the reconstruction seems to not work that well. This happens because the events simulated with zenith angle $\leq 10^o$ are very few, 7 to be more precise. For larger zenith values the events are significantly more, which constitutes a better statistical sample. The weighted average of the zenith bins where as weight the number of events is defined, are also summarized in table 23 and table 24.

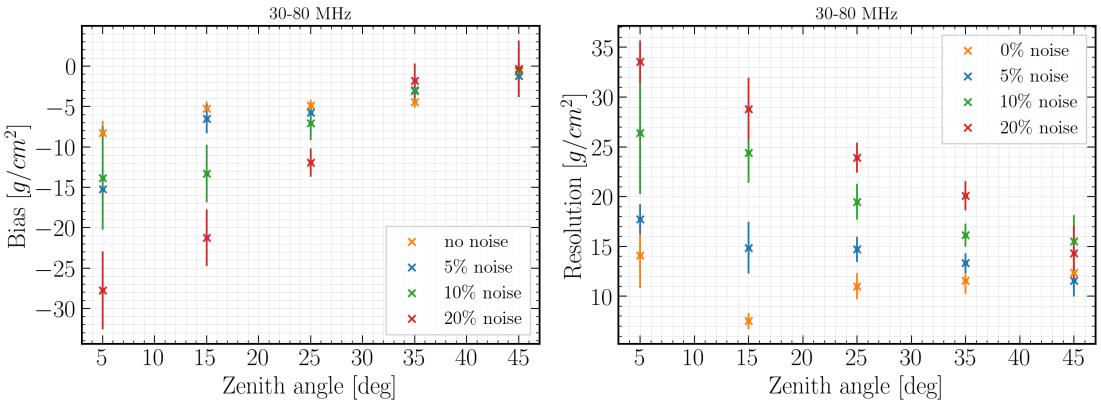


Figure 94: The simulated events are organized in zenith bins with a 10^o step for the 30 MHz to 80 MHz frequency band. The reconstruction scheme is tested for the case of no noise and for generated noise levels of 5%, 10% and 20%. On the left the bias is plotted and on the right the corresponding resolution is shown.

Benchmark results (30 MHz to 80 MHz)		
Noise level	Bias ($g\,cm^{-2}$)	Resolution ($g\,cm^{-2}$)
No noise	-4.5	11.1
5% noise	-4.9	14.1
10% noise	-5.9	18.8
20% noise	-8.9	22.6

Table 24: The weighted average values of the bias and the resolution for the 30 MHz to 80 MHz frequency band are summarized in this table. The weight is dictated by the number of events in each zenith bin.

The results as summarized in table 23 and table 24 are encouraging. One needs to take into account that the number of times that the reconstruction scheme took place for the library of events, amounts to over 3500. This implies that an automated algorithm was implemented and the parabola fitting procedure could not be inspected manually. When manually inspecting every parabola fitting one is able to perform optimizations such as a reasonable choice of the "grammage window" and to confirm that there is a cluster of simulations around the *reference* $X_{\max}$. For these reasons, a universal "grammage window" of $100\,g\,cm^{-2}$ was chosen for the fits. In a real world analysis, where one has experimental data of 1 event, these parameters can be adjusted to give better results because one can generate a suitable amount of simulations.

This reconstruction scheme could potentially be used for both frequency bands that were inspected. This goes to show that there is indeed information contained in the pulse shape accessible through the "fluence ratio" in the frequency domain and that it can be used for $X_{\max}$ reconstruction. Although, the effect of noise cannot go unnoticed. The way the noise was taken

into account in this work is rather simplistic. However, an advanced method to extract the noise out of the measured pulses together with a good understanding of the antenna model used by SKA has the potential to lead to encouraging results.

The way this reconstruction scheme could be potentially used at its current state would be as a second order analysis of the standard χ^2 minimization reconstruction method already used in LOFAR. The χ^2 minimization gives a resolution in $X_{\max}$ of $\sim 17\,\mathrm{g\,cm^{-2}}$ [49]. A cluster of simulations can be produced with $X_{\max}$ values close to the estimated $X_{\max}$ by the χ^2 minimization procedure for the specific antenna locations of the experiment. In the case of SKA, the antennas are of the order of tens of thousands which would of course make the production of these simulations very computationally heavy. The way to overcome this could potentially be to produce simulations with standard star-shaped antenna arrays and use a pulse interpolation method to determine the pulses on the specific antenna locations. Such an interpolation method has already been developed for the LOFAR analysis [84].

Having presented, benchmarked and discussed potential use of the $X_{\max}$ reconstruction scheme it is instructive to discuss a few things that were tried before arriving to the final results. In [85], a very similar study was presented with 2 distinct differences. First, a different more simplistic parabola fitting method was used. Second, the contribution of noise to the fluence percentage was not accounted for in the form of the correction that was presented here. On the parabola fitting method used in [85], results were worse and improving on that was a major step. The most distinct discrepancy in the benchmarks presented in [85] was omitting to treat the noise in some way. This worsened the values for the bias and the resolution. Before coming up with the noise correction procedure, I also tried to take into account the effect of noise in other ways.

Instead of correcting for the noise of the *reference* simulation, I simply generated the same level of noise to all the simulations that were used for the comparisons. A smaller time window of 30 ns around the pulse's peak was chosen this time. The idea of using a smaller time window is that the fluence percentage would capture less contributions from the noise. Since the window is small though, this was tested in the 50 MHz to 350 MHz band. The benchmark where the noise was not corrected and simply added is shown in fig. 95 and the weighted mean averages of bias and resolution are summarized in table 25. Similarly, the benchmark where the noise is corrected for the smaller 30 ns window is shown in fig. 96 and the weighted mean average bias and resolution values can be found in table 26. The point of presenting these alternative methods of treating the noise is to showcase once more that a more sophisticated and effective way of treating noise, renders this $X_{\max}$ reconstruction scheme a promising technique.

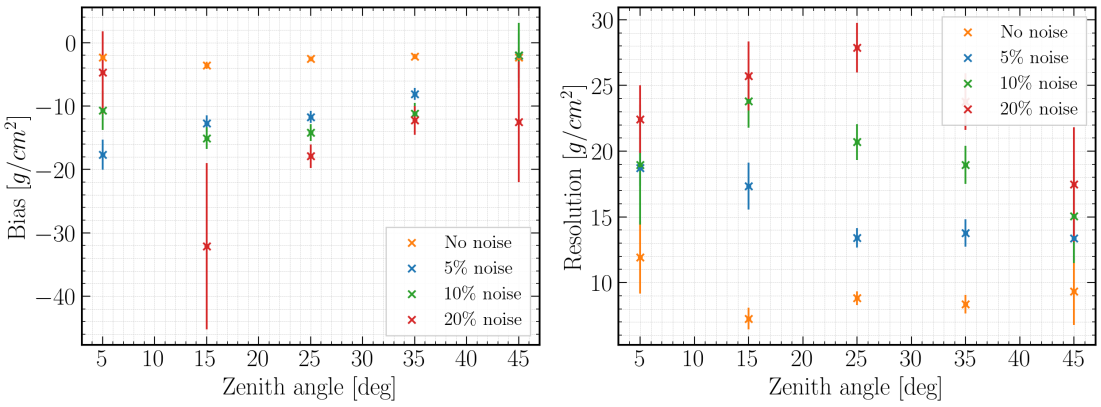


Figure 95: The simulated events are organized in zenith bins with a 10^o step for the 50 MHz to 350 MHz frequency band. The reconstruction scheme is tested for the case of no noise and for generated noise levels of $5\%, 10\%$ and 20%. On the left the bias is plotted and on the right the corresponding resolution is shown. In this benchmark, the noise is not "corrected", it is rather added to all simulations. Hence "noisy" simulations with the same level of noise are compared among each other. The time window around the pulse's peak is chosen to be 30 ns.

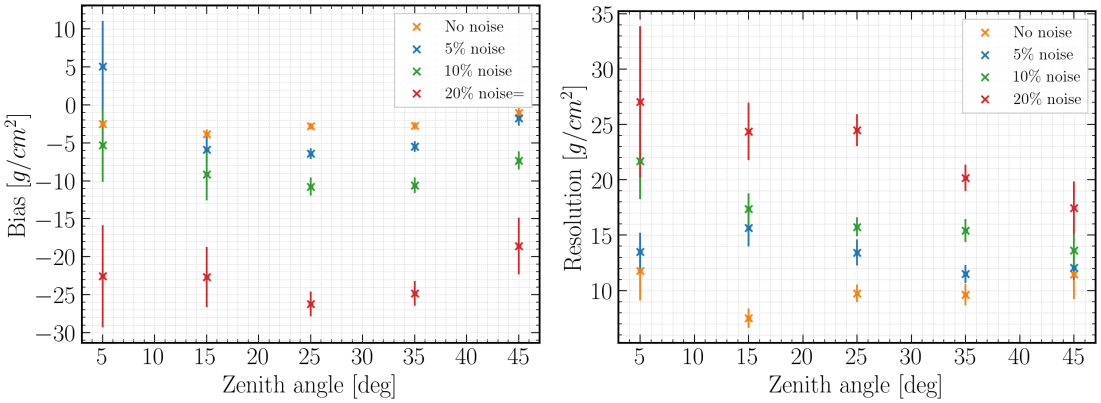


Figure 96: The simulated events are organized in zenith bins with a 10^o step for the 50 MHz to 350 MHz frequency band. The reconstruction scheme is tested for the case of no noise and for generated noise levels of $5\%, 10\%$ and 20%. On the left the bias is plotted and on the right the corresponding resolution is shown. In this benchmark, the noise is "corrected", according to eq. (39). The time window around the pulse's peak is chosen to be 30 ns instead of 55 ns that is shown in fig. 93.

Benchmark results (50 MHz to 350 MHz)		
Noise level	Bias ($\mathrm{g\,cm^{-2}}$)	Resolution ($\mathrm{g\,cm^{-2}}$)
No noise	-2.5	8.7
5% noise	-9.9	14.2
10% noise	-11.9	19.8
20% noise	-16.4	24.9

Table 25: The weighted average values of the bias and the resolution for the 50 MHz to 350 MHz frequency band are summarized in this table. The weight is dictated by the number of events in each zenith bin. In this benchmark, the noise is not "corrected", it is rather added to all simulations. Hence "noisy" simulations with the same level of noise are compared among each other. The time window around the pulse's peak is chosen to be 30 ns.

Benchmark results (30 MHz to 80 MHz)		
Noise level	Bias ($\mathrm{g\,cm^{-2}}$)	Resolution ($\mathrm{g\,cm^{-2}}$)
No noise	-2.6	10.0
5% noise	-2.9	13.2
10% noise	-8.6	16.7
20% noise	-22.9	22.7

Table 26: The weighted average values of the bias and the resolution for the 30 MHz to 80 MHz frequency band are summarized in this table. The weight is dictated by the number of events in each zenith bin. In this benchmark, the noise is "corrected", according to eq. (39). The time window around the pulse's peak is chosen to be 30 ns instead of 55 ns that is shown in fig. 93.

11.5 Antenna diagnostics utilizing the pulse shape information

Before arriving to the reconstruction scheme that takes advantage of the fluence percentage, the information of the pulse shape was studied in the time domain. The main tool for that was the Pearson correlation coefficient (r_{xy}) [86] defined for a sample in the formula below:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{41}$$

where $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean and $\overline{y}$ is defined correspondingly. A mathematical property of eq. (41) is that it is invariant under separate changes in location and amplitude of the variables x and y. So when comparing 2 pulses, one can slide one of them to match the other pulse's peak and calculate r_{xy}. This effectively takes out the amplitude information and correlates 2 pulses in terms of their shape.

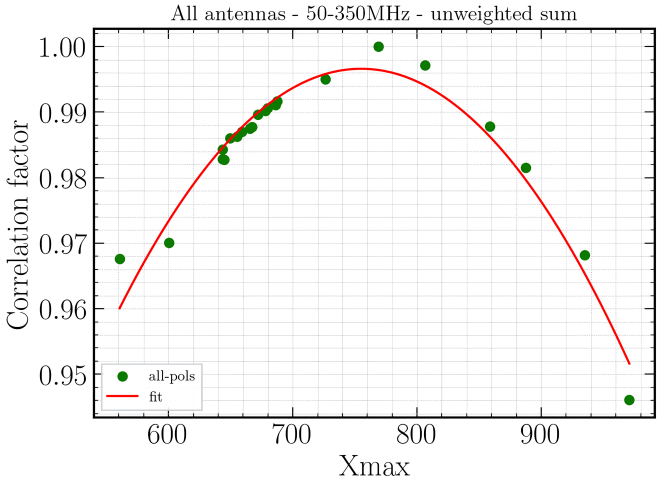


Figure 97: Parabola fitting scheme using the correlation factor eq. (41) as a metric. The *reference* simulation is left intentionally in this plot to showcase that the correlation factor is 1 when the *reference* is compared with itself. Naturally, not excluding this point affects the fitting procedure favorably.

The Pearson correlation coefficient was considered as a metric to identify pulses that matched on $X_{\max}$. A library of star-shaped antenna array simulations were used to test this hypothesis in an identical way as the one presented in the previous sections. Many simulations with identical settings that differ only in random number seed were used. One of them was treated as the *reference* simulation. Each polarization from each pulse from each antenna location was compared to the corresponding antenna location of the rest of the simulations. So, for each antenna of each simulation there are 3 correlation coefficients (also mentioned as correlation factors), one per polarization. Their mean value of all polarizations per antenna is calculated. Then each simulation has a number of correlation factors equal to the number of antennas. Once again, the mean value of the correlation factors of all antennas per simulation is calculated. As a result, in each simulation corresponds a correlation factor. The closer this factor is to 1 the closer the simulation is to the *reference* $X_{\max}$. An example study of what was just described is shown in fig. 97. These are simulations with no noise included and it is already evident that such an idea provides a clear structure with the correlation factors close to 1 being the ones closer to the *reference* $X_{\max}$ indeed. One could then fit a parabola and treat its maximum as the reconstructed $X_{\max-\mathrm{reco}}$. While this looks promising, it was quickly realised that the correlation values were very close numerically. In real world scenarios in the presence of noise such a simplistic approach would probably not survive, and hence this procedure was not studied further with the addition of noise.

An interesting idea arose though through inspecting what one could do with the Pearson correlation coefficient. For this a few measured events from LOFAR and their corresponding simulations were used. When the event is identified and most of the detected shower properties,

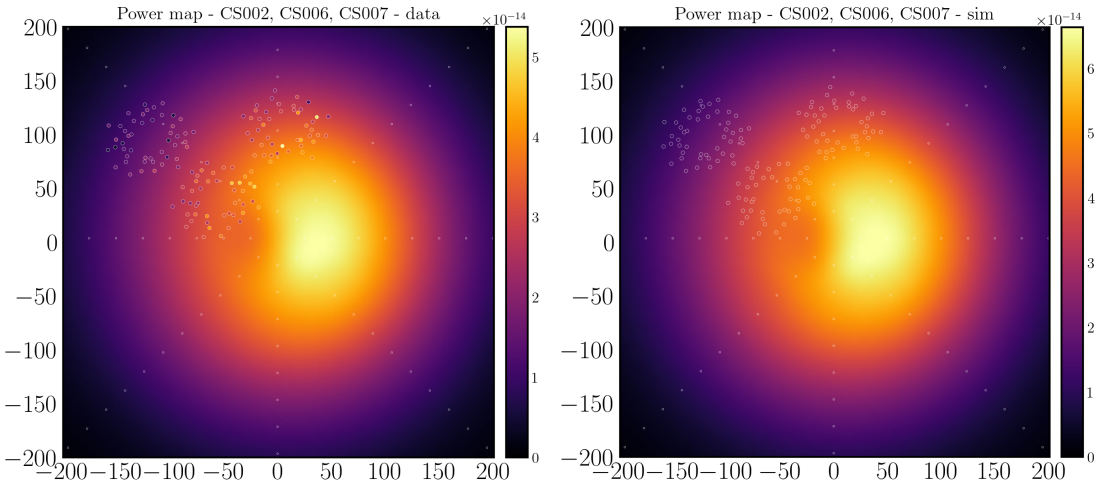


Figure 98: On the background of both simulation the simulated time integrated power map is shown. The dots are the antenna locations from the triggered LOFAR stations. On the left, the dots are the the reconstructed measured data and one can compare how well the colors of the dots match the background. On the right, the dots are the simulated data and as expected their colors bled completely in the background.

simulations with the exact same settings and $X_{\max}$ are generated for the exact antenna locations for the LOFAR stations that were triggered. One can then fit on top of the ideal simulation the power detected by the antennas triggered. Time integrated power per antenna in this context is defined as the following quantity:

$$P_i = \epsilon_0 c \Delta t \sum_{t_{start}}^{t_{end}} V_i(t)^2 \tag{42}$$

where $V_i(t)$ is the antenna response measured in Volts. This is a very similar formula to eq. (11) and its meaning is also very similar. Instead of measuring fluence as energy deposited to the antenna on the ground coming for the electric field, the antenna response is used. Some example power maps are shown in fig. 98. The dots shown in fig. 98, represent the antenna locations. On the background of both plots, a star-shaped antenna array simulation with interpolated values between the antennas is plotted. On top of that, data from the reconstructed measurements in the triggered antennas are plotted (left) and simulations of the antenna specific locations (and not a star-shape) that have been folded with the LOFAR antenna model are shown (right). As was anticipated, on the right the dots match very well with background, while on the left the agreement is very good but one can spot that for some locations the dots do not match completely the background.

Maps like fig. 98 can be made for the Pearson correlation coefficient as well. Once the reconstructed measurements and the corresponding simulations that include also the antenna

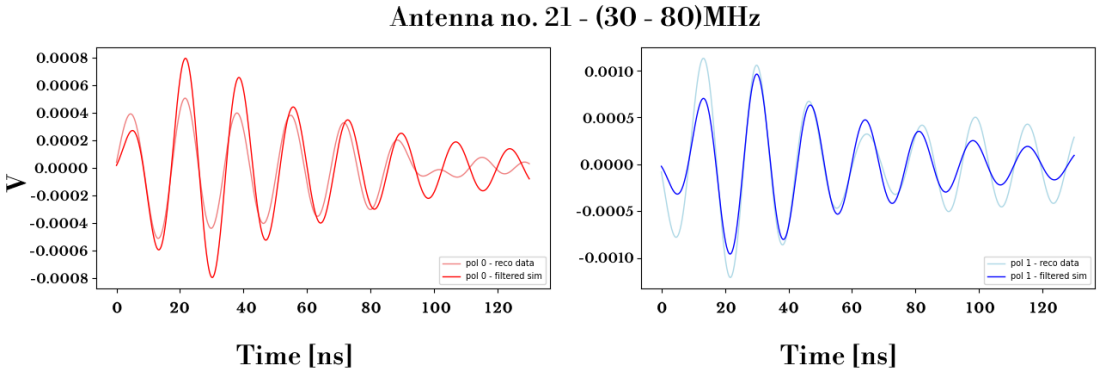


Figure 99: A random antenna from a random LOFAR station that was triggered for an event is selected. On the left, polarization 0 (pol 0) is plotted and on the right, the dominant polarization (pol 1) is plotted. With the bold colors the reconstructed measurements are shown, while the simulations folded with the antenna response are indicated with the faint colors. The purpose of the correlation factor is to inspect how well measurements and simulations match in terms of pulse shape.

response are in place, the following procedure can happen. The pulse coming from the cosmic ray event is identified in the reconstructed data. The event is usually being put in the middle of the reconstructed data once the pipeline procedures take place. Using r_{xy} one can identify the event. The simulated trace is utilized with a time window of 55 ns around the maximum. One can now choose a time window around the "middle" of the reconstructed measurement and scan the simulated trace in a 0.1 ns step, which corresponds to the sampling period. For each step, a correlation factor and its time information is calculated and stored. The value with the highest correlation factor is considered the best match between measurement and simulation. An example of this "pulse matching" for both polarizations of a random antenna of a random LOFAR station is shown in fig. 99. In the dominant polarization (annotated as pol 1 here) the signal is stronger and the correlation factor is usually higher. A high correlation factor indicates also a good understanding of the antenna model. One could use the same procedure for different antenna models to test how well they agree with measurements. One could also use this pulse matching procedure for all antennas of all triggered stations to make correlation factor maps fig. 100. Exact agreement is designated with the red color on the background (correlation factor is equal to 1). The dots indicate the antenna positions, and how well their color blends with background shows how well the measurements agree with simulations. An interesting use case of such maps would be for antenna diagnostics. Especially for the case of SKA which consists of thousands of antennas and is expected to have a large increase in the number of events, making a color map at the end of reconstruction of every event could be a very useful way to potentially reveal antenna malfunctions. If the correlation factor of a specific antenna location is consistently below a specified value, then that could indicate malfunction or even damage to that antenna. This would be an interesting, visual and easy way to estimate the operational state of antennas all at once.

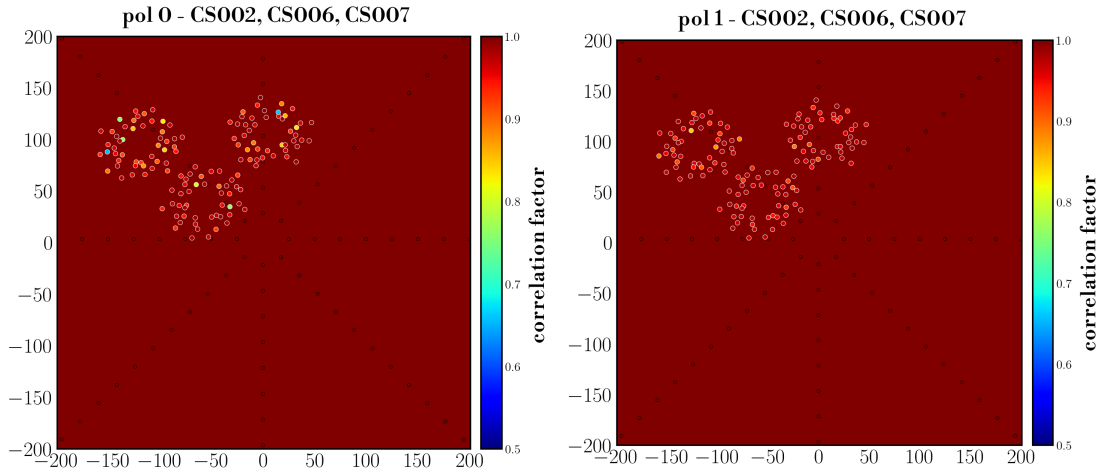


Figure 100: A perfect agreement means that the correlation factor has a value of 1 and this is depicted by the red color plotted on the background. Every colored dot, represents the correlation factor for every antenna. On the left, the correlation factors for polarization 0 are plotted and on the right for polarization 1. Signals are stronger in the dominant polarization (right) and their correlation factors have higher values. Stronger signals are easier to be identified and the calculation of the correlation factor is more optimal in that case.

11.6 Discussion

In this part of the thesis the signal pulses emitted from air showers were inspected regarding the information contained on their shape. So far, for $X_{\max}$ reconstruction only the information in fluence is used. In this work, a reconstruction scheme was devised through a simulation study that does not rely on fluence, rather in the pulse shape information. Considering the ideal case where no noise is present, the reconstruction scheme showcases promising results. When noise is added though, one needs to take it into account and correct for it. With advanced techniques to identify and correct for the noise the scheme shows encouraging results. Such a scheme could be used as a second order analysis complementary to the standard χ^2 minimization technique to potentially improve the accuracy of $X_{\max}$ reconstruction. During this study another avenue was investigated, where a procedure that uses the pulse shape was developed that could be potentially used for antenna diagnostics.

12 Summary and Outlook

The growing experimental needs and the advent of new technologies constitute modern and reliable software crucial for our understanding of cosmic rays. The focus of this thesis was the radio emission from Extensive Air Showers and the establishment that it is a well understood process through simulations. For this, a radio module as an integral part of CORSIKA 8 that utilizes modern technologies like multithreading was implemented, validated and can be used by the scientific community. The design choices ensure flexibility and upgradability making it future-proof and adjustable to the modern experimental needs. The development of this module contributed to CORSIKA 8 and pushed the project as a whole. It was used to diagnose and find logical errors and bugs in the core code. Architectural changes took place that overall improved the accuracy of CORSIKA 8. Due to the sensitive nature of the radio emission extensive comparisons between CORSIKA 8 and CORSIKA 7 were made. It was found that for the default value of the track length of the simulated particle tracks the two codes can differ up to more than $\sim 50\%$ in radiation energy. However, if the simulations are done finely enough, i.e. small enough tracks, it was found that CORSIKA 7 and CORSIKA 8 agree within $\sim 8.3\%$ in the 30 MHz to 80 MHz frequency range and within $\sim 1.7\%$ in 50 MHz to 350 MHz. Given the preliminary state of CORSIKA 8 this is considered as a good result because it means this leads to a systematic uncertainty of less than 5% when using radio. Calibration uncertainties are of order $10 - 15\%$, so this will not be the dominating uncertainty. Suggestions on further studies to improve the agreement between the two simulation codes were given.

The Endpoints and ZHS formalisms were implemented directly into the radio module. This allows for a direct comparison of the two on the exact same shower. In the past, comparisons that were made between the two formalisms were burdened by the underlying uncertainties of the different simulation codes that were used. In this thesis, a definite answer on their level of agreement is given. It was shown that for simulations made with fine enough tracking, the Endpoints and ZHS formalisms as implemented in CORSIKA 8, practically converge with an agreement in radiation energy of less than $\sim 2\%$. The fact that two formalisms derived from first principles with different assumptions agree to that extent solidifies the idea that the radio emission calculation is well understood. Studying the effect of different track lengths in radio simulations for both CORSIKA 8 and CORSIKA 7, revealed indeed that both the radiation energy and the fluence are affected by it. The radiation energy is essential for setting an energy scale and the fluence is used to determine $X_{\max}$. Apart from the differences in radiation energy, extensive fluence ratio maps for different track lengths and different formalisms were shown. It is shown that the signal distribution does change with different track lengths for both formalisms. Considering that there is a discrepancy between LOFAR and AERA on their $X_{\max}$ results, and that for their analyses the standard value of track length in CORSIKA 7 was used, it could be worth considering to redo their analysis with simulations that utilize very small, fine tracks.

Finally, in this thesis an alternative $X_{\max}$ reconstruction scheme was developed that utilizes the information contained on the pulse shape. On standard χ^2 minimization reconstruction

procedures the pulse shape information is discarded. In this work, it was shown that the pulse shape information can indeed be used for X_{max} reconstruction with encouraging results. It is suggested to be used as a complementary method χ^2 minimization in order to potentially improve the accuracy of the overall reconstruction.

Appendix

For the purposes of this thesis libraries of simulations were produced with CORSIKA 8 and CORSIKA 7 with different settings. In section 8.2, a total of 200 iron induced vertical air shower simulations generated with CORSIKA 8 were validated with 200 iron induced vertical air shower simulations generated with CORSIKA 7 for different track lengths. Half of these showers from both simulation codes were generated with the maximally allowed magnetic deflection angle set to 0.2 rad and the other half set to 0.001 rad. The level of agreement between CORSIKA 8 and CORSIKA 7 was inspected for both values.

In section 9, the 200 iron induced shower simulations were utilized to systematically compare the CoREAS and ZHS formalisms. Both formalisms were compared for both selected values of the maximally allowed deflection angle. In addition, for the same comparison 200 photon induced vertical air showers with the maximally allowed magnetic deflection angle set to 0.2 rad were generated with CORSIKA 8. For the first 100, the multiple scattering approximation that was used was Molière, while for the rest the Highland approximation was chosen. The level of agreement between the CoREAS and ZHS formalisms was also inspected for photon showers and different multiple scattering approaches.

In section 10, the CORSIKA 8 iron induced vertical air shower simulations with the maximally allowed magnetic deflection angle set to 0.2 rad were compared with the CORSIKA 8 iron induced vertical air shower simulations with the maximally allowed magnetic deflection angle set to 0.001 rad. The same comparison took place for the CORSIKA 7 generated simulations. Finally, the CORSIKA 8 photon induced vertical air shower simulations with multiple scattering approach set to Molière were compared with the CORSIKA 8 photon induced vertical air shower simulations with multiple scattering approach set to Highland, for both CoREAS and ZHS.

For all these simulations the energy of the primary particle was set to 1 PeV. The rest of the settings are listed on the chapters the libraries are presented. All these simulations were used to compare fluence and radiation energy throughout section 8, section 9 and section 10. In these sections the fluence ratio maps were presented in two frequency bands, namely 30 MHz to 80 MHz and 50 MHz to 350 MHz. For the sake of completeness and documentation purposes the mean values of fluence maps per group of showers that were compared, are plotted for total fluence, geomagnetic and charge excess emission in the figures below.

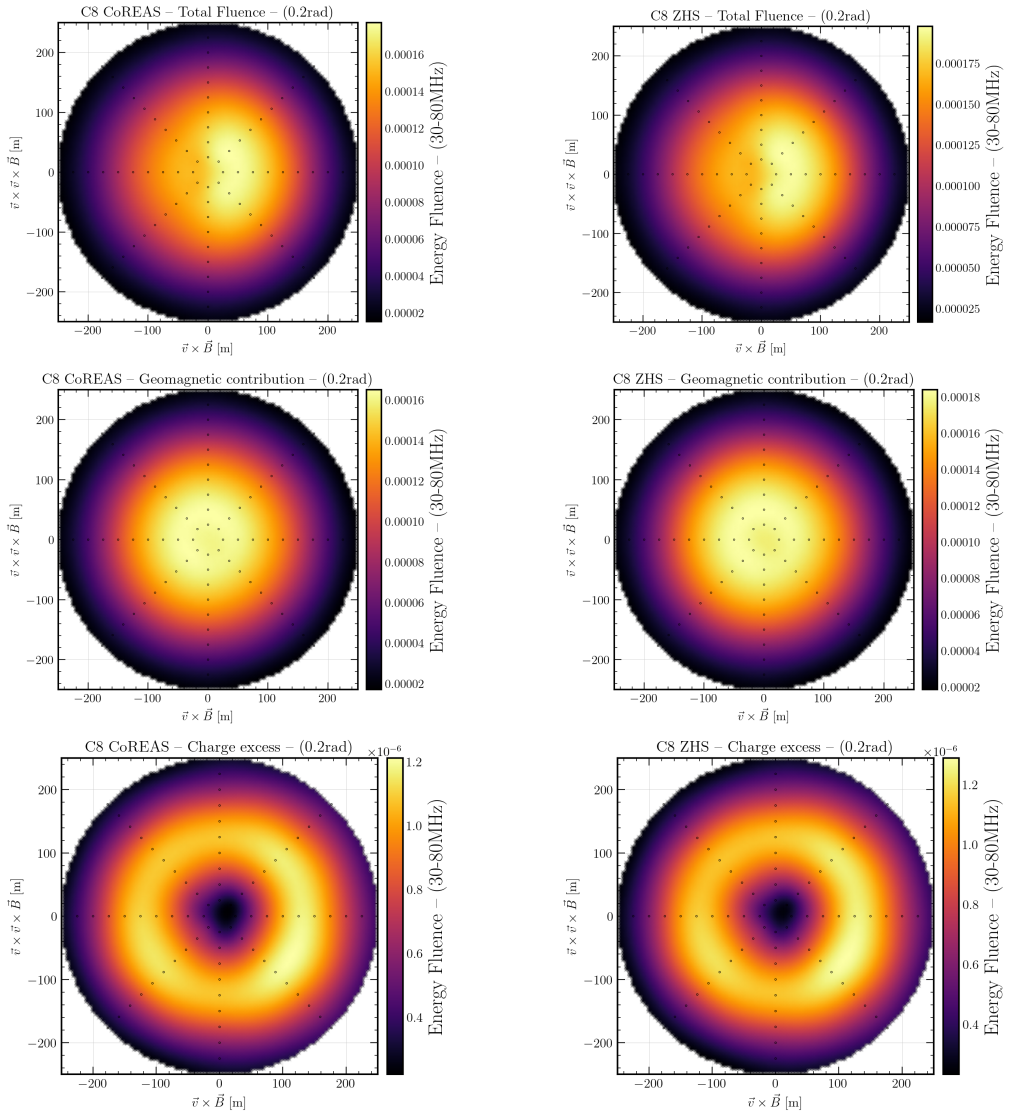


Figure 101: Fluence maps for 100, 1 PeV, iron induced vertical showers in the 30 MHz to 80 MHz band for CoREAS (left column) and ZHS (right column). In the first row, the total fluence map is plotted, the second the geomagnetic contribution and the last row the charge excess. The maximally allowed deflection angle is set to 0.2 rad.

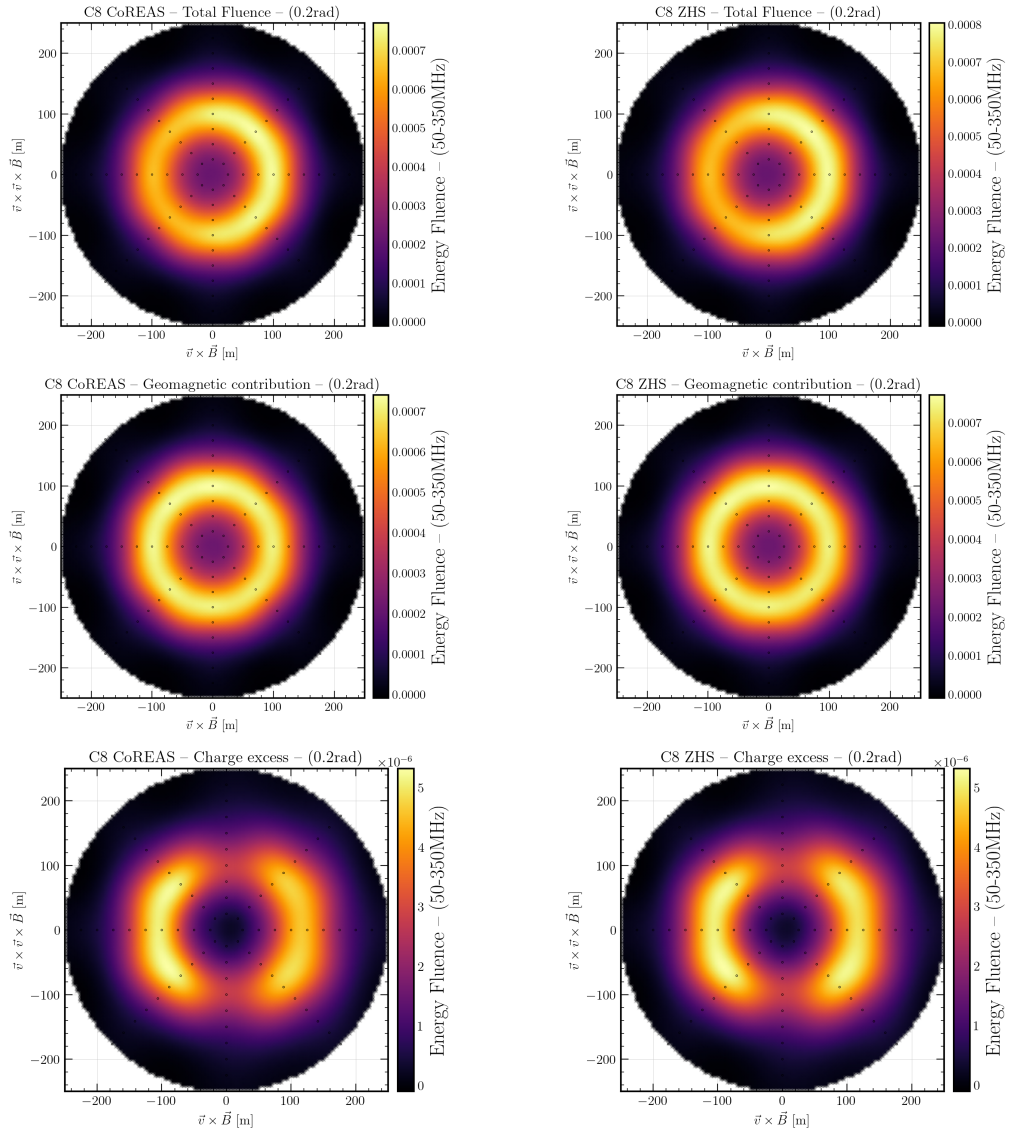


Figure 102: Fluence maps for 100, 1 PeV, iron induced vertical showers in the 50 MHz to 350 MHz band for CoREAS (left column) and ZHS (right column). In the first row, the total fluence map is plotted, the second the geomagnetic contribution and the last row the charge excess. The maximally allowed deflection angle is set to 0.2 rad.

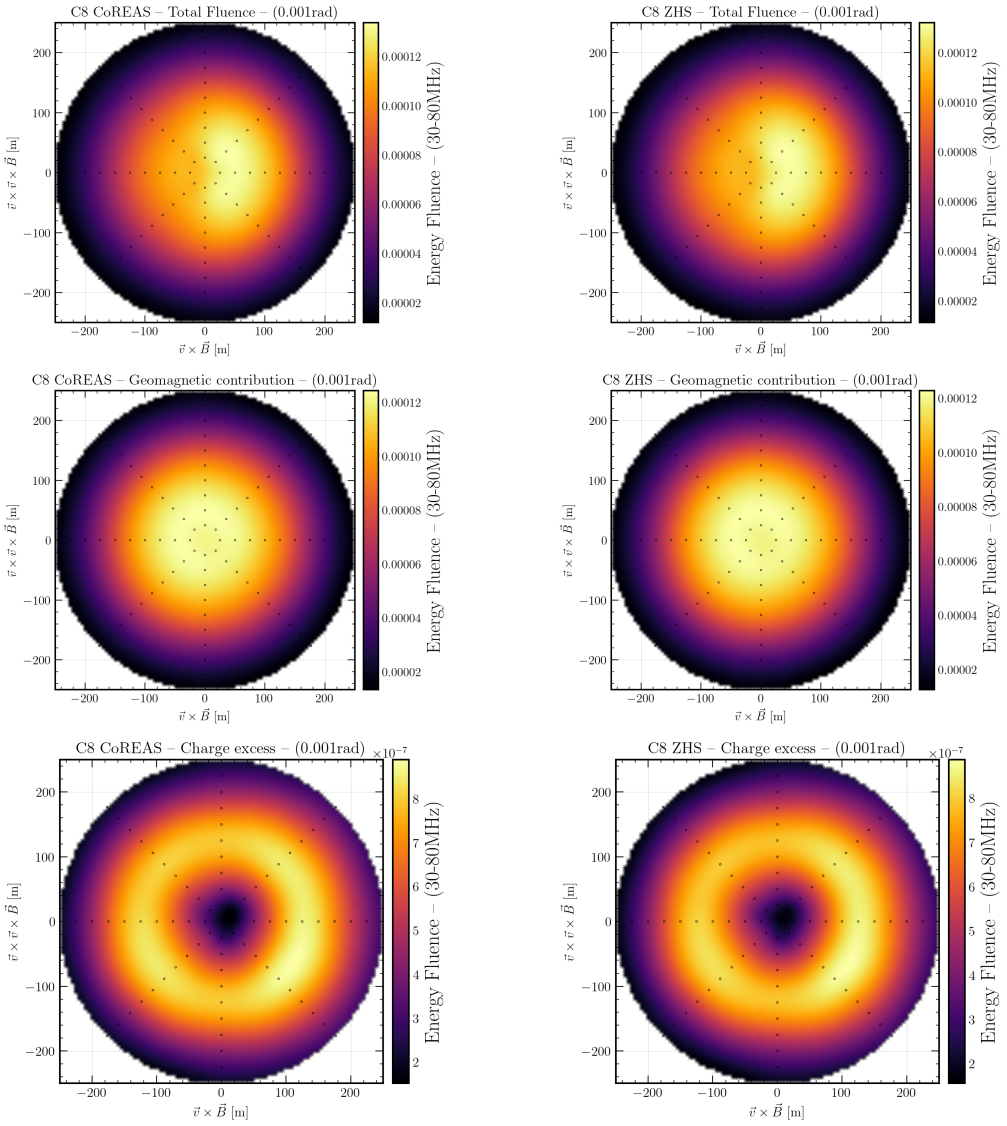


Figure 103: Fluence maps for 100, 1 PeV, iron induced vertical showers in the 30 MHz to 80 MHz band for CoREAS (left column) and ZHS (right column). In the first row, the total fluence map is plotted, the second the geomagnetic contribution and the last row the charge excess. The maximally allowed deflection angle is set to 0.001 rad.

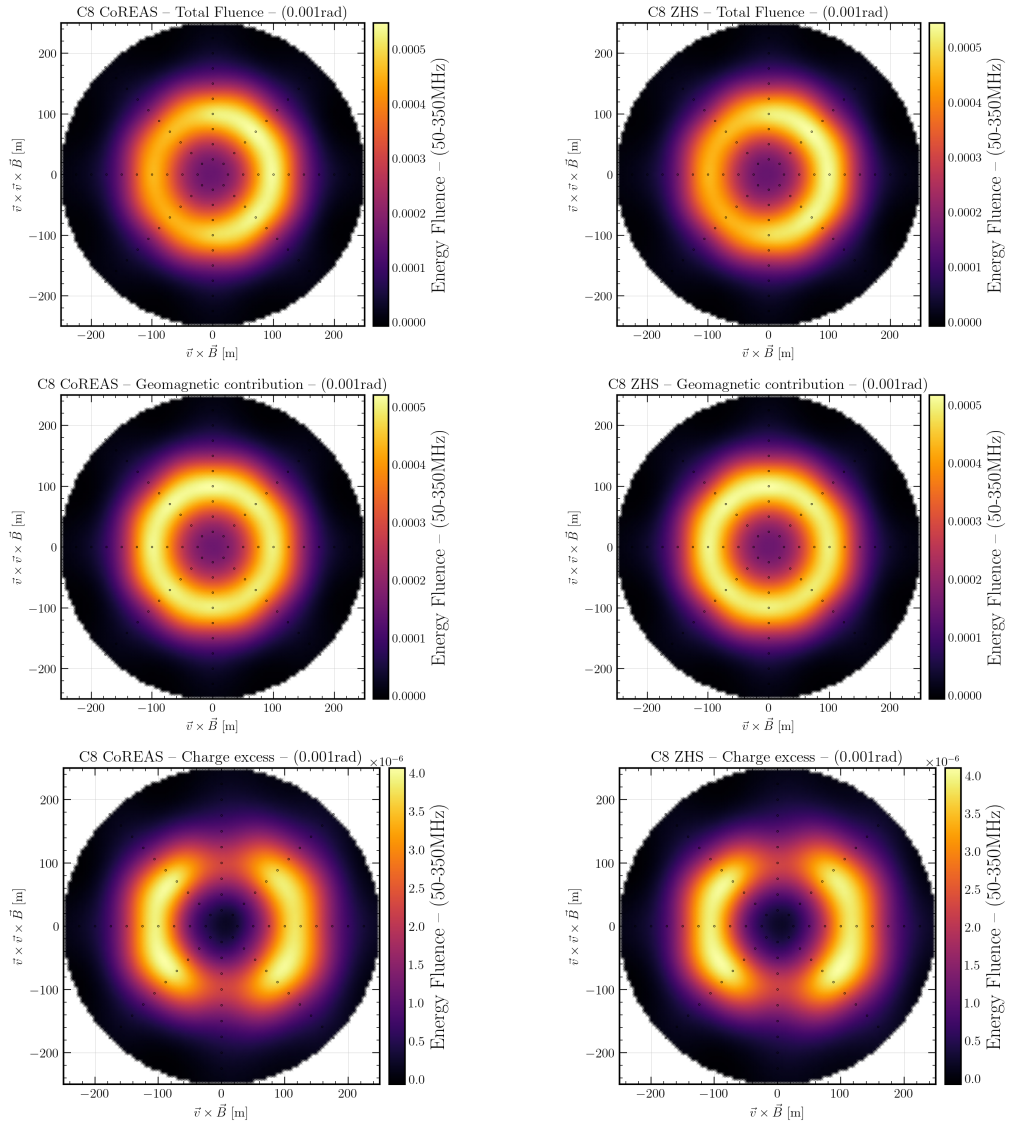


Figure 104: Fluence maps for 100, 1 PeV, iron induced vertical showers in the 50 MHz to 350 MHz band for CoREAS (left column) and ZHS (right column). In the first row, the total fluence map is plotted, the second the geomagnetic contribution and the last row the charge excess. The maximally allowed deflection angle is set to 0.001 rad.

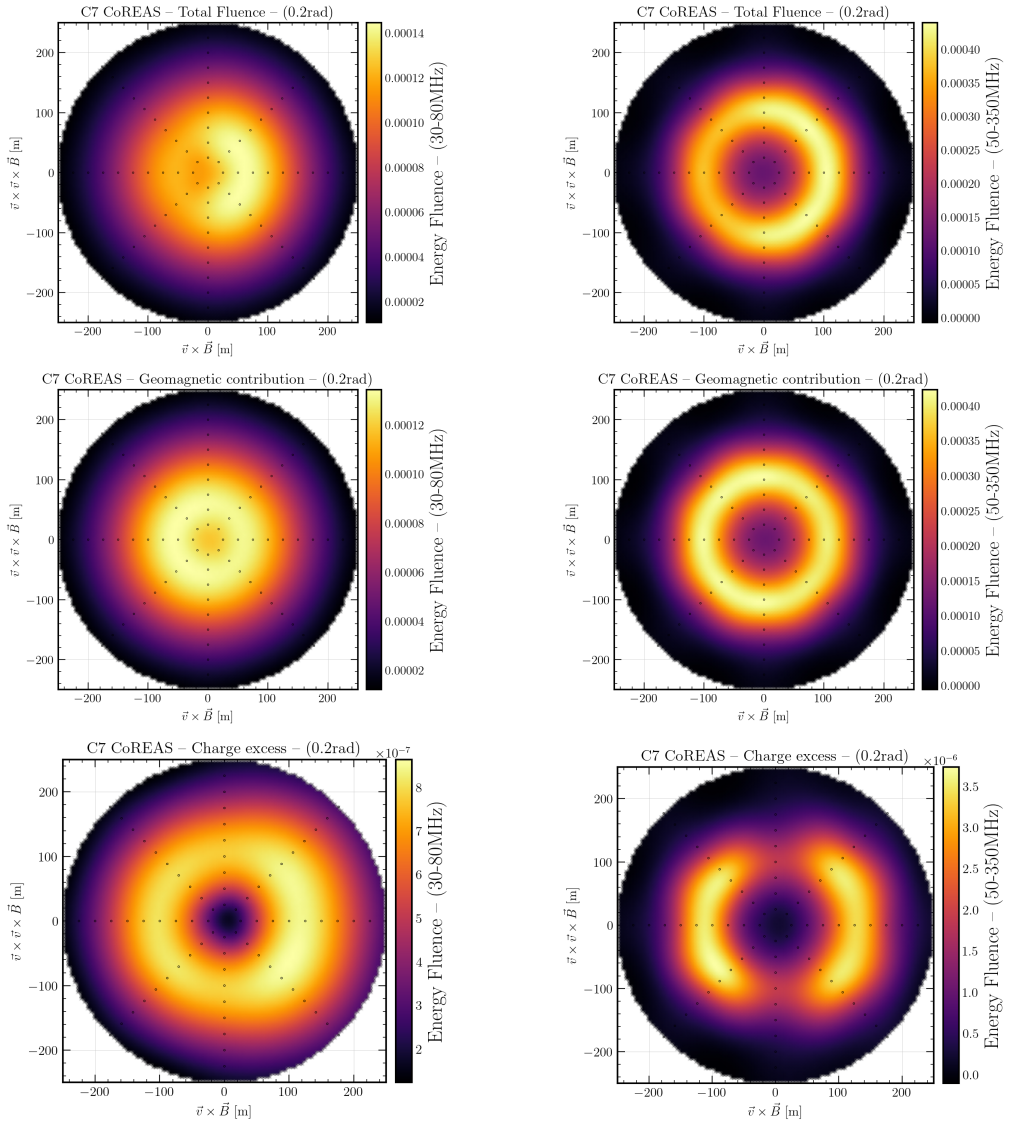


Figure 105: Fluence maps for 100, 1 PeV, iron induced vertical showers in the 30 MHz to 80 MHz (left column) and the 50 MHz to 350 MHz band (right column). In the first row, the total fluence map is plotted, the second the geomagnetic contribution and the last row the charge excess. The maximally allowed deflection angle is set to 0.2 rad and these showers were simulated with CORSIKA 7.

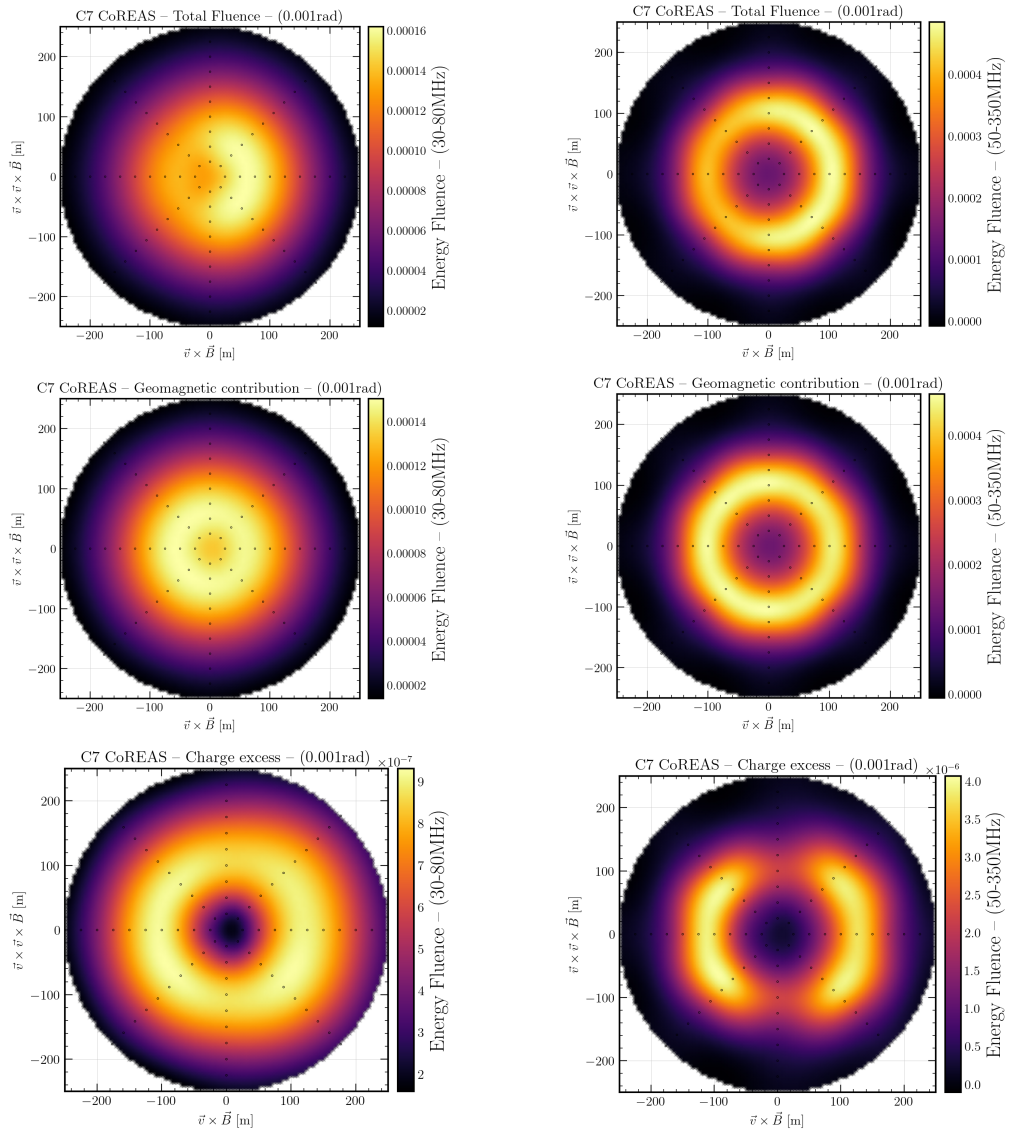


Figure 106: Fluence maps for 100, 1 PeV, iron induced vertical showers in the 30 MHz to 80 MHz (left column) and the 50 MHz to 350 MHz band (right column). In the first row, the total fluence map is plotted, the second the geomagnetic contribution and the last row the charge excess. The maximally allowed deflection angle is set to 0.001 rad and these showers were simulated with CORSIKA 7.

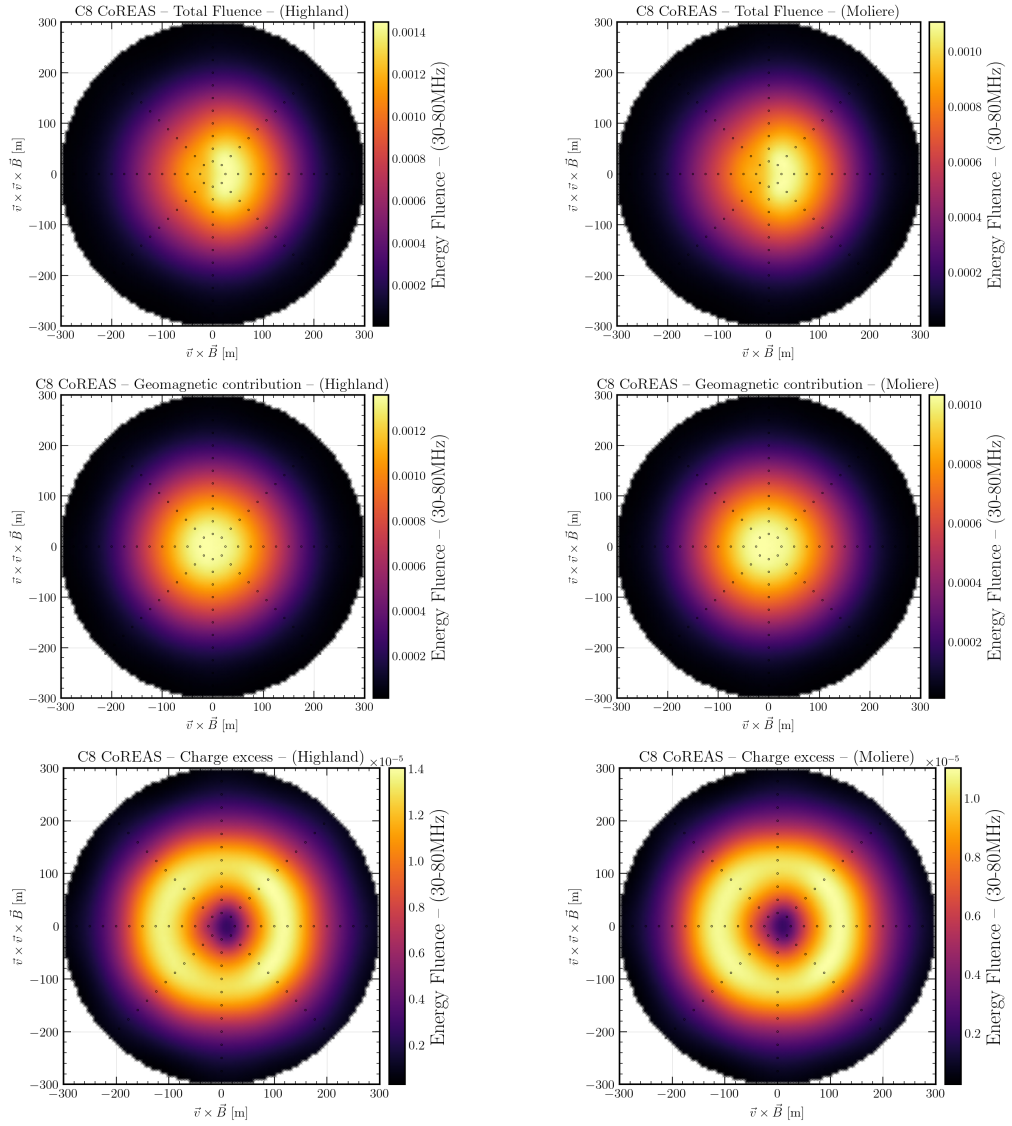


Figure 107: Fluence maps for 100, 1 PeV, photon induced vertical showers in the 30 MHz to 80 MHz frequency band. For the simulations, Highland (left column) and Molière (right column) multiple scattering mechanisms. In the first row, the total fluence map is plotted, the second the geomagnetic contribution and the last row the charge excess. The maximally allowed deflection angle is set to 0.2 rad and the radio emission was simulated with CoREAS.

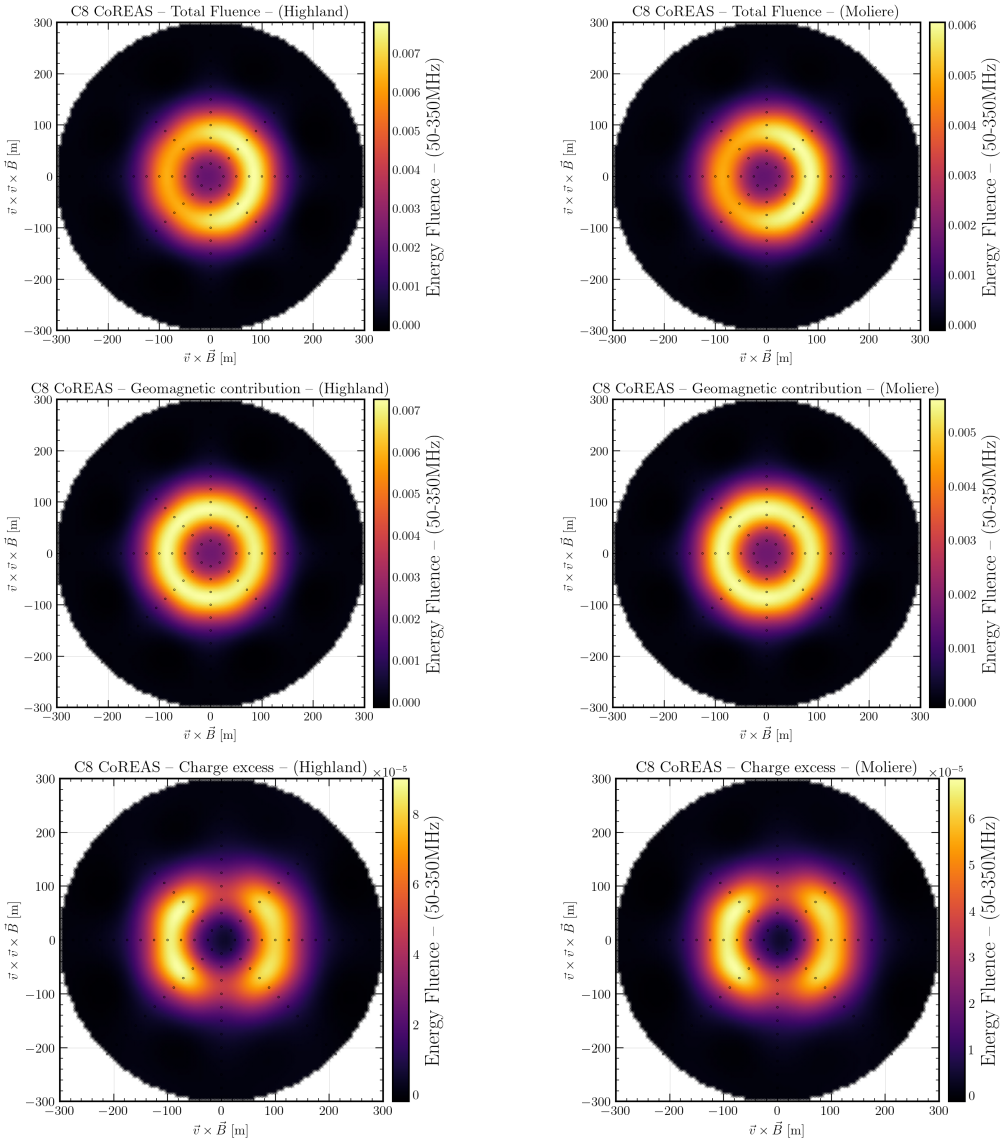


Figure 108: Fluence maps for 100, 1 PeV, photon induced vertical showers in the 50 MHz to 350 MHz frequency band. For the simulations, Highland (left column) and Molière (right column) multiple scattering mechanisms. In the first row, the total fluence map is plotted, the second the geomagnetic contribution and the last row the charge excess. The maximally allowed deflection angle is set to 0.2 rad and the radio emission was simulated with CoREAS.

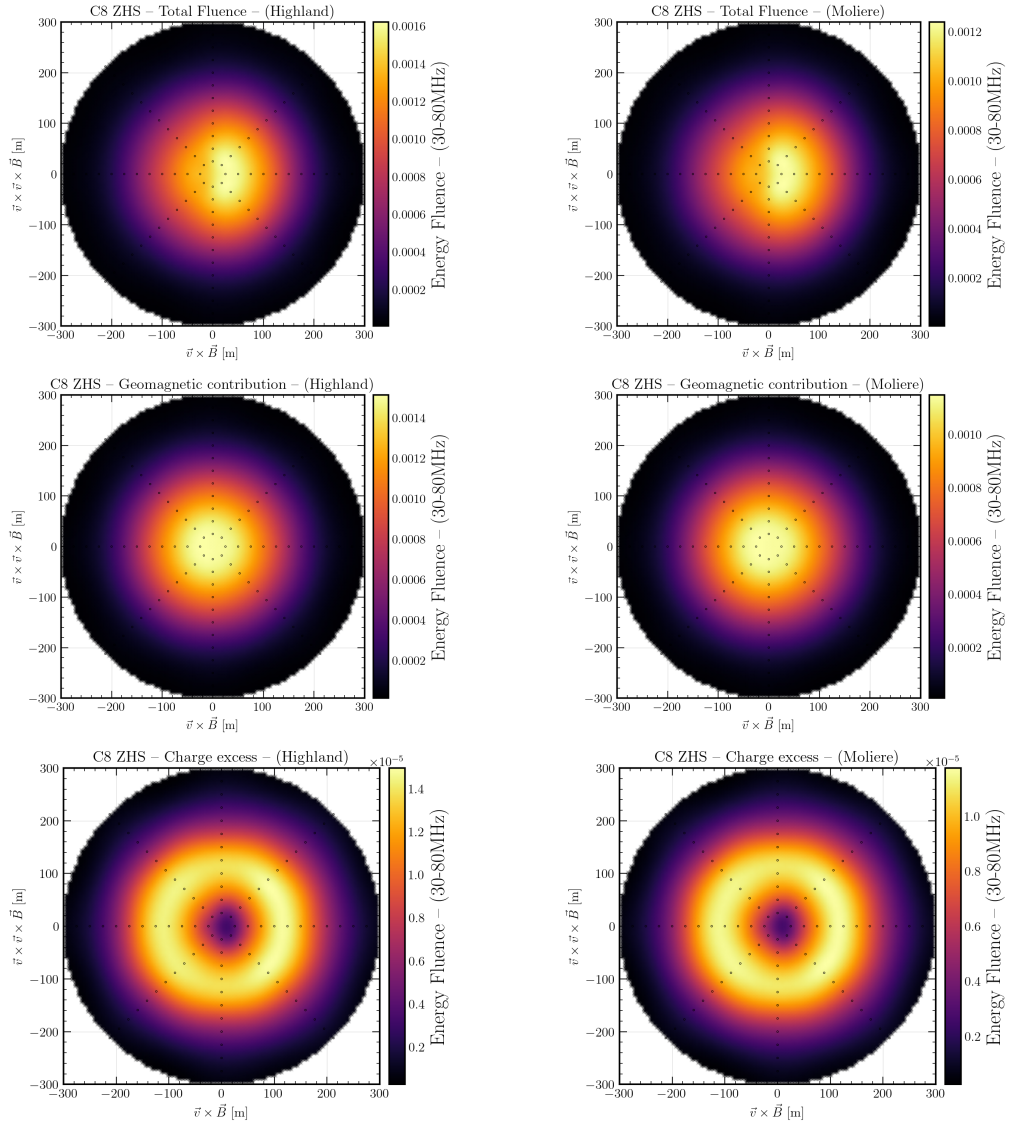


Figure 109: Fluence maps for 100, 1 PeV, photon induced vertical showers in the 30 MHz to 80 MHz frequency band. For the simulations, Highland (left column) and Molière (right column) multiple scattering mechanisms. In the first row, the total fluence map is plotted, the second the geomagnetic contribution and the last row the charge excess. The maximally allowed deflection angle is set to 0.2 rad and the radio emission was simulated with ZHS.

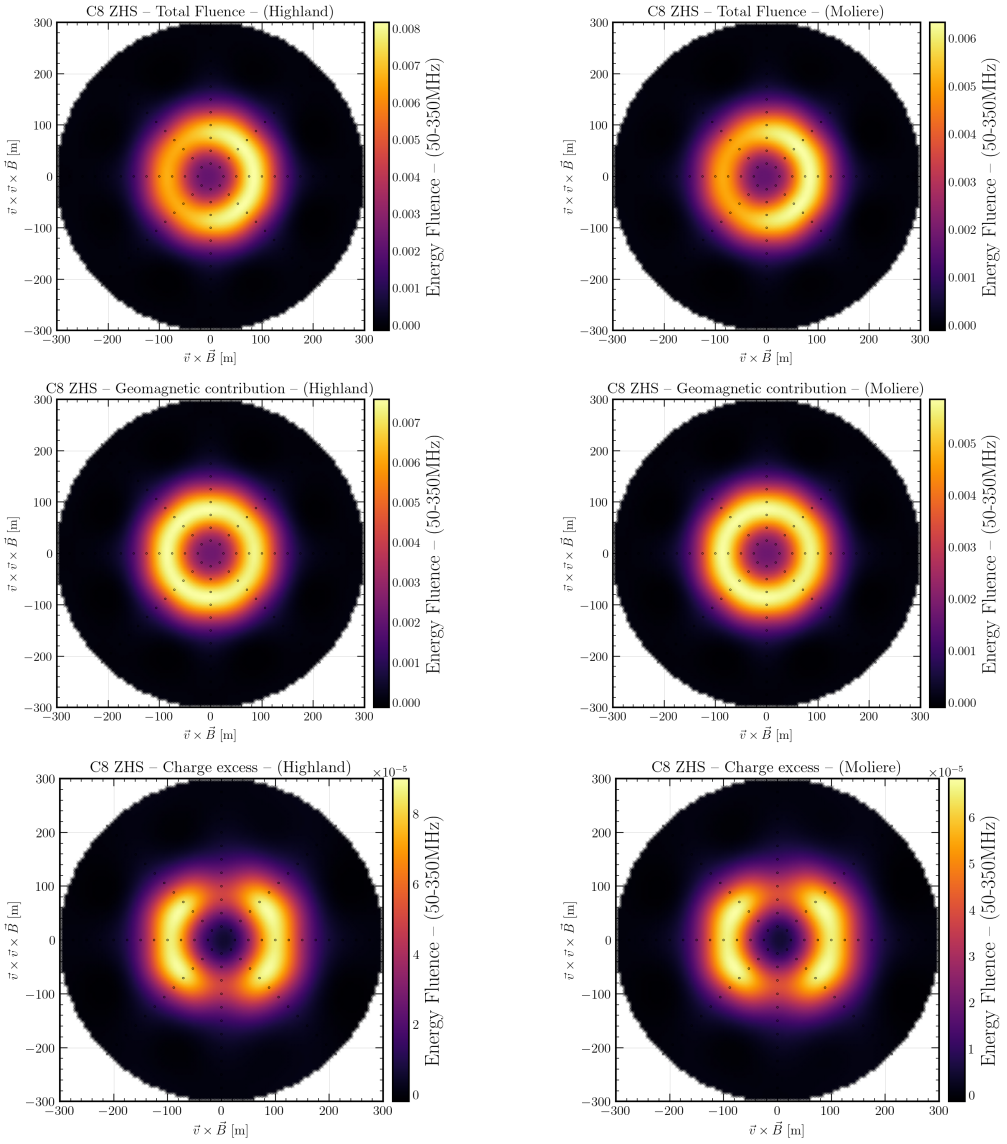


Figure 110: Fluence maps for 100, 1 PeV, photon induced vertical showers in the 50 MHz to 350 MHz frequency band. For the simulations, Highland (left column) and Molière (right column) multiple scattering mechanisms. In the first row, the total fluence map is plotted, the second the geomagnetic contribution and the last row the charge excess. The maximally allowed deflection angle is set to 0.2 rad and the radio emission was simulated with ZHS.

Acknowledgements

I would like to thank Professors Ralph Engel and Tim Huege for giving me the opportunity to work on such an exciting topic. I am truly indebted to Tim Huege who acted as my daily supervisor throughout my studies and was besides me on every step of the way. He was always happy to advise me, provide me his insights and share his knowledge.

I would also like to thank all the CORSIKA 8 collaborators that I worked with for all this time. It was our team effort that made CORSIKA 8 go so far. Special thanks to Remy Prechelt, Maximilian Reininghaus, Juan Ammerman-Yebra, Jean-Marco Alameddine and Antonio Augusto Alves Jr., fellows that I worked closely with and learned a lot from them through countless fruitful discussions.

I am very glad to say that I made new friends at the Institute of Astroparticle Physics in Karlsruhe, which I thank each one of them individually for their companionship. Kudos to Jelena Köhler for helping with the translation of the abstract in German and special thanks to Raphael Toscano for the shower illustration of the thesis cover.

Finally, I want to thank my family and my girlfriend Dina Karakosta for their infinite love and support throughout my endeavors.

References

[1] Thomas K. Gaisser, Ralph Engel, and Elisa Resconi. *Cosmic Rays and Particle Physics.* Cambridge University Press, 2 edition, 2016.

[2] A. M. Hillas. The Origin of Ultra-High-Energy Cosmic Rays. *Annual review of astronomy and astrophysics*, 22, 1984, 1984.

[3] Mauricio Bustamante, Pavel Je, J. A. Montanez, G. D. Carrillo Montoya, Giuseppe Romeo, Fátima Padilla Cabal, J A Duarte Chavez, Ulises Saldaña-Salazar, Maite Rozas, Mauricio Velasquez, Diego Julián Rodríguez Patarroyo, Antonio Ortiz Velasquez, Wayne de Paula, A. M. Gago, Hayk Hakobyan, and Marina von Steinkirch. High-energy cosmic-ray acceleration. 2010.

[4] Markus Ackermann et al. Cosmic Ray Anisotropy with 11 Years of IceCube Data. *PoS*, ICRC2023:360, 2023.

[5] Bergman, D. R., Deligny, O., Fenu, F., Fujii, T., Fujita, K., Kim, J. H., Lhenry-Yvon, I., Maris¸, I., Luce, Q., Roth, M., Salamida, F., Tsunesada, Y., Verzi, V., for the Pierre Auger, and the Telescope Array. The energy spectrum of ultra-high energy cosmic rays measured at the pierre auger observatory and the telescope array. *EPJ Web Conf.*, 283:02003, 2023.

[6] Alessandro De Angelis and Mario Pimenta. *Introduction to Particle and Astroparticle Physics: Multimessenger Astronomy and its Particle Physics Foundations.* Undergraduate Lecture Notes in Physics. Springer Nature, Heidelberg, 2018.

[7] Claus Grupen. *Astroparticle Physics.* Springer, Siegen, 2005.

[8] P. A. Zyla et al. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020.

[9] Marco Stein Muzio, Luis A. Anchordoqui, and Michael Unger. Peters cycle at the end of the cosmic ray spectrum? *Phys. Rev. D*, 109(2):023006, 2024.

[10] Antonella Castellina. Augerprime: the pierre auger observatory upgrade. *EPJ Web of Conferences*, 210:06002, 2019.

[11] Florian Gate, Richard Dallier, Lilian Martin, and Benoit Revenu. Reconstruction of the parameters of cosmic ray induced extensive air showers using radio detection and simulation. *PoS*, ICRC2015:397, 2016.

[12] Ariel Bridgeman. *Determining the Mass Composition of Ultra-high Energy Cosmic Rays Using Air Shower Universality.* PhD thesis, Karlsruhe U., 2018.

[13] M.A. Srednicki. *Quantum Field Theory.* Cambridge University Press, 2007.

[14] Felix Riehn, Ralph Engel, Anatoli Fedynitch, Thomas K. Gaisser, and Todor Stanev. Hadronic interaction model Sibyll 2.3d and extensive air showers. *Phys. Rev. D*, 102(6):063002, 2020.

[15] S. Ostapchenko. Monte carlo treatment of hadronic interactions in enhanced pomeron scheme: Qgsjet-ii model. *Phys. Rev. D*, 83:014018, Jan 2011.

[16] T. Pierog, Iu. Karpenko, J. M. Katzy, E. Yatsenko, and K. Werner. Epos lhc: Test of collective hadronization with data measured at the cern large hadron collider. *Physical Review C*, 92(3), September 2015.

[17] A Ferrari, P R Sala, Milan /CERN /INFN, A Fasso, /SLAC, J Ranft, and /Siegen U. Fluka: A multi-particle transport code.

[18] Giuseppe Battistoni et al. Overview of the FLUKA code. *Annals Nucl. Energy*, 82:10–18, 2015.

[19] M Bleicher, E Zabrodin, C Spieles, S A Bass, C Ernst, S Soff, L Bravina, M Belkacem, H Weber, H Stöcker, and W Greiner. Relativistic hadron-hadron collisions in the ultra-relativistic quantum molecular dynamics model. *Journal of Physics G: Nuclear and Particle Physics*, 25(9):1859–1896, September 1999.

[20] 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.

[21] A. B. Migdal. Bremsstrahlung and pair production in condensed media at high energies. *Phys. Rev.*, 103:1811–1820, Sep 1956.

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

[23] Johannes Albrecht, Lorenzo Cazon, Hans Dembinski, Anatoli Fedynitch, Karl-Heinz Kampert, Tanguy Pierog, Wolfgang Rhode, Dennis Soldin, Bernhard Spaan, Ralf Ulrich, and Michael Unger. The muon puzzle in cosmic-ray induced air showers and its connection to the large hadron collider. *Astrophysics and Space Science*, 367(3), March 2022.

[24] Tim Huege. Radio detection of cosmic ray air showers in the digital era. *Physics Reports*, 620:1–52, March 2016.

[25] T. Huege, W.D. Apel, T. Asch, A.F. Badea, L. Bähren, K. Bekk, A. Bercuci, M. Bertaina, P.L. Biermann, J. Blümer, H. Bozdog, I.M. Brancus, S. Buitink, M. Brüggemann, P. Buchholz, H. Butcher, A. Chiavassa, F. Cossavella, K. Daumiller, F. Di Pierro, P. Doll, R. Engel, H. Falcke, H. Gemmeke, P.L. Ghia, R. Glasstetter, C. Grupen, A. Hakenjos, A. Haungs, D. Heck, J.R. Hörandel, A. Horneffer, P.G. Isar, K.H. Kampert, Y. Kolotaev, O. Krömer, J. Kuijpers, S. Lafebre, H.J. Mathes, H.J. Mayer, C. Meurer, J. Milke, B. Mitrica, C. Morello, G. Navarra, S. Nehls, A. Nigl, R. Obenland, J. Oehlschläger, S. Ostapchenko, S. Over, M. Petcu, J. Petrovic, T. Pierog, S. Plewnia, H. Rebel, A. Risse, M. Roth, H. Schieler, O. Sima, K. Singh, M. Stümpert, G. Toma, G.C. Trinchero, H. Ulrich, J. van Buren, W. Walkowiak, A. Weindl, J. Wochele, J. Zabierowski, J.A. Zensus, and D. Zimmermann. Radio detection of cosmic ray air showers with lopes. *Nuclear Physics B - Proceedings Supplements*, 165:341–348, March 2007.

[26] P. Schellart and A et al. Nelles. Detecting cosmic rays with the lofar radio telescope. *Astronomy amp; Astrophysics*, 560:A98, December 2013.

[27] Tim Huege. Radio detection of cosmic rays with the auger engineering radio array. *EPJ Web of Conferences*, 210:05011, 2019.

[28] Frank G. Schröder. Radio detection of cosmic-ray air showers and high-energy neutrinos. *Progress in Particle and Nuclear Physics*, 93:1–68, March 2017.

[29] Frank G. Schröder. Status of the radio technique for cosmic-ray induced air showers. *Nuclear and Particle Physics Proceedings*, 279-281:190–197, 2016. Proceedings of the 9th Cosmic Ray International Seminar.

[30] Simon Chiche, Chao Zhang, Kumiko Kotera, Tim Huege, Krijn D. de Vries, Felix Schlüter, and Matias Tueros. New features in the radio-emission of very inclined air-showers. *PoS*, ARENA2022:035, 2023.

[31] T. Huege and H. Falcke. Radio emission from cosmic ray air showers: Simulation results and parametrization. *Astroparticle Physics*, 24(1–2):116–136, September 2005.

[32] Clancy W. James. Nature of radio-wave radiation from particle cascades. *Physical Review D*, 105(2), jan 2022.

[33] Tim Huege and Clancy W. James. Full monte carlo simulations of radio emission from extensive air showers with coreas, 2013.

[34] Christian Glaser, Martin Erdmann, Jörg R. Hörandel, Tim Huege, and Johannes Schulz. Simulation of radiation energy release in air showers. *Journal of Cosmology and Astroparticle Physics*, 2016(09):024–024, September 2016.

[35] Christian Glaser, Sijbrand de Jong, Martin Erdmann, and Jörg R. Hörandel. An analytic description of the radio emission of air showers based on its emission mechanisms. *Astroparticle Physics*, 104:64–77, 2019.

[36] Tim Huege, Marianne Ludwig, Olaf Scholten, and Krijn D. de Vries. The convergence of eas radio emission models and a detailed comparison of reas3 and mgmr simulations, 2010.

[37] O. Scholten, K. Werner, and F. Rusydi. A macroscopic description of coherent geo-magnetic radiation from cosmic-ray air showers. *Astroparticle Physics*, 29(2):94–103, March 2008.

[38] K WERNER and O SCHOLTEN. Macroscopic treatment of radio emission from cosmic ray air showers based on shower simulations. *Astroparticle Physics*, 29(6):393–411, July 2008.

[39] John David Jackson. *Classical Electrodynamics*. Wiley, 1998.

[40] Clancy W. James, Heino Falcke, Tim Huege, and Marianne Ludwig. General description of electromagnetic radiation processes based on instantaneous charge acceleration in “endpoints”. *Physical Review E*, 84(5), November 2011.

[41] Jaime Alvarez-Muñiz, Andrés Romero-Wolf, and Enrique Zas. Čerenkov radio pulses from electromagnetic showers in the time domain. *Phys. Rev. D*, 81:123009, Jun 2010.

[42] Atlas Obscura.

[43] Christian Glaser, Anna Nelles, Ilse Plaisier, Christoph Welling, Steven W. Barwick, Daniel García-Fernández, Geoffrey Gaswint, Robert Lahmann, and Christopher Persichilli. Nuradioreco: a reconstruction framework for radio neutrino detectors. *The European Physical Journal C*, 79(6), June 2019.

[44] Lisa Gerhardt, Spencer Klein, Thorsten Stezelberger, Steve Barwick, Kamlesh Dookayka, Jordan Hanson, and Ryan Nichol. A prototype station for arianna: A detector for cosmic neutrinos. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 624(1):85–91, 2010.

[45] S.W. Barwick, D.Z. Besson, A. Burgman, E. Chiem, A. Hallgren, J.C. Hanson, S.R. Klein, S.A. Kleinfelder, A. Nelles, C. Persichilli, S. Phillips, T. Prakash, C. Reed, S.R. Shively, J. Tatar, E. Unger, J. Walker, and G. Yodh. Radio detection of air showers with the arianna experiment on the ross ice shelf. *Astroparticle Physics*, 90:50–68, 2017.

[46] Tim Huege, Justin D. Bray, Stijn Buitink, David Butler, Richard Dallier, Ron D. Ekers, Torsten Enßlin, Heino Falcke, Andreas Haungs, Clancy W. James, Lilian Martin, Pragati Mitra, Katharine Mulrey, Anna Nelles, Benoît Revenu, Olaf Scholten, Frank G. Schröder, Steven Tingay, Tobias Winchen, and Anne Zilles. Ultimate precision in cosmic-ray radio detection — the ska. *EPJ Web of Conferences*, 135:02003, 2017.

[47] A. Corstanje, S. Buitink, J. Bhavani, M. Desmet, H. Falcke, B. M. Hare, J. R. Hörandel, T. Huege, N. Karasthatis, G. K. Krampah, P. Mitra, K. Mulrey, A. Nelles, K. Nivedita, H. Pandya, J. P. Rachen, O. Scholten, S. Thoudam, G. Trinh, and S. ter Veen. Prospects for measuring the longitudinal particle distribution of cosmic-ray air showers with ska, 2023.

[48] S. Buitink, A. Corstanje, J.E. Enriquez, H. Falcke, J.R. Hörandel, T. Huege, A. Nelles, J.P. Rachen, P. Schellart, O. Scholten, S. ter Veen, S. Thoudam, and T.N.G. Trinh. A method for high precision reconstruction of air shower xmax using two-dimensional radio intensity profiles. *Physical Review D*, 90(8), October 2014.

[49] A. Corstanje, S. Buitink, H. Falcke, B.M. Hare, J.R. Hörandel, T. Huege, G.K. Krampah, P. Mitra, K. Mulrey, A. Nelles, H. Pandya, J.P. Rachen, O. Scholten, S. ter Veen, S. Thoudam, G. Trinh, and T. Winchen. Depth of shower maximum and mass composition of cosmic rays from 50 pev to 2 eev measured with the lofar radio telescope. *Physical Review D*, 103(10), May 2021.

[50] Tanguy Pierog, Ralph Engel, Dieter Heck, and Ralf Ulrich. 3D Hybrid Air Shower Simulation in CORSIKA. In *32nd International Cosmic Ray Conference*, volume 2, page 222, 2011.

[51] K. Mulrey, S. Buitink, A. Corstanje, H. Falcke, B.M. Hare, J.R. Hörandel, T. Huege, G.K. Krampah, P. Mitra, A. Nelles, H. Pandya, J.P. Rachen, O. Scholten, S. ter Veen, S. Thoudam,

T.N.G. Trinh, and T. Winchen. On the cosmic-ray energy scale of the lofar radio telescope. *Journal of Cosmology and Astroparticle Physics*, 2020(11):017–017, November 2020.

[52] Marvin Gottowik, Christian Glaser, Tim Huege, and Julian Rautenberg. Determination of the absolute energy scale of extensive air showers via radio emission: Systematic uncertainty of underlying first-principle calculations. *Astroparticle Physics*, 103:87–93, December 2018.

[53] The Pierre Auger Collaboration, A. Abdul Halim, P. Abreu, M. Aglietta, I. Allekotte, K. Almeida Cheminant, A. Almela, R. Aloisio, J. Alvarez-Muñiz, J. Ammerman Yebra, G. A. Anastasi, L. Anchordoqui, B. Andrada, S. Andringa, Anukriti, L. Apollonio, C. Aramo, P. R. Araújo Ferreira, E. Arnone, J. C. Arteaga Velázquez, P. Assis, G. Avila, E. Avocone, A. Bakalova, F. Barbato, A. Bartz Mocellin, J. A. Bellido, C. Berat, M. E. Bertaina, G. Bhatta, M. Bianciotto, P. L. Biermann, V. Binet, K. Bismark, T. Bister, J. Biteau, J. Blazek, C. Bleve, J. Blümer, M. Boháčová, D. Boncioli, C. Bonifazi, L. Bonneau Arbeletche, N. Borodai, J. Brack, P. G. Brichetto Orchera, F. L. Briechle, A. Bueno, S. Buitink, M. Buscemi, M. Büsken, A. Bwembya, K. S. Caballero-Mora, S. Cabana-Freire, L. Caccianiga, R. Caruso, A. Castellina, F. Catalani, G. Cataldi, L. Cazon, M. Cerda, A. Cermenati, J. A. Chinellato, J. Chudoba, L. Chytka, R. W. Clay, A. C. Cobos Cerutti, R. Colalillo, A. Coleman, M. R. Coluccia, R. Conceição, A. Condorelli, G. Consolati, M. Conte, F. Convenga, D. Correia dos Santos, P. J. Costa, C. E. Covault, M. Cristinziani, C. S. Cruz Sanchez, S. Dasso, K. Daumiller, B. R. Dawson, R. M. de Almeida, J. de Jesús, S. J. de Jong, J. R. T. de Mello Neto, I. De Mitri, J. de Oliveira, D. de Oliveira Franco, F. de Palma, V. de Souza, B. P. de Souza de Errico, E. De Vito, A. Del Popolo, O. Deligny, N. Denner, L. Deval, A. di Matteo, M. Dobre, C. Dobrigkeit, J. C. D'Olivo, L. M. Domingues Mendes, Q. Dorosti, J. C. dos Anjos, R. C. dos Anjos, J. Ebr, F. Ellwanger, M. Emam, R. Engel, I. Epicoco, M. Erdmann, A. Etchegoyen, C. Evoli, H. Falcke, J. Farmer, G. Farrar, A. C. Fauth, N. Fazzini, F. Feldbusch, F. Fenu, A. Fernandes, B. Fick, J. M. Figueira, A. Filipčič, T. Fitoussi, B. Flaggs, T. Fodran, T. Fujii, A. Fuster, C. Galea, C. Galelli, B. García, C. Gaudu, H. Gemmeke, F. Gesualdi, A. Gherghel-Lascu, P. L. Ghia, U. Giaccari, J. Glombitza, F. Gobbi, F. Gollan, G. Golup, M. Gómez Berisso, P. F. Gómez Vitale, J. P. Gongora, J. M. González, N. González, I. Goos, D. Góra, A. Gorgi, M. Gottowik, T. D. Grubb, F. Guarino, G. P. Guedes, E. Guido, L. Gülzow, S. Hahn, P. Hamal, M. R. Hampel, P. Hansen, D. Harari, V. M. Harvey, A. Haungs, T. Hebbeker, C. Hojvat, J. R. Hörandel, P. Horvath, M. Hrabovský, T. Huege, A. Insolia, P. G. Isar, P. Janecek, V. Jilek, J. A. Johnsen, J. Jurysek, K. H. Kampert, B. Keilhauer, A. Khakurdikar, V. V. Kizakke Covilakam, H. O. Klages, M. Kleifges, F. Knapp, J. Köhler, N. Kunka, B. L. Lago, N. Langner, M. A. Leigui de Oliveira, Y. Lema-Capeans, A. Letessier-Selvon, I. Lhenry-Yvon, L. Lopes, L. Lu, Q. Luce, J. P. Lundquist, A. Machado Payeras, M. Majercakova, D. Mandat, B. C. Manning, P. Mantsch, S. Marafico, F. M. Mariani, A. G. Mariazzi, I. C. Mariş, G. Marsella, D. Martello, S. Martinelli, O. Martínez Bravo, M. A. Martins, H. J. Mathes, J. Matthews, G. Matthiae, E. Mayotte, S. Mayotte, P. O. Mazur, G. Medina-Tanco, J. Meinert, D. Melo, A. Menshikov, C. Merx, S. Michal, M. I. Micheletti, L. Miramonti, S. Mollerach, F. Montanet, L. Morejon, C. Morello, K. Mulrey, R. Mussa, W. M. Namasaka, S. Negi, L. Nellen, K. Nguyen, G. Nicora, M. Niechciol, D. Nitz, D. Nosek, V. Novotny, L. Nožka, A. Nucita, L. A. Núñez, C. Oliveira, M. Palatka, J. Pallotta,

S. Panja, G. Parente, T. Paulsen, J. Pawlowsky, M. Pech, J. Pękala, R. Pelayo, L. A. S. Pereira, E. E. Pereira Martins, J. Perez Armand, C. Pérez Bertolli, L. Perrone, S. Petrera, C. Petrucci, T. Pierog, M. Pimenta, M. Platino, B. Pont, M. Pothast, M. Pourmohammad Shahvar, P. Privitera, M. Prouza, A. Puyleart, S. Querchfeld, J. Rautenberg, D. Ravignani, J. V. Reginatto Akim, M. Reininghaus, J. Ridky, F. Riehn, M. Risse, V. Rizi, W. Rodrigues de Carvalho, E. Rodriguez, J. Rodriguez Rojo, M. J. Roncoroni, S. Rossoni, M. Roth, E. Roulet, A. C. Rovero, P. Ruehl, A. Saftoiu, M. Saharan, F. Salamida, H. Salazar, G. Salina, J. D. Sanabria Gomez, F. Sánchez, E. M. Santos, E. Santos, F. Sarazin, R. Sarmento, R. Sato, P. Savina, C. M. Schäfer, V. Scherini, H. Schieler, M. Schimassek, M. Schimp, D. Schmidt, O. Scholten, H. Schoorlemmer, P. Schovánek, F. G. Schröder, J. Schulte, T. Schulz, S. J. Sciutto, M. Scornavacche, A. Segreto, S. Sehgal, S. U. Shivashankara, G. Sigl, G. Silli, O. Sima, K. Simkova, F. Simon, R. Smau, R. Šmída, P. Sommers, J. F. Soriano, R. Squartini, M. Stadelmaier, S. Stanič, J. Stasielak, P. Stassi, S. Strähnz, M. Straub, T. Suomijärvi, A. D. Supanitsky, Z. Svozilikova, Z. Szadkowski, F. Tairli, A. Tapia, C. Taricco, C. Timmermans, O. Tkachenko, P. Tobiska, C. J. Todero Peixoto, B. Tomé, Z. Torrès, A. Travaini, P. Travnicek, C. Trimarelli, M. Tueros, M. Unger, L. Vaclavek, M. Vacula, J. F. Valdés Galicia, L. Valore, E. Varela, A. Vásquez-Ramírez, D. Veberič, C. Ventura, I. D. Vergara Quispe, V. Verzi, J. Vicha, J. Vink, S. Vorobiov, C. Watanabe, A. A. Watson, A. Weindl, L. Wiencke, H. Wilczyński, D. Wittkowski, B. Wundheiler, B. Yue, A. Yushkov, O. Zapparrata, E. Zas, D. Zavrtanik, and M. Zavrtanik. Demonstrating agreement between radio and fluorescence measurements of the depth of maximum of extensive air showers at the pierre auger observatory, 2023.

[54] D. Heck, J. Knapp, J. N. Capdevielle, G. Schatz, and T. Thouw. CORSIKA: A Monte Carlo code to simulate extensive air showers. 2 1998.

[55] E. Zas, F. Halzen, and T. Stanev. Electromagnetic pulses from high-energy showers: Implications for neutrino detection. *Phys. Rev. D*, 45:362–376, Jan 1992.

[56] P. Allison, J. Auffenberg, R. Bard, J.J. Beatty, D.Z. Besson, S. Böser, C. Chen, P. Chen, A. Connolly, J. Davies, M. DuVernois, B. Fox, P.W. Gorham, E.W. Grashorn, K. Hanson, J. Haugen, K. Helbing, B. Hill, K.D. Hoffman, E. Hong, M. Huang, M.H.A. Huang, A. Ishihara, A. Karle, D. Kennedy, H. Landsman, T.C. Liu, L. Macchiarulo, K. Mase, T. Meures, R. Meyhandan, C. Miki, R. Morse, M. Newcomb, R.J. Nichol, K. Ratzlaff, M. Richman, L. Ritter, C. Rott, B. Rotter, P. Sandstrom, D. Seckel, J. Touart, G.S. Varner, M.-Z. Wang, C. Weaver, A. Wendorff, S. Yoshida, and R. Young. Design and initial performance of the askaryan radio array prototype eev neutrino detector at the south pole. *Astroparticle Physics*, 35(7):457–477, 2012.

[57] J.A. Aguilar, P. Allison, J.J. Beatty, H. Bernhoff, D. Besson, N. Bingefors, O. Botner, S. Buitink, K. Carter, B.A. Clark, A. Connolly, P. Dasgupta, S. de Kockere, K.D. de Vries, C. Deaconu, M.A. DuVernois, N. Feigl, D. García-Fernández, C. Glaser, A. Hallgren, S. Hallmann, J.C. Hanson, B. Hendricks, B. Hokanson-Fasig, C. Hornhuber, K. Hughes, A. Karle, J.L. Kelley, S.R. Klein, R. Krebs, R. Lahmann, M. Magnuson, T. Meures, Z.S. Meyers, A. Nelles, A. Novikov, E. Oberla, B. Oeyen, H. Pandya, I. Plaisier, L. Pyras,

D. Ryckbosch, O. Scholten, D. Seckel, D. Smith, D. Southall, J. Torres, S. Toscano, D.J. Van Den Broeck, N. van Eijndhoven, A.G. Vieregg, C. Welling, S. Wissel, R. Young, and A. Zink. Design and sensitivity of the radio neutrino observatory in greenland (rno-g). *Journal of Instrumentation*, 16(03):P03025, March 2021.

[58] Jaime Álvarez Muñiz, Rafael Alves Batista, Aswathi Balagopal V., Julien Bolmont, Mauricio Bustamante, Washington Carvalho, Didier Charrier, Ismaël Cognard, Valentin Decoene, Peter B. Denton, Sijbrand De Jong, Krijn D. De Vries, Ralph Engel, Ke Fang, Chad Finley, Stefano Gabici, QuanBu Gou, JunHua Gu, Claire Guépin, HongBo Hu, Yan Huang, Kumiko Kotera, Sandra Le Coz, Jean-Philippe Lenain, GuoLiang Lü, Olivier Martineau-Huynh, Miguel Mostafá, Fabrice Mottez, Kohta Murase, Valentin Niess, Foteini Oikonomou, Tanguy Pierog, XiangLi Qian, Bo Qin, Duan Ran, Nicolas Renault-Tinacci, Markus Roth, Frank G. Schröder, Fabian Schüssler, Cyril Tasse, Charles Timmermans, Matías Tueros, XiangPing Wu, Philippe Zarka, Andreas Zech, B. Theodore Zhang, JianLi Zhang, Yi Zhang, Qian Zheng, and Anne Zilles. The giant radio array for neutrino detection (grand): Science and design. *Science China Physics, Mechanics amp; Astronomy*, 63(1), August 2019.

[59] T Antoni, W.D Apel, F Badea, K Bekk, A Bercuci, H Blümer, H Bozdog, I.M Brancus, C Büttner, A Chilingarian, K Daumiller, P Doll, J Engler, F Feßler, H.J Gils, R Glasstetter, R Haeusler, A Haungs, D Heck, J.R Hörandel, A Iwan, K.-H Kampert, H.O Klages, G Maier, H.J Mathes, H.J Mayer, J Milke, M Müller, R Obenland, J Oehlschläger, S Ostapchenko, M Petcu, H Rebel, M Risse, M Roth, G Schatz, H Schieler, J Scholz, T Thouw, H Ulrich, A Vardanyan, J Weber, A Weindl, J Wentz, J Wochele, J Zabierowski, and S Zagromski. The cosmic-ray experiment kascade. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 513(3):490–510, 2003.

[60] Maximilian Reininghaus and Ralf Ulrich. Corsika 8 – towards a modern framework for the simulation of extensive air showers. *EPJ Web of Conferences*, 210:02011, 2019.

[61] Maximilian Reininghaus. *The air shower simulation framework CORSIKA 8: Development and first applications to muon production.* PhD thesis, Karlsruher Institut für Technologie (KIT), 2022. 51.13.03; LK 01.

[62] Apache Software Foundation. Parquet.

[63] YAML. Yaml.

[64] T. Huege, M. Ludwig, and C. W. James. Simulating radio emission from air showers with coreas. In *AIP Conference Proceedings*. AIP, 2013.

[65] Juan Ammerman-Yebra, Jean-Marco Alameddine, Johannes Albrecht, Jaime Alvarez-Muniz, Luisa Arrabito, Jannik Augscheller, Antonio Augusto Alves Jr., Dominik Baack, Konrad Bernlöhr, Marcus Bleicher, Alan Coleman, Hans Dembinski, Dominik Elsässer, Ralph Engel, Alfredo Ferrari, Chloé Gaudu, Christian Glaser, Dieter Heck, Fan Hu, Tim Huege, Karl-Heinz Kampert, Nikolaos Karastathis, Uzair Abdul Latif, Hualin Mei, Lukas Nellen, Tanguy Pierog,

Remy Prechelt, Maximilian Reininghaus, Wolfgang Rhode, Felix Riehn, Maximilian Sackel, Paola Sala, Pranav Sampathkumar, Alexander Sandrock, Jan Soedingrekso, Ralf Ulrich, Donglian Xu, Enrique Zas, and Simon De Kockere. Simulations of cross media showers with corsika 8. In *Proceedings of 38th International Cosmic Ray Conference — PoS(ICRC2023)*, ICRC2023. Sissa Medialab, August 2023.

[66] The pandas development team. pandas-dev/pandas: Pandas, February 2020.

[67] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[68] Nikolaos Karastathis, Remy Prechelt, Tim Huege, and Juan Ammerman-Yebra. Simulations of radio emission from air showers with CORSIKA 8. *PoS*, ICRC2021:427, 2021.

[69] Nikolaos Karastathis, Remy Prechelt, Juan Ammerman-Yebra, and Tim Huege. Simulating radio emission from air showers with CORSIKA 8. *PoS*, ARENA2022:050, 2023.

[70] Jean-Marco Alameddine, Lars Bollmann, Mario Dunsch, Tomasz Fuchs, Pascal Gutjahr, Jan-Hendrik Köhne, Claudio Kopper, Kai Krings, Chung-Yun Kuo, Thorben Menne, Maximilian Nöthe, Alex Olivas, Wolfgang Rhode, Maximilian Sackel, Alexander Sandrock, Austin Schneider, Jan Soedingrekso, and Jacob van Santen. tudo-astroparticlephysics/proposal: v7.6.1, 2023.

[71] Gene M. Amdahl. Validity of the single processor approach to achieving large scale computing capabilities, reprinted from the afips conference proceedings, vol. 30 (atlantic city, n.j., apr. 18–20), afips press, reston, va., 1967, pp. 483–485, when dr. amdahl was at international business machines corporation, sunnyvale, california. *IEEE Solid-State Circuits Society Newsletter*, 12(3):19–20, 2007.

[72] Antonio Augusto Alves Jr. Gyges.

[73] 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(1):012021, 2020.

[74] Tim Huege and Maximilian Reininghaus. The particle-shower simulation code corsika 8, 2023.

[75] John Hall Gladstone and T. P. Dale. Xiv. researches on the refraction, dispersion, and sensitiveness of liquids. *Phil. Trans. Roy. Soc. Lond.*, 153:317–343, 1863.

[76] Jean-Marco Alameddine. *Enabling Next-generation Particle Cascade Simulations Implementing the Propagation Software -PROPOSAL as an Interaction Model for CORSIKA 8.* PhD thesis, TU Dortmund, 2024.

[77] Anne Zilles. *Emission of Radio Waves in Particle Showers: Validation of microscopic simulations with the SLAC T-510 experiment and their potential in the future Square Kilometer Array.* PhD thesis, Karlsruhe Institute of Technology, 2016.

[78] A. et al. Aab. Energy estimation of cosmic rays with the engineering radio array of the pierre auger observatory. *Physical Review D*, 93(12), June 2016.

[79] Benedikt Zimmermann. Cosmic ray physics with the auger engineering radio array (aera). *Nuclear and Particle Physics Proceedings*, 291-293:90–95, 2017. "New eyes on the Universe" CRIS 2016 Cosmic Rays International Seminars Proceedings of the Cosmic Rays International Seminars.

[80] Anna Nelles, Stijn Buitink, Heino Falcke, Jörg R. Hörandel, Tim Huege, and Pim Schellart. A parameterization for the radio emission of air showers as predicted by coreas simulations and applied to lofar measurements. *Astroparticle Physics*, 60:13–24, 2015.

[81] Felix Schlüter, Marvin Gottowik, Tim Huege, and Julian Rautenberg. Refractive displacement of the radio-emission footprint of inclined air showers simulated with coreas. *The European Physical Journal C*, 80(7), July 2020.

[82] Stefan Jansen. *Radio for the masses: Cosmic ray mass composition measurements in the radio frequency domain.* PhD thesis, Nijmegen U., 2016.

[83] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.

[84] A. Corstanje, S. Buitink, M. Desmet, H. Falcke, B.M. Hare, J.R. Hörandel, T. Huege, V.B. Jhansi, N. Karastathis, G.K. Krampah, P. Mitra, K. Mulrey, A. Nelles, K. Nivedita, H. Pandya, O. Scholten, K. Terveer, S. Thoudam, G. Trinh, and S. ter Veen. A high-precision interpolation method for pulsed radio signals from cosmic-ray air showers. *Journal of Instrumentation*, 18(09):P09005, September 2023.

[85] Nikolaos Karastathis et al. Using pulse-shape information for reconstructing cosmic-ray air showers and validating antenna responses with LOFAR and SKA. *PoS*, ICRC2023:487, 2023.

[86] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. Pearson correlation coefficient. In *Noise reduction in speech processing*, pages 37–40. Springer, 2009.