Measurements in the Forward Phase-Space with the CMS Experiment and their Impact on Physics of Extensive Air Showers

Zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

an der Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)

genehmigte

Dissertation

von

Colin Baus
aus Mannheim

Tag der mündlichen Prüfung: 17.04.2015
Referent: Prof. Dr. J. Blümer
Korreferent: Prof. Dr. G. Quast
Betreuer: Dr. R. Ulrich

ABSTRACT

The astrophysical interpretation of ultra-high energy cosmic rays is based on detection of extensive air showers in indirect measurements. Hadronic interaction models that are needed for such analyses require parameters to be adjusted to collider data since soft particle production cannot be calculated from first principles. Within this work, the program CRMC was developed that unifies all air shower hadronic interaction models and supports the output formats used by collider experiments. Almost all LHC experiments have adopted the use these hadronic interaction models thanks to CRMC. The program can even be used in detector simulations to make direct comparison to reconstructed quantities from which the cosmic ray and the particle physics communities benefit immensely. Furthermore, nuclear effects were studied with the CMS experiments at the LHC. The production cross section was derived in recent proton-lead collision data at $\sqrt{s_{NN}} = 5.02\,\text{TeV}$ in order to study nuclear effects. The measurement constrains corrections on the cross section calculation known from theory. Diffractive collisions that influence the energy transport in air showers were studied with the CASTOR calorimeter. Only half of the tested models show a good agreement to the data. The highlight of this work is a global likelihood analysis that not only directly connects measurements in the forward phase-space to the physics of air showers but achieves to constrain parameters of hadronic interaction models from astrophysical priors. The analysis method has the potential to revolutionise how model parameters are tuned in the future and it is planned to extend this further.

ZUSAMMENFASSUNG

Die astrophysikalische Interpretation ultrahochenergetischer kosmischer Strahlung basiert auf der Detektion von ausgedehnten Luftschauern mit indirekten Messmethoden. Die hadronischen Wechselwirkungsmodelle müssen für die Analyse von Luftschauermessungen mit Daten von Beschleunigerexperimenten angepasst werden, da keine Theorie existiert, die Teilchenendzustände bei kleinen Impulsüberträgen beschreiben kann. In dieser Arbeit wurde das Programm CRMC entwickelt, das alle Luftschauermodelle vereint und die von Beschleunigerexperimenten benötigten Standardausgabeformate unterstüzt. Dank diesem Programm werden die Wechselwirkungsmodelle an fast allen Experimenten am LHC direkt in der Analyse eingesetzt und können sogar direkt mit Detektordaten verglichen werden. Die Welt der Teilchenphysik profitiert in gleicher Maßen, wie die Welt der Luftschauerphysik. Darüber hinaus wurden in dieser Arbeit Messungen am CMS-Experiments am LHC-Beschleuniger durchgeführt, die Kerneffekte in Teilchenkollisionen untersuchen. In erst kürzlich aufgezeichneten Daten von Proton-Blei-Kollisionen, bei noch nie zuvor

erreichten Schwerpunktsenergien pro Nukleonpaar von 5.02 TeV, wurde der Wechselwirkungsquerschnitt für Teilchenproduktion mit Vorwärtskalorimetern gemessen. Das Ergebnis schränkt die Größe von Korrekturen auf theoretische Berechnungen des Wirkungsquerschnitts ein und hat dadurch direkte Bewandtnis für die Luftschauerphysik. Ebenso wichtig sind diffraktive Wechselwirkungen, die den Energietransport in Luftschauern beeinflussen. Eine Messung mit dem CASTOR-Kalorimeter konnte zeigen, dass nicht alle Wechselwirkungsmodelle die diffraktiven Prozesse in Proton-Kern-Kollisionen beschreiben können. Das Glanzstück der Arbeit ist eine globale Likelihood-Analyse, die nicht nur eine direkte Verbindung zwischen Luftschauern und Messungen im Vorwärtsbereich herstellt, sondern zum ersten Mal ermöglicht, auf systematische Weise Modellparameter der hadronischen Wechselwirkungsmodelle anhand astrophysikalischer Messdaten einzuschränken. Diese Analyse hat das Potenzial die Anpassungen von Modellparametern grundlegend zu verändern und soll in der Zukunft dahingehend ausgebaut werden.

CONTENTS

1 INTRODUCTION

More than 100 years have passed since the discovery of cosmic rays, yet, the most important questions about Ultra-High-Energy Cosmic Rays (UHECR) are still open [BEH09]. What type of particles are these exactly, how do they interact within our atmosphere, how are they accelerated, and where are the extreme environments that achieve this? UHECR are studied in huge experiments like the Pierre Auger Observatory [Aab+15b] and the Telescope Array [Tok+12; Abu+12]. These experiments gather large statistics of high-quality data recording Extensive Air Showers (EAS), which are cascades of secondary particles produced in hadronic interactions of cosmic rays with the atmosphere.

The interpretation of measurements based on EAS is limited by large systematic uncertainties rooted in our limited understanding of the modelling of hadronic interactions. The theory of strong interaction (QCD) is well established but still no predictions can be made for soft particle production with low momentum transfer. Models based on phenomenologies derived from data are necessary to describe such processes. Nonetheless, even straightforward quantities of the EAS cannot be described well. Most noteworthy in this context is that measurements of the muon production in EAS are not in agreement with any of the predictions of current interaction models [Ape+11b; Aab+15a; Aab+14b], suggesting the presence of unknown processes in these particle cascades. UHECR also offer a unique environment to search for phenomena of particle physics beyond the reach of particle accelerators. Theories describing new physics that can be probed include the violation of Lorentz-invariance [CG99; Lib13] and restoration of chiral symmetry [FA13].

One of the results derived from EAS is the proton-proton cross section measured in cosmic ray collisions at energies far above collider energies [Abr+12]. To extract information about proton-proton interactions from proton-air collisions, nuclear effects are subtracted using the Glauber theory. Nuclear shadowing and antishadowing effects are known from theory [Fra+93; Alv+10; Cio+11; Alv+13] but are not included in the Glauber framework. It is one of the goals of this thesis to asses the nuclear corrections with LHC data. This is also of paramount importance for the better understanding of uncertainties on the determination of the UHECR mass composition [Aab+14a; Abb+14b], where the inelastic cross section, nuclear effects, and diffraction in proton-nucleus collisions are significant sources of uncertainties.

Only the low-energy extensions of the UHECR experiments (HEAT and TALE) observe cosmic ray interactions at the same centre-of-mass energies as foreseen in the upcoming run phase at the most powerful man-made accelerator, the LHC at CERN [EB08]. At LHC, both proton and lead ion collisions give insights into hadronic physics. Within the work of this thesis the two worlds of air showers and particle physics are united by providing

a software interface for the air shower models to communicate with collider experiments. With the CRMC program, detector simulations at the LHC can be explicitly performed with the phenomenological models that are used for air shower simulations, and, thus, direct comparisons to reconstructed quantities are possible. Furthermore, this class of models describe well the irreducible QCD background to all measurements at the LHC [dEn+11]. Thus, both fields, UHCER as well as LHC, benefit from this contribution. CRMC is already used not only by the ATLAS, CMS, LHCb, IceCube, and the Pierre Auger Collaborations but also by theorists.

The hadronic particle production that is driving the development of EAS is dominant at very small angles to the beam, in the so called forward phase-space, since the majority of primary energy is directed there. At the same time, this makes it difficult to study them at collider experiments. The CMS experiment [Cha+08] at the LHC has very good calorimetric coverage of the forward phase-space. Particularly noteworthy are the Hadronic Forward (HF) calorimeters and the CASTOR calorimeter that all are located close to the beam pipe. The high-radiation and high-magnetic field environment provides challenges that need to be accounted for in the detector design. One important effect is that small movements of the CASTOR detector due to the magnetic field can lead to large corrections on the measured energy deposit. To reduce systematic uncertainties on the energy measurement, necessary improvements are made as part of this thesis in the determination of the actual position. The approach is based on a combination of information from multiple distance sensor techniques. This is presented together with a new description of the detector geometry in the GEANT4 simulation [BCG94] of CMS, which allows to account for the observed movements in event reconstruction and simulation.

In 2013 proton-nucleus collisions took place at the LHC for the first time. A novel measurement of the production cross section is performed within this thesis that explicitly determines different contributions from photo-hadronic and diffractive processes. This measurement is based on information from the CMS forward calorimeters. With this unique result, the size of the corrections to the Glauber model is constrained. Moreover, data from the CASTOR detector are used to study very large rapidity gaps in nuclear collisions at $\sqrt{s_{NN}} = 5\,\text{TeV}$, which has not been possible before.

Finally, a global analysis technique is developed that combines measurements from UHECR and collider experiments. Constraints on parameters of hadronic interaction models are derived by comparing them to simulations obtained with modified event generators. It is remarkable and very innovative to obtain constraining power from an astrophysical prior. This is achieved by studying the UHECR mass composition. For the first time, the UHECR and collider data are systematically investigated over the boundaries of many different centre-of-mass energies and collision systems including nuclear interactions. The approach has the potential to simultaneously investigate the comprehensive and high-dimensional phase-space of astroparticle physics as well as the high-energy QCD multi-particle production. It is foreseen to further develop and extent this analysis tool into a major large-scale framework for scientific research in this field.

2

THEORETICAL BACKGROUND

> Thirty years can seem like an eternity to anyone who actually lives through thirty years of adulthood. But for the development of physics, thirty years is only a moment.
>
> K. Wilson

2.1 FROM UNDERSTANDING EXTENSIVE AIR SHOWERS TO PROBING THE UNIVERSE

Cosmic rays are charged particles accelerated at Galactic and extragalactic sources with energies reaching more than hundreds of EeV [BEH09]. On Earth, we detect these particles after they have traversed vast distances in Space and we are struggling for more than 100 years to find out where and how they are produced. This information is contained only indirectly since it is convolved with the environmental conditions through which they pass as they travel through and interact with the universe. Therefore, cosmic rays are a highly interesting research topic since the beginning of the 19th century. Despite the investment in many experiments it is still extremely difficult to study the properties of cosmic rays, in particular at the highest energies. The main reason lies with the steeply falling particle flux with increasing energy. The flux has been measured by many experiments with high statistical precision. Measurements of the flux (see Fig. 1 and a zoom of the highest energies in Fig. 2) show a falling power law spectrum $\propto E^{-\alpha}$ with several regions rich in so far unexplained features.

Multiple breaking points of the power law generate a structure often compared to the anatomy of the human leg: a knee-like structure where α increases and an ankle-like structure where it decreases. At the very end of the measurable spectrum is a strong suppression of the flux, which is typically described by the energy $E_{1/2}$ defined as the point where the flux dropped to half of its expected value from an extrapolated flux with a spectral index given by the value just above the ankle. The Pierre Auger Observatory measured $\log_{10}(E_{1/2}/\text{eV}) \simeq 19.6$ [Aab+13].

The breaking point at $E_{\text{knee}} \simeq 4\times10^{15}$ eV has been identified by direct and indirect measurements as a feature of the proton particle flux [Agl+04; Ape+11a; Ame+08; Bud+13]. Heavy particles with large charge, Z, gain more energy assuming the same acceleration mechanism. Therefore, the slope of heavy particles is expected not to change until higher

energies of $Z\,E_{\mathrm{knee}}$. At E_{knee}, α changes from $\simeq$2.7 to $\simeq$3.2. Data from many experiments [Ape+11a; Abb+13; Bud+13] show more substructure in the spectrum in the energy region between the knee and the ankle. However, so far no precise cosmic ray particle identification can be done at these energies. At the ankle, $E_{\mathrm{ankle}} = 10^{18.7}$ eV, the spectrum softens to a power law with a slope of $\alpha \simeq 2.6$ [Aab+13].

Most theories predict a transition at E_{th}, at which the flux of particles produced within our Galaxy, typically accelerated by supernova remnants, becomes smaller than an extragalactic flux of yet undiscovered sources [ABG12]. The "dip" model provides one explanation for the ankle, in which a heavy Galactic component dies out at $E_{\mathrm{th}} \simeq 5\times10^{17}$ eV. Extragalactic protons experience pair production losses ($p + \gamma_{\mathrm{CMB}} \rightarrow p + e^+ + e^-$) starting at around 10^{18} eV due to interaction with photons of the Cosmic Microwave Background (CMB). The extragalactic flux reaches its maximum at around 10^{18} eV. A second knee appears in the transition region a bit lower in energy than the ankle, at which the spectral index increases. An alternative theory is the "ankle" model, where the transition to extragalactic cosmic rays takes place at higher energies of around $E_{\mathrm{th}} \simeq 10^{18.1}$ eV. It requires that the extragalactic proton component has an index $\alpha \simeq 2$ to be able to flatten the spectrum at ankle energies. Also "mixed-composition" models exists where multiple species of nuclei contribute to the composition. Due to freedom of the mixture of the species they can reproduce experimental data well. In all cases a build-up just before the cut-off is predicted.

While current experimental evidence cannot rule out any of these models, a broad agreement exists that a shift from a Galactic to an extragalactic components must take place since the confinement of the Galactic magnetic field is not sufficient and high energy particles leave the plane.

Multiple interpretations exist also for the highest energies. In the event that the particles at the highest energies are mostly protons, cosmic rays need to have their origin in sources closer than 100 Mpc for energies above 6×10^{19} eV [KO11] due to the Greisen and Zatsepin-Kuzmin (GZK) effect [ZK66; Gre66]. For higher energies, protons interact with the cosmic microwave background radiation and lose energy by radiating off pions after producing a Δ-resonance

$$p + \gamma_{\mathrm{CMB}} \rightarrow \Delta^+ \rightarrow p + \pi^0$$
$$p + \gamma_{\mathrm{CMB}} \rightarrow \Delta^+ \rightarrow n + \pi^+ .$$

The charged pions will eventually produce neutrinos as decay products and photons from the neutral pion are also expected to arrive at Earth. Both have not yet been discovered, but several experiments could become sufficiently sensitive within the next years.

If cosmic rays are instead dominated by heavier nuclei, a quite different effect exists that allows for a similar cut-off in the energy spectrum and is hard to differentiate from the excitation of the Δ-resonance. In this case, the cut-off is due to photo disintegration, where the nuclei lose mass via the giant dipole resonance. This leads to an energy reduction over distance travelled by the nuclei, indistinguishable from photo-pion production. For example, the energy of iron nuclei with an initial momentum of 6×10^{19} eV is reduced by 50 % after

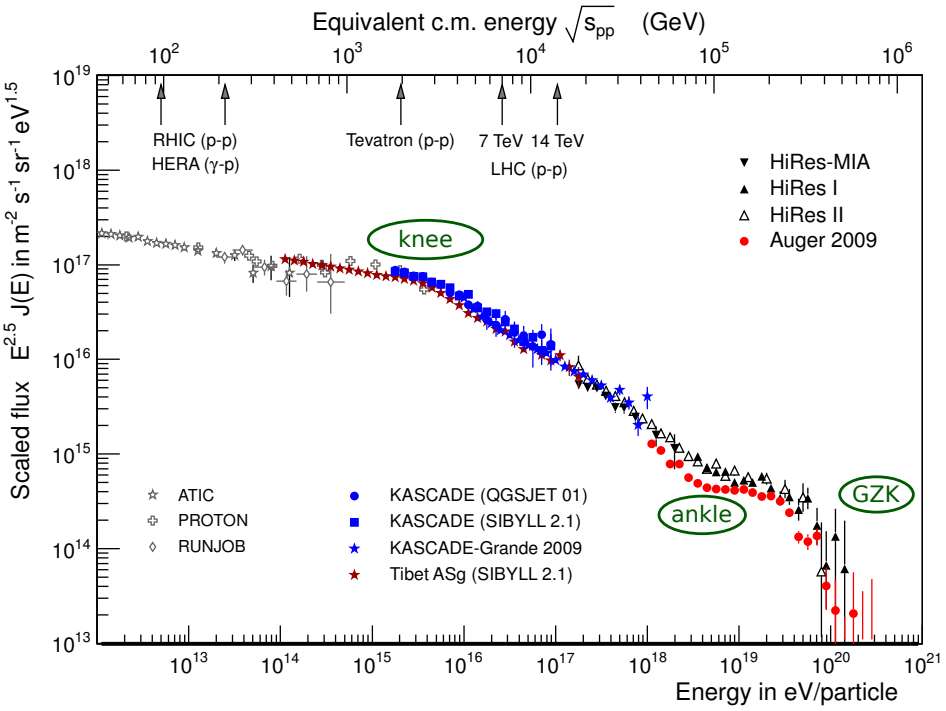


Figure 1: The cosmic ray flux over a wide range of particle energies (given in laboratory frame for the bottom axis and centre-of-mass frame for the top axis). Measurements from various experiments are included. The energies of recent collider experiments are marked at the top axis. Note that the flux is scaled by a factor $E^{2.5}$. (from [Bau11] and references therein)

travelling 100 Mpc; leading to a similar effect as seen from photo-pion emission [KO11].

The lack of clear-cut explanations of the measured phenomena can at least be partly blamed on difficulties in direct detection of cosmic rays with energies above 10^{16} eV. Direct detection requires a high flux or a large detection area; neither of these are technically feasible on high-altitude balloons or satellites outside the influence of the shielding atmosphere. Indirect measurements, on the other hand, rely on the observation of extensive air showers, which are cascades of secondary particles produced when primary cosmic ray particles interact with air molecules. To maximise the information gain from the detection of air showers, experiments pursue a hybrid technique that includes direct sampling of the secondaries at ground level, measurement of the fluorescence light emitted during propagation in the atmosphere, and emission of radiation in the radio or microwave frequency ranges.

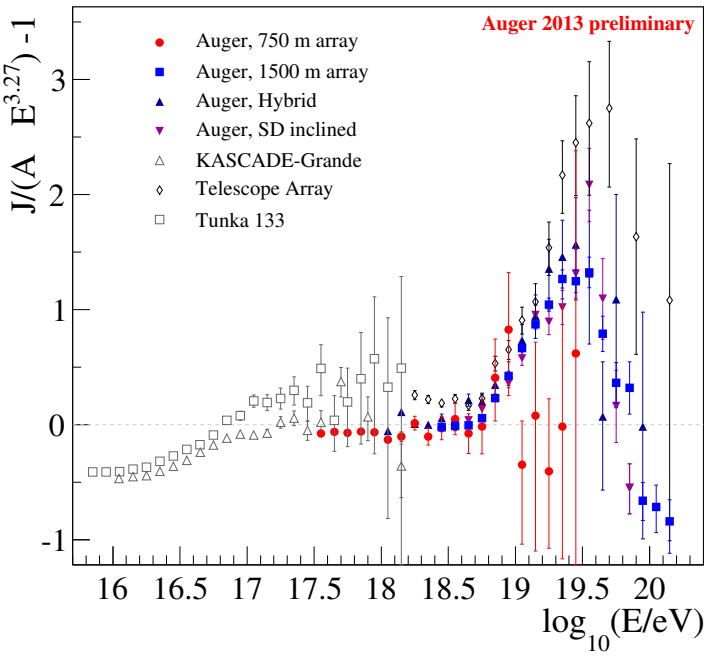


Figure 2: Cosmic ray flux as a function of energy in laboratory frame measured by different experiments. (from [Aab+13] and references therein)

2.1.1 ACCELERATION AND ORIGIN OF ULTRA-HIGH ENERGY COSMIC RAYS

Air showers induced by Ultra-High Energy Cosmic Rays (UHECR) have the potential to answer many open questions: whether UHECR are of Galactic or extragalactic origin, what are the sources and the acceleration mechanisms, and what is the abundance of masses of the particles (composition). In this work we will focus on improving the understanding of particle collisions within the air shower cascades to enable the possibility of answering these questions. The knowledge of collision processes comes from accelerator experiments that are executed under well known and reproducible conditions in the laboratory. This microscopic understanding is transferred to simulations of cosmic rays colliding in our atmosphere and producing air showers. With more accurate simulations, the systematic uncertainties of the astrophysical interpretation of cosmic ray data decrease. In the past, there have been many collider experiments (`LEP collider`: Aleph, Delphi, Opal, and L3; `RHIC collider`: STAR, PHENIX, PHOBOS, and BRAHMS; `TEVATRON collider`: CDF and DØ) at energies below $\sqrt{s_{NN}} = 2\,\text{TeV}$ that recorded collisions of electrons, positrons, protons, or ions. The quantity $\sqrt{s_{NN}}$ refers here to the centre-of-mass energy of the nucleon-nucleon system of the collision. At the Large Hadron Collider (LHC), collisions of hadrons can be probed with energies up to $\sqrt{s_{NN}} = 14\,\text{TeV}$. As of 2014, the design energy has not yet been reached but collision data have been recorded with a maximum energy of $\sqrt{s_{NN}} = 8\,\text{TeV}$. These remarkable data offer valuable insights to the cosmic ray community.

Sources of comic rays can accelerate particles to significantly higher energies. For protons, the energy of the ankle region corresponds to $\sqrt{s_{NN}} = 10^{14}$ eV or 100 TeV in the centre-of-mass frame of the individual nucleons that collide within the atmosphere. This energy surpasses the design energy of the largest man-made accelerator by more than a factor of 7. The first collisions of the cascade started by an UHECR particle take place at these extremely high centre-of-mass energies and cannot be simulated from experimental data without extrapolations. This leads to a fundamental problem since only phenomenological models are available to calculate this [Kna97]. The collision energy of later generations is distributed to many secondary particles and falls back into the regime, where experimental data is available. Measurements at these energies can be used to adjust the models used in simulations, but due to phase-space constraints of experimental detectors, the models are never free of extrapolations.

Man-made particle collider designs are based on linear and circular accelerators. The former have the advantage of low synchrotron radiation losses. In circular colliders, a magnetic field of strength B is applied to keep the particles on their track. The total power radiated by a particle with charge q is

$$P_{\text{loss}} = \frac{q^4}{6\pi\epsilon_0 m^4 c^5} E^2 B^2 \,, \tag{1}$$

where ϵ_0 is the dielectric constant, m is the rest mass of the particle, c is the speed of light, and E is the energy of the particle. Especially for low-mass particles, these losses are very high. For electrons, whose rest mass is roughly a factor of 2000 smaller than the rest mass of protons, linear accelerators are a competitive design. For the heavier hadrons it is preferable to accelerate them in ring-like structures where they can pass the devices used for acceleration (usually radio-frequency cavities) multiple times. The maximum velocity of a particle depends on the magnetic field strength and the size of the accelerator. The same statement also holds for astrophysical accelerators as Hillas pointed out in [Hil84] reasoning that a particle must be confined to the accelerator region for a certain amount of time to reach a certain velocity. A charged particle with Larmor radius, r_{L}, larger than half of the acceleration region of the source, L, will leave the accelerator for geometrical reasons and not gain any additional velocity. The Larmor radius is

$$r_{\text{L}} = \frac{\gamma m v_{\perp}}{|q|\, B} \,, \tag{2}$$

where γm is the relativistic mass. It follows that a particle can reach a maximum energy of

$$E_{\text{max}} = \frac{1}{2} LB\,|q| \,. \tag{3}$$

Figure 3 shows the relationship of the magnetic field and the size of different types of source candidates. Two lines are drawn, which indicate the minimum field strength at size L needed to accelerate a proton or an iron nucleus to 10^{20} eV. We can derive that only a few sources like active Galactic nuclei or neutron stars can possibly contribute to the cosmic

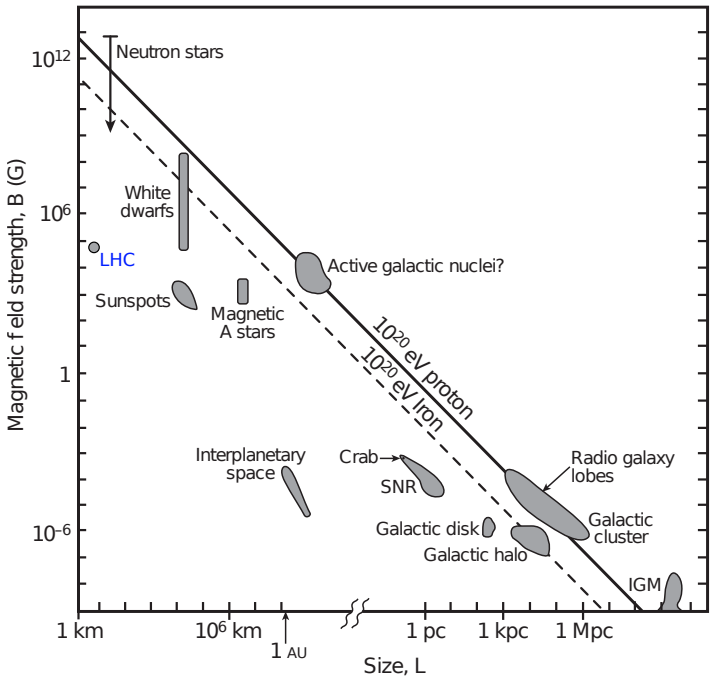


Figure 3: The LHC and astrophysical objects are shown according to their size r_S and magnetic field strength B. Accelerators found above the solid (dashed) line can accelerate a proton (iron) particle to 10^{20} eV. Crab indicates the Crab nebula; SNR, supernova remnants, and IGM, the intergalactic magnetic field (modified from [BM09]).

ray flux in the GZK energy region. The point for the CERN LHC marked in blue stresses that man-made accelerators are by far not capable of accelerating particles to such high energies. The field strength would need to increase by about 6 orders of magnitude for an accelerator with the size of LHC.

Apart from geometrical constraints due to magnetic confinement, the acceleration mechanism also depends on the size of the source. In the same paper [Hil84], Hillas showed that the mean escape time, which depends on diffusion processes, should be small compared to the acceleration time, which is determined by the speed in units of c of the shock wave β_{sh}. This modifies Eq. (3) to $E_{\max} = \frac{\beta_{\mathrm{sh}}}{2} LB|q|$. The optimal acceleration is achieved by $\beta_{\mathrm{sh}} = 1$ but, for example, the brightest recorded supernova SN1006 (Tycho remnant) is found to have a mean shock velocity of (5160 ± 70) km/s [HFB07]. Even if the LB requirement would be met, with $\beta_{\mathrm{sh}} \simeq 1/60$ it could accelerate particles only to 60 times lower energies than derived from the confinement constraint.

Consequently, sources and acceleration of cosmic rays remain a field of active research. In fact, to be able to draw a more complete picture one has to understand an interplay of many vital components of source properties and observables measurable at Earth. The list of components, which is by no means exhaustive, includes:

- the distance of the source to the observer is important since cosmic rays interact with a magnetic field and photons of the CMB and background in the radio frequency,
- the metallicity of the source environment, which is also connected to

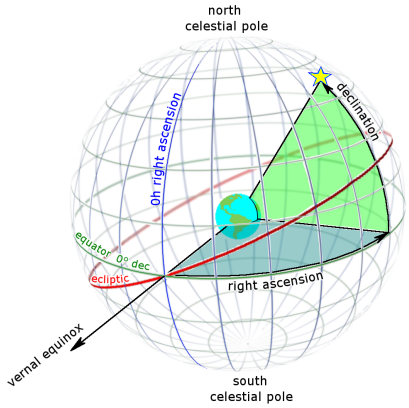


Figure 4: Example for right ascension and declination angles to an object (yellow star). (from wikimedia.org)

- the mass composition and the spectral shape produced for each nucleus with a certain mass. Additional observables are affected by the aforementioned items, e. g.,
- the observed mass composition, which differs from the composition at source due to propagational effects and diffusion processes in galaxies, and lastly
- large-scale anisotropies that can identify arrival directions with origins from large structures like the Galactic plane.

Composition studies and measuring the large-scale anisotropy are the most widely used UHECR analyses to scrutinise astrophysical scenarios.

2.1.2 LARGE-SCALE ANISOTROPY

Variations in the rate of cosmic ray particles at different solid angles can be measured by means of periodic modulation in time due to the rotation of Earth around the sun and around itself. The sidereal time based on the rotation of Earth relative to fixed stars as well as a coordinate system expressed by right ascension α and declination (compare Fig. 4) are the basic tools for determining the anisotropies. In this coordinate system the Galactic centre is located at right ascension $\alpha_{GC} \simeq 268.4^{\circ}$.

The key is to study the right ascension of all cosmic rays (recent studies, e. g. in [Pie+11]) and determine the amplitude and phase of the Fourier expansion. In the left panel of Fig. 5 the amplitude is shown as a function of energy. The measurement does not deviate greatly from the isotropic prediction (dashed line) and it follows that no anisotropies have been discovered yet. However the phase shows a clear transition from $\alpha \simeq \alpha_{GC}$ at $1{\times}10^{18}$ eV to $\alpha \simeq 100^{\circ}$ above $5{\times}10^{18}$ eV. The right panel of Fig. 5 shows the limits on anisotropy that, in fact, are in some energy ranges sensitive to current predictions for Galactic or extragalactic components. One can conclude that large-scale anisotropy due to UHECR has not been found but the current measurements can be used to constrain astrophysical scenarios.

Finding point sources could establish a direct link to the source objects. At the highest energies, e. g., the Pierre Auger Collaboration searches for excess fluxes of cosmic rays or

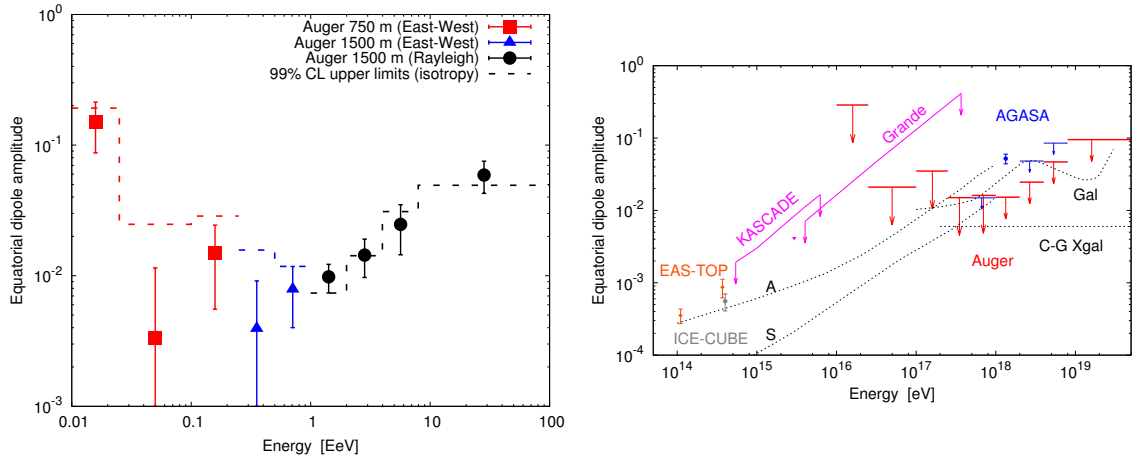


Figure 5: Left panel: Measured amplitude of large scale anisotropy as derived by the Pierre Auger Collaboration. The isotropic expectation is shown as a dashed line. Right panel: Upper limits for a wider energy region obtained by different experiments. Two model predictions that differ in their magnetic field model are shown for a purely Galactic origin and the horizontal line corresponds to an extragalactic component isotropic in the CMB rest frame. (from [Aab+13] and references therein)

so-called hotspots from potential sources. No deviation from isotropy within the statistical uncertainties is found at the studied parts of the sky. Fig. 6 shows that neither an excess is found for the closest radio galaxy to Earth, Centaurus A, nor for the Galactic centre on any angular scale.

The Telescope Array Collaboration has recently revealed a study of the northern part of the sky, which is not visible to the Pierre Auger Observatory, and discovered a clustering in the arrival direction of cosmic rays above 57 TeV. The clustering is confirmed with 5.1σ. Nevertheless, there is still a 0.04 % probability that such a clustering appears from an isotropic sky. If the hotspot was real, the interpretation is difficult as no astrophysical structure in the list of astrophysical source candidates is known along the trajectory [Abb+14a].

2.1.3 MASS COMPOSITION

With the current statistics gathered by the experiments measuring air showers in the energy region of the ankle, the mass composition upon arrival at Earth can be probed with good accuracy. Results from multiple analyses constrain the mass composition. Most notably those based on measuring the distribution of the maximum of the longitudinal profile of electromagnetic particles X_{max} and those measuring the muon content of the air shower at the ground level. Recent results from the Pierre Auger Collaboration and Telescope Array Collaboration can be found in [Aab+14a; Abb+14b]. The experimental results differ between the Telescope Array, which is situated in the northern hemisphere, and the Pierre Auger Observatory, which is situated in the southern hemisphere. Whereas the former sees

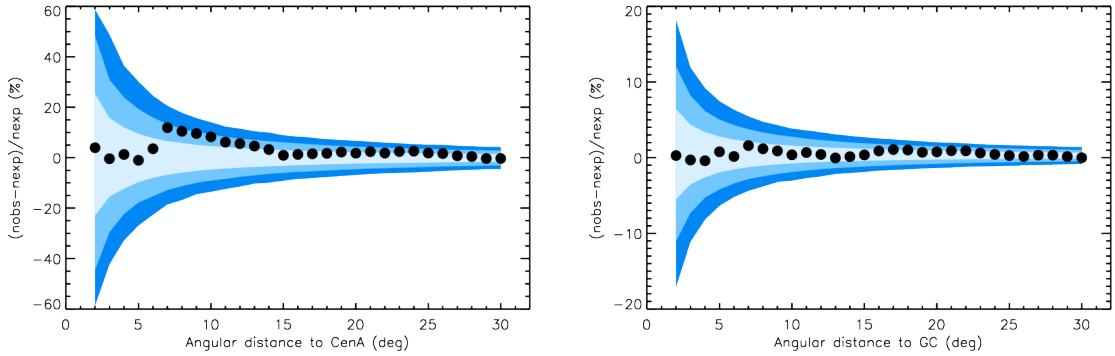


Figure 6: Difference in observed events from expected isotropic background. The colours indicate 1, 2, and 3 standard deviations. Left panel: Centaurus A for energies above 8×10^{18} eV. Right panel: Galactic centre for energies between 1 and 2×10^{18} eV. (from [Aab+13])

a dominating light component at energies above 10^{19} eV, the latter sees an increase in the mass of the particles.

This measurement of $X_{\max}$ is based on how deeply air showers penetrate the atmosphere. When a heavy nucleus initiates the shower, it will develop more quickly compared to one induced by a light ion with the same energy. Given the same primary energy of the nucleus, each individual nucleon only carries $1/A$ its energy. The cross section increases, however, only slowly (roughly logarithmically) with $\sqrt{s_{NN}}$. The geometric effect is therefore more important resulting in larger cross sections at the same primary energies for heavier nuclei. Multiple nucleons can simultaneously be involved in one collision. Consequently, the shower maxima of iron primaries are observed at shallower depths compared to proton primaries. The maxima can be directly studied with fluorescence telescopes. These data are compared to model predictions of shower maxima of Fig. 7. The data of the Pierre Auger Observatory and the Telescope Array are close to the model predictions for proton primaries and agree with predictions of heavier nuclei for energies above 10^{19} eV. Another well-established method is comparing the standard deviation of the $X_{\max}$ distributions, which leads to similar results. The detector effects are not unfolded in the data of the Telescope Array to maximise the number of events. Thus, the absolute values cannot be compared between the two experiments.

One way to circumvent this problem is to study the average mass of the cosmic ray ensemble or the average of its natural logarithm that is favoured by a particular hadronic interaction model. There is a linear relation between $\langle X_{\max}\rangle$ and a model dependent estimator of the mass, $\langle \ln(A)\rangle$, of the cosmic rays.

The observed $\langle \ln(A)\rangle$ is as high as nitrogen for some models at $10^{19.4}$ eV [Aab+14a]. Looking at the variance of $\ln(A)$, one finds that the composition is remarkably pure with a variance of typically less than 1. This variable can easily be compared between different experiments. For the highest energies the results are shown in Fig. 8. The uncertainties of the Telescope Array measurements are large but the HiRes experiment finds consistent

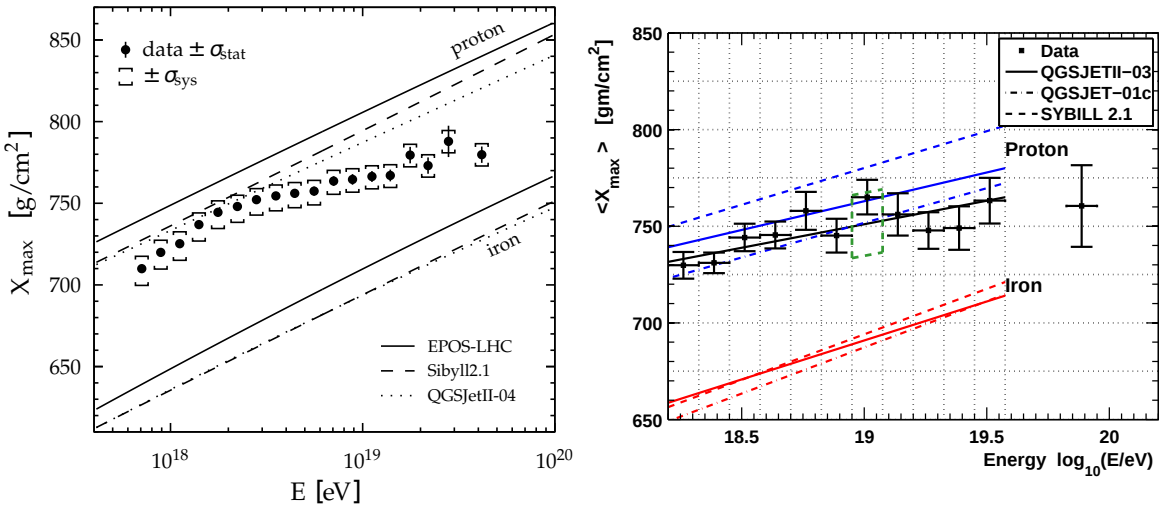


Figure 7: Measured average of the distribution of shower maxima ($\langle X_{\max} \rangle$) as a function of energy. Left panel: Pierre Auger Observatory. Right panel: Telescope Array. (from [Aab+14a; Abb+14b])

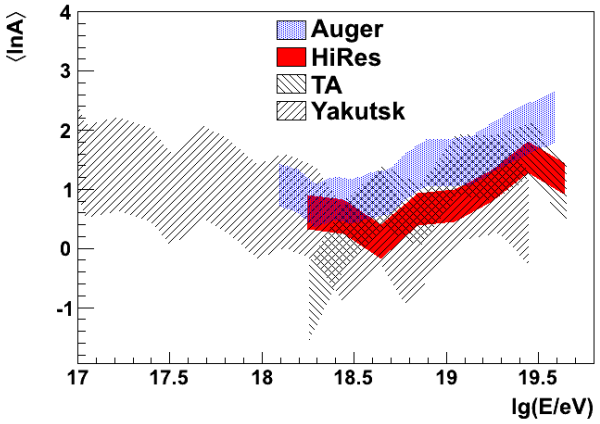


Figure 8: Comparison of average composition ($\langle \ln(A) \rangle$) measured by four experiments. (from [Bar+13])

$\langle X_{\max} \rangle$ values. HiRes was located at the same site in the northern hemisphere. Both results favour a lighter composition and are in tension with the data from the Pierre Auger Observatory at the highest energies. A close source seen only from either one hemisphere can produce a bias in one of the observed compositions. The collaborations of the two running experiments work together in an effort to understand the differences between their measurements. The authors of [Abu+13] came to the conclusion that the Telescope Array analysis is not sensitive to distinguish between a pure proton composition and an ad-hoc mixture of four components that describe well the mass composition seen by the Pierre Auger Observatory. On one hand, improved event statistics are needed and on the other the spread of the predictions of hadronic interaction models of the shower maxima hamper the interpretation of the composition. The differences of the predicted $X_{\max}$ distributions for one particular primary show large deviations as can be seen in Fig. 7. Assuming perfect

hadronic interaction models, this spread should be zero. The differences emerge from the underlying theory of how the collisions in air showers are described in interaction models and they can only be reduced by a better understanding of the physics at small scales in an interaction. We will come back to this notion later but already here it is clear that a more accurate and consistent prediction of $X_{\max}$ will improve the measurement of the cosmic ray mass composition.

The importance of the interaction models is studied further by looking at the muon content. It shall serve as an example to the plethora of methods based on observables that are sensitive to the mass composition. We choose this example since it is closely connected to a major caveat in simulations of air showers. In an analysis involving the surface and fluorescence detectors, the amount of muon content at the ground level with inclined air showers ($62^\circ < \theta < 80^\circ$) [Aab+15a]. For inclined showers many of the charged pions that are produced around the shower maximum have already decayed into the muonic component. The muon dominated signals are measured as a function of the zenith angle and distance from shower core with the surface stations. The signal is scaled by N_{19} to match that of QGSJETII-03 at 10^{19} eV and relates the density of muons in the reference model. The estimator is called R_μ when it is corrected for detector effects. The advantage of using this estimator is a weak dependence on mass and energy. The result of the measurement is shown in Fig. 9 at a primary energy of $E = 10^{19}$ eV. The measurement of R_μ lies outside of the allowed region by the interaction models for any possible primary mass. EPOS-LHC, which generates events with the largest muon content, deviates by 1.4σ. The different versions of QGSJET all show stronger deviation with the latest version QGSJETII-04 being the closest to the measurement. The number of muons, under the assumption that the uncertainties on the Auger energy scale are correct, need to be scaled by 30 % to 80 %. This points to a serious problem in the description of muons in hadronic interaction models and is not the only observation showing this [Ape+11b].

2.1.4 MULTI-MESSENGER OBSERVATIONS: NEUTRINOS AND PHOTONS

At lower energies, it is a well-established method to combine multiple experimental observations. Examples are full sky surveys at infrared, visible, and radio frequencies or the follow-up studies of sources found by very high energy gamma ray telescopes with X-ray observations to study local phenomena. For ultra-high energies, the field of multi-messenger astronomy is only in its beginning. One can combine studying photons, neutrinos, and cosmic rays to benefit from the very different information gain of their detection. Where as photons and neutrinos point back to their sources, the charged cosmic rays are deflected in the magnetic field of the interstellar medium. Consequently, they contain information about this medium and their mass also reflects the environments at their origin. A problem for studying photons are many different emission processes that result in similar and very low expected rates. Also very high energy neutrinos have a low flux; nevertheless, experiments with large fiducial volumes can detect them, which leads to strong constraints

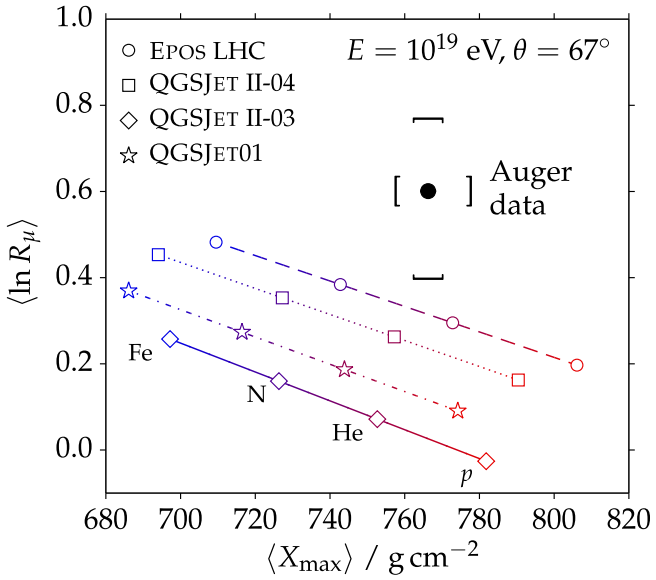


Figure 9: $\langle \ln(R_\mu) \rangle$ (see text for description of R_μ) compared to $\langle X_{\max} \rangle$. The outer model predictions are from CORSIKA simulations whereas the inner points are estimated by a Heitler model. (from [Aab+15a])

on their propagation. Pinpointing the source with neutrinos is difficult for some of the events (e. g., ν_τ interactions). All three messengers combined provide the best picture of astrophysical scenarios.

Plans have therefore been made to study these messengers in combined analyses. For instance, common hotspot searches for which IceCube, a neutrino experiment located at the South Pole, could play a major role. Recently, astrophysical neutrinos [Aar+15] have been discovered in the IceCube data [Aar+15], which adds a new piece to the great puzzle. The neutrinos are produced in the same source region, where cosmic rays are accelerated. Candidates of sources are active Galactic nuclei, radio galaxies, star-forming galaxies, and clusters. From, for example, diffuse shock acceleration one expects a flux $\propto E^{-2}$ for neutrinos of all three generations. They are produced when protons interact in the radiation fields. The neutrinos oscillate between three generations on their way to Earth changing the ratio from $\nu_e : \nu_\mu : \nu_\tau = 1 : 2 : 0$ at the source to an observable ratio of $1 : 1 : 1$. At energies of 10^{13} eV to 10^{15} eV, IceCube detects such neutrinos, which are typically produced from protons that have energies about two orders of magnitude above the neutrino energy. These observations must be consistent with measurements of UHECR. In Fig. 10, the IceCube measured neutrino flux is brought in perspective with other measurements of the cosmic ray flux. The simulation curves deliberately neglect neutrino production in the source region and, thus, show the requirement of such a missing component as the predicted flux is magnitudes below the measured flux by IceCube. All other parameters have been set to agree with the measured cosmic ray flux by the Pierre Auger Observatory.

Combined hotspot searches are in preparation merging data of the Pierre Auger Observatory, IceCube, and Telescope Array experiments. With multi-messenger input the sky is covered completely, and stronger constraints can be derived. Most notable is the AMON system [Smi+13], which will enable highly-efficient exotic searches for neutrons or black

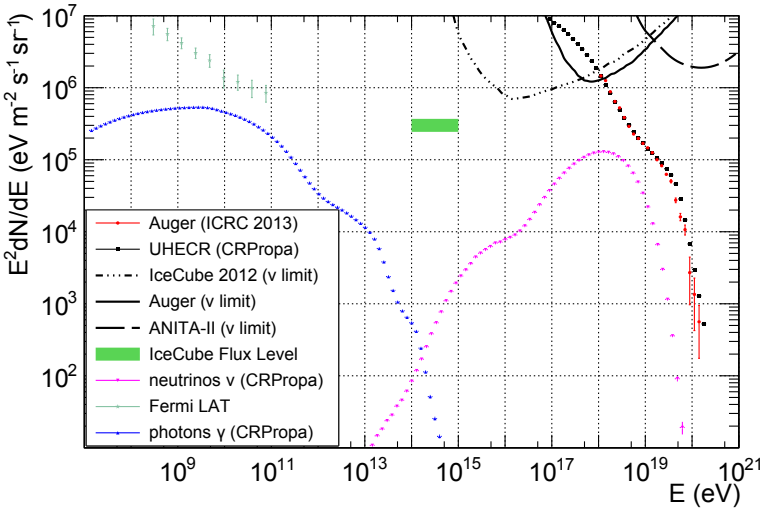


Figure 10: Comparison of measurements of the cosmic ray and neutrino fluxes with simulations. Proton sources are simulated with an injection spectral index of 2.4 and $E_{\max} = 200\,\text{EeV}$ within the GRB2 scenario. (from [SV14] and references therein)

holes as it provides a possibility for rapid follow-up investigation on interesting events triggered by the Pierre Auger Observatory. The sensitivity to such rare particles can increase dozens of times when requiring coincidence in events of two different experiments. It is these kinds of measurements that will help to constrain the astrophysical scenarios in the near future.

2.1.5 NECESSITY OF PREDICTIONS FROM HADRONIC INTERACTION MODELS

Before constructing an expensive UHECR experiment, hadronic interaction models are employed to explore its science case and its predicted sensitivities. The design reports of the Pierre Auger Observatory and the Telescope Array contain constraints derived from simulated air shower events that provide valuable information on the construction, layout, and detector design.

More importantly, later, when data is available by the constructed UHECR experiment, the interpretation of those measured relies on hadronic interaction models in various parts of the analysis. Calibration constants and reconstruction biases are estimated from Monte Carlo (MC) based methods that make use of these models. One example is the determination of the invisible energy to the detectors of fluorescence telescopes. It is derived from simulating the response of the fluorescence detector to the produced fluorescence light of various particle types that are predicted by hadronic interaction models. Nowadays, methods exist to reduce uncertainties with data from the complementary surface detectors in a hybrid experiment but parameters like the critical energy for pions is still estimated with models introducing uncertainties around one percent.

Detector efficiencies and the fine-tuning of specific selection thresholds also involve MC methods. With typically three to five models being available, the analyser has to rely on the assumption that the calculated values bracket the true value on both sides and no

bias is introduced by taking the central value of the predictions. This has two caveats: Firstly, the theory behind soft hadronic interactions needs input from data as is shown later and is not always well constrained by it. Secondly, the underlying phenomenological theory is identical for multiple models, as is often the case for hadronisation of particles or approaches to treat low-mass diffraction.

The most relevant dependence on model predictions is the astrophysical interpretation of air shower data. This is evident from the example of the mass composition measurement based on the longitudinal shower development. Figure 7 shows data from the Pierre Auger Observatory at energies above 10^{18} eV. The measurement of $\langle X_{\mathrm{max}} \rangle$ is compared to predictions of two primary components. If data are in better agreement with the proton predictions, one can assume a lighter component, and if closer to the iron prediction, a heavier component. Recent publications compare to predictions for more different masses at lower energies where a few million recorded events are available for analysis, many different mixtures of mass components can be fit to the measured flux [Pro+14; Ape+13].

There is only a limited number of models available that are actively developed and give a good description of data measured by collider and cosmic ray experiments. The list comprises: DPMJET, EPOS, QGSJET, and SIBYLL (Section 3.3). Often comparisons to data are made with older versions of the models for technical reasons, and corrected errors in the code of the model and retuned parameters to more recent data are neglected. By tuning, one refers to the process of changing available parameters of the model in order to describe measurement results. In general it is preferable to use recent model versions for astrophysical interpretation. The exception remains when correcting for detector effects: here a model generating events as close to the measured distributions seen in the detector quantities should be the preferred choice.

To quantify how important the hadronic interaction models really are, an example of determining the muon production depth [Aab+14b] is studied in more detail. The maximum of the muon production, X^{μ}_{max}, is also a mass sensitive parameter and can be calculated by the arrival times of muons at the detectors at ground combined with an accurate reconstruction of the direction of the shower axis. This measurement is performed on inclined showers, that travel trough a bigger part of Earth's atmosphere and, thus, contain more muons at ground. In terms of model dependencies, the analysis is an extreme case and shows well where the uncertainties can arise from. Uncertainties due to (1) detector effects and (2) astrophysical interpretation shed light on where hadronic interaction models matter.

Attributable to the first category is the difference in the reconstructed X^{μ}_{max} compared to a simulated input X^{μ}_{max} of an air shower. Here, model differences introduce an effect of $10\,\mathrm{g\ cm^{-2}}$ on the measured X^{μ}_{max} that is almost energy independently found to be around $500\,\mathrm{g\ cm^{-2}}$. Thus, 2 % uncertainty is introduced by the models. The total systematic uncertainty is 3.4 %. There are further hidden uncertainties that depend on MC generators. One example is the parametrisation of the velocity of the muons depending on the production

point within the air shower. The method used introduces small biases when compared to generator predictions [CVZ05].

The interpretation of $X^{\mu}_{\max}$ has much larger uncertainties. The result of the measurement (Fig. 8 in [Aab+14b]) is presented in the energy range from roughly 2×10^{19} eV to 6×10^{19} eV overlaid with model predictions from EPOS-LHC and QGSJETII-04. Both models have been tuned to data measured by experiments at the LHC. Nevertheless, there is an almost energy independent spread of $50\,\mathrm{g\ cm^{-2}}$ between the two models for both proton and iron primaries. Iron induced showers have a lower $X^{\mu}_{\max}$ but the measurement is at all energies below the iron prediction from EPOS. The authors cannot present a conclusion on the mass composition since up to now it is unclear what causes the muon production depth to be too high for the EPOS model.

It is evident that hadronic interaction models need to be improved. To do this, one needs to understand why they are needed in air shower simulations and which theoretical framework they are based on.

2.2 PHYSICS OF EXTENSIVE AIR SHOWERS

Extensive Air Showers (EAS) are initiated by ultra-high energy cosmic ray primary particles that interact with the atmosphere to form huge particle cascades. In such collisions, many secondary particles are produced that lead to a rapid growth of the particle cascade. The shower particles are concentrated in a thin layer at the front of the shower development, which propagates at a speed close to that of light in vacuum. The number of secondary particles per slant depth, X, increases with depth up to the shower maximum, $X_{\max}$, and then decreases due to absorption. The electromagnetic ($e^{\pm}$ and γ), muonic, and hadronic components develop differently as they are associated with different physical processes. The final properties of the air shower are mainly determined by the type and energy of the cosmic ray.

The slant depth given in units of $\mathrm{g\ cm^{-2}}$ expresses through how much matter a particle passed. It includes already the density of the atmosphere in relation to the height above sea level. The shower axis of an air shower is defined by the point of entry in the atmosphere and the incident angle, θ, of a cosmic ray. In the flat atmosphere approximation, the depth along the axis to the ground is $\propto 1/\cos(\theta)$. For vertical showers it is around $1035\,\mathrm{g\ cm^{-2}}$ at sea level and one can express this in the mean number of hadronic interactions that take place: $12\,\lambda_{\mathrm{I}}$, with the hadronic interaction length λ_{I}. In simulations, the density profile of the atmosphere is modelled in layers by parametrisations of the exponentially decreasing densities, $\rho(h)$, with height h. One prominent example is the US standard atmosphere with 7 different layers for which ρ is determined from measurements. For a known ρ, the slant depth is expressed as an integral over the traversed distance, e. g., for a vertical shower $X(h) = \int_h^{\infty} \mathrm{d}h'\,\rho(h')$.

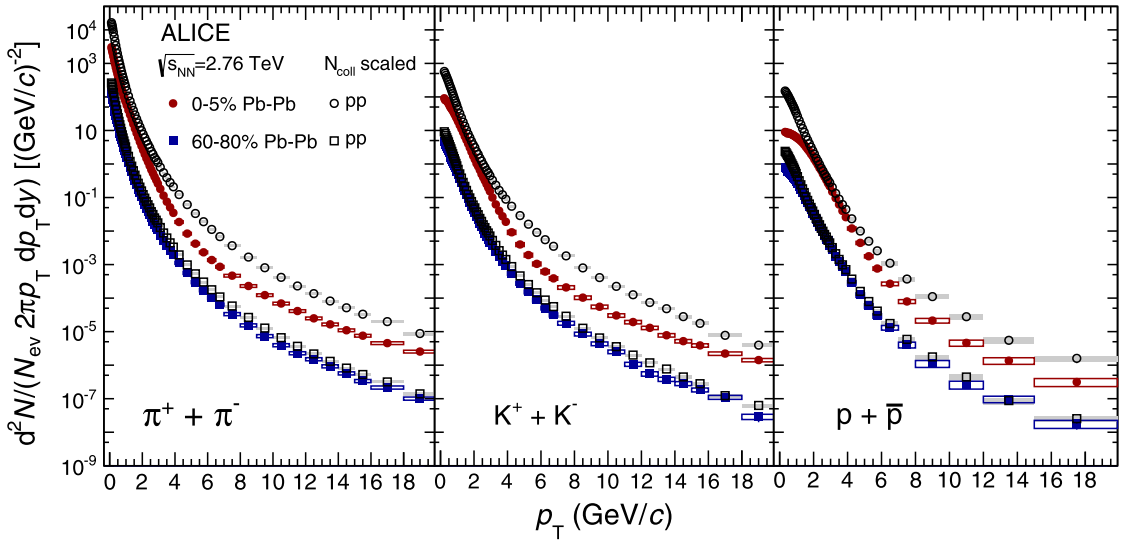


Figure 11: Transverse momenta of different particles measured by the ALICE experiment for proton-proton and lead-lead collisions (Pb-Pb), both with $\sqrt{s_{NN}} = 2.76\,\text{TeV}$. (from [Abe+14b])

At the highest energies cosmic rays are protons or heavier nuclei. Photons and neutrinos are theoretically predicted but have so far not been found. Interactions of hadronic particles are governed by processes that can be described with Quantum-Chromo-Dynamics (QCD), the theory of strong interactions. It describes the interaction of quarks and the force carriers, gluons, which carry the quantum number colour. The colour is introduced as an additional gauge degree of freedom to explain how quarks with the same spin can be confined to the small space of particles like baryons with 3 constituent quarks (firstly introduced in the 1960s to explain the Δ^{++} that has three spin-parallel up quarks) and mesons that contain two quarks. In air showers, QCD processes with low momentum transfers between the quarks take place that have to be treated with non-perturbative models. In such a collision, mainly protons (p), neutrons (n), pions (π), and kaons (K) are produced.

The longitudinal (along the shower axis) development of an air shower is determined by the specifics of these collisions, e. g., how often these particles scatter and how many particles are produced in each interaction. Usually, large numbers of particles are produced per collision, but often diffractive processes produce one *leading particle* that carries a substantial amount of energy compared to the other products. The leading particle can efficiently transport energy to large slant depths. The lateral (perpendicular to the shower axis) distribution of an air shower is determined by the transverse component of the momentum, p_T, of produced particles. The distribution of p_T in a collisions is measured at collider experiments for different particle types and decreases rapidly with p_T (Fig. 11). Combining the longitudinal and lateral characteristics leads to a thin shower front of $\simeq$1 m thickness at the centre of the front with diffuse tails that can spread to many kilometres in size. To sample an extensive air shower at the highest energies, large detector arrays with distance on the order of kilometres between measurement stations are built.

Charged pions decay via $\pi^{\pm} \rightarrow \mu^{\pm} + \nu_{\mu}$. The resulting muons ($\mu$) travel in straight lines without immediate further interactions while slowly losing energy via ionisation. This applies to all muons with energies close to the minimum-ionising-particle energy of a few hundred MeV up to almost 1 TeV. In this energy region energy loss per depth is almost constant. Neutral pions will decay via the preferred (98.8 %) branching ratio $\pi^0 \rightarrow \gamma + \gamma$ with a very short lifetime of $c\tau = 8 \times 10^{-17}$ s $= 25.1$ nm compared to 7.8 m for their charged counterparts. In an air shower, they will immediately transfer energy into photons in support of the electromagnetic component of the cascade. The photons will undergo pair production $\gamma + \mathrm{A} \rightarrow e^+ + e^- + \mathrm{A}'$.

It is useful to describe interactions in the centre-of-mass frame of the participating particles. One can change the reference frame with a Lorentz transformation. In the following numerical values are given for a charged pion with $E = 10^{15}$ eV as an example for these calculations. At this energy, the pion has a Lorentz or γ-factor of

$$\gamma = \frac{E}{m} = \frac{1}{\sqrt{1-\beta^2}} = 7.2 \times 10^6 \,, \tag{4}$$

with the rest mass m and $\beta = v/c$. According to special relativity, the observed distances the pion traverses are contracted by this factor. Thus, the mean travelled distance, ℓ, before it decays is

$$\ell = \gamma c \tau = 57\,753 \,\mathrm{km} \,, \tag{5}$$

where τ is the mean decay time. The high-energy pions have a very small probability to decay within the atmosphere and are more likely to interact. The Lorentz boost can be expressed by the Lorentz-invariant rapidity of the pion with

$$y = \frac{1}{2} \log \left(\frac{E+p}{E-p} \right) . \tag{6}$$

The rapidity is also related to γ by $e^y = \gamma(1+\beta)$. A boost of a particle moving in x-direction is then defined as

$$P_{\mathrm{lab}} = \begin{pmatrix} \cosh y & -\sinh y & 0 & 0 \\ -\sinh y & \cosh y & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} P_{\mathrm{cm}} \,, \tag{7}$$

where P is the four-momentum vector and its subscript denotes whether it is given in the lab or the centre-of-mass (cm) frame. We can derive that a high energy pion will travel further than a low energy pion before decaying. At high altitudes, where pions have larger energies, the density of the atmosphere is small and an interaction is, thus, less likely. This leads to muons being created even early on during shower development. With a lifetime of $\tau = 2 \times 10^{-6}$ s, high energy muons can penetrate the ground and are detectable even hundreds of meters below the surface. For inclined showers, they are the most dominant component at the ground level.

Table 1: Properties for electromagnetic and hadronic processes for different materials.

Material	X_0 [g cm^{-2}]	E_{crit} [MeV]	λ_{I} [g cm^{-2}]	ρ [g cm^{-3}]	R_{M} [g cm^{-2}]
N_2	37.99	75.42	89.7	gas	8.78
W	6.76	7.97	191.9	19.3	18.0
Pb	6.3	6.9	131.0	11.34	13.89
SiO_2	27.5	50.58	97.8	2.6	11.34

At the slant depth X_{max} the shower reaches its maximum number of charged particles, N_{ch}. This happens when it becomes less likely for the charged particle to interact than to decay or recombine with a nucleus under emission of light. Whereas the former is relevant for pions, absorption is the main reason why the number of electrons that actually dominate the charged component decreases. The transition energy, for which ionisation losses become the dominant processes, is called critical energy and for electrons it is about 82 MeV and determines X_{max}.

2.2.1 INTERACTIONS OF ELECTROMAGNETIC PARTICLES

Two processes are important for electromagnetic particles: bremsstrahlung and pair production. While interactions of relativistic electrons with $\beta \simeq 1$ with a nucleus lead to ionisation losses, the more important effect is bremsstrahlung due to the low rest mass of the electrons of 512 keV. The electric field of the nucleus makes them lose kinetic energy by radiating photons. The ratio of the two effects is [Kle05]

$$R\left(\frac{\text{Brems}}{\text{Ion}}\right) \propto Z \times E/580\,\text{MeV}\,. \tag{8}$$

This can be written as $E_{\text{crit}}(e^-) = 580\,\text{MeV}/Z$, from which the previous mentioned critical energy in air is obtained for $Z = 7$. The radiation length, X_0, is the distance, after which electrons lose all but $1/e$ of their energy or $\simeq 7/9$ of the mean free path length μ_γ^0 for a photon. For high energy photons (>1 GeV), the relation between the radiation length an the mean free path is a good approximation since pair production dominates (cf. Fig. 12). We obtain the probability of a photon to interact with matter

$$P(X) = 1 - e^{-\mu_\gamma^0}\,. \tag{9}$$

Electromagnetic particles undergo scattering processes while they lose energy in matter. The Molière radius, R_{M}, is defined as the radius of a cylinder, in which 90 % of the energy of the initiated shower cascade is contained. It is also related to X_0 by $R_{\text{M}} \simeq 0.0265 X_0(Z + 1/2)$. An overview of these properties for different materials can be found in Table 1.

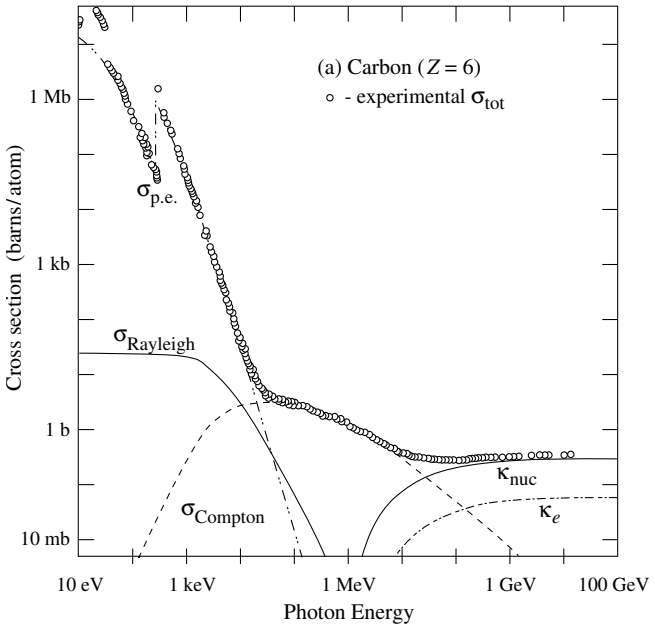


Figure 12: Cross section of a photon passing through carbon. The components of the total cross section are shown explicitly: photoelectric (p.e.) effect, Rayleigh scattering, Compton effect, and pair production (κ) due to the nuclear (nuc) and electric (e) field of a nucleus. (from [Nak+10])

2.2.2 DESCRIPTION OF HADRONIC AIR SHOWERS WITH AN ANALYTIC MODEL

Hadronic processes are not as well understood as electromagnetic processes and Section 2.3 is devoted to the theoretical overview. Nonetheless, some properties of air showers can be deduced from the two discussed electromagnetic processes, which is summarised in the well-known Heitler model [Hei50]. The Heitler model has been extended to describe hadronic processes [Mat05]. In this analytical calculation, a hadronic interaction produces only pions. Pions are the lightest mesons that have a charge of -1, 0, or 1. The charged pions have a mass of $139.6\,\mathrm{MeV}/c^2$, which is only $4.5\,\mathrm{MeV}/c^2$ heavier than their neutral counterparts. Therefore, the three types of pions are assumed to be produced with equal probability in the calculation. As neutral pions with a mean lifetime of $\tau = 8 \times 10^{-17}\,\mathrm{s}$ decay almost immediately into two photons, 1/3 of the energy in the hadronic component is transferred into the electromagnetic component at each interaction. In the model, the energy dependent interaction length, λ_I, defines after which depth the next interaction takes place. In [Mat05] it is actually $\lambda_\mathrm{I}/\ln 2$, where the superfluous factor $\ln(2)$ was kept from the original Heitler model. The multiplicity of secondary particle production is approximated by an energy dependent fit to hadronic interaction models. For hadronic interactions, the energy dependence of X_max for different primaries in the high-energy limit is calculated to be

$$\frac{X_\mathrm{max}}{\mathrm{g\ cm^{-2}}} = \begin{cases} -140 + 40\ \lg(E_0/\mathrm{eV}) & \text{for p} \\ -305 + 45\ \lg(E_0/\mathrm{eV}) & \text{for Fe}\,. \end{cases} \tag{10}$$

Similar dependencies are found within the full treatment of hadronic interaction MC programs, but the values for X_max are of lower absolute values (cf. Fig. 7). This could be

attributed to the effect of the leading particle as its occurrence is neglected in the assumption of equally shared energy between the secondaries. The step size and multiplicity were both fixed analytically in the original Heitler model due to the simplicity of the electromagnetic processes. It is still remarkable that properties such as $X_{\max}$ but also the number of muons and electrons can be derived from these basic assumptions.

2.3 INTERACTION OF HADRONIC PARTICLES

The technical framework to properly describe hadronic interactions in extensive air showers is based on the theory of QCD. Interactions at hadron collider experiments are of the same nature and QCD is an important ingredient in understanding proton-proton collisions at the LHC. Here, we will explain how the elastic, inelastic, and total cross sections can be predicted within this framework. Accurate measurements of soft collisions (small momentum exchange) are needed for a good description of cross section data. Regge proposed a theory how to describe high-energy hadron scattering in 1959, however, at that time, data were available at much lower energies. We will start from his early point of view and establish the theoretical need for a description with a *reggeon* exchange particle. Cross sections at high energies or even UHECR energies could only be described with the addition of the *pomeron* exchange particle. In the late 1960s, Gribov managed to formulate the Regge theory as a field theory. All these phenomenological tools explain soft interactions well and are still employed today. The Glauber model, which treats interactions between nuclei, is also important for UHECR interactions and lead ion collisions at the LHC. We show not only how its application leads to values for nuclear cross sections but also mention known caveats.

Hadronic interactions refer to the strong forces between quarks and gluons. These interactions are more relevant than electromagnetic processes for UHECR air showers. At high energies only protons and nuclei are observed to initiate air showers, and they create many hadronic secondary particles. The proton consists of point-like constituents called partons, a name introduced by Feynman. Point-like constituents were predicted by Bjorken, who suggested that variables, which later became part of QCD, scale with energy. At the SLAC and HERA colliders, extensive Deep Inelastic Scattering (DIS) experiments confirmed the scaling property. At these experiments, processes like $e^- + p \rightarrow e' + X$ are studied. The measurement leads to the conclusion that the variables are independent of the increasing resolution gained by higher energies and that the constituents are point-like. Scaling-violation exists only in some processes but the behaviour can be described by theory, yet, it is still an open research topic. The proton is made up of 3 valence quarks (up, up, down), a "sea" of quark-antiquark pairs, and gluons. Their individual number densities are measured by DIS experiments and are described by the parton distribution functions (PDFs) of the proton. They are derived dependent on the squared momentum transfers between the colliding particles, $Q = -q^2$. An example for $Q = 10\,\mathrm{GeV}^2$ is shown in Fig. 13. The

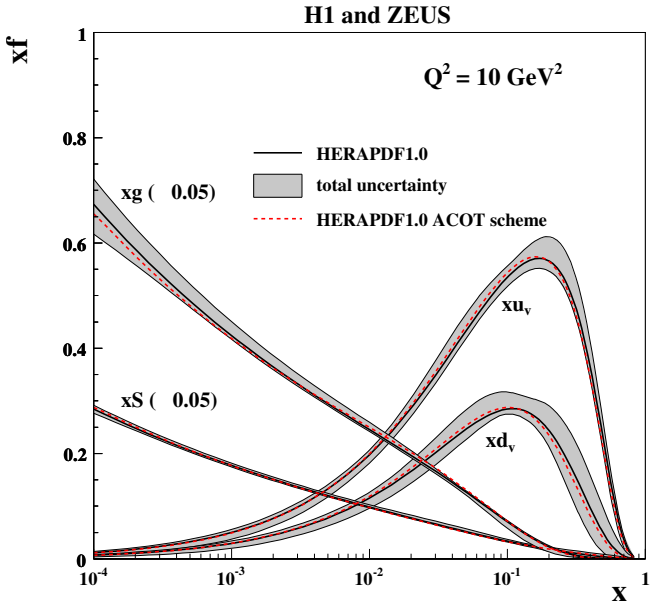


Figure 13: Parametrised PDFs multiplied by x shown as a function Bjorken x. The number density xf is given for the gluons distribution xg, the valence quark distributions xu_v and xd_v, and the sea quark distribution xS. (modified from [Aar+10])

figure shows the number densities as a function of Bjorken-x. The quantity is defined as $x = Q^2/(2pq)$, where p is the momentum of the electron and pq/M is the energy loss the electron experiences when being scattered. It is the ratio of the momentum of the parton to the total momentum of scattered particle. For elastic collisions $x = 1$ and $0 < x < 1$ for inelastic collisions. With this information, the PDFs shown in Fig. 13 can be interpreted. The valence quarks play the most important role at about $x = 1/3$ but towards low-x values the distribution decreases quickly. In this region, many gluons and sea quarks share the total momentum of the proton. The gluon density rises for $x \to 0$ and some theories predict that the gluon density reaches saturation for high energies. This is one of the puzzles to be solved by the experiments at the LHC and future colliders. Previous experiments at the Hera collider measured these distributions to high accuracy only for $x > 10^{-4}$.

The PDFs are necessary input for any hadronic interaction model. For MC generators they are used in selecting the momentum of the partons between which the interaction occurs. First of all, one needs to calculate the probability for an interaction to take place and this probability is expressed as the cross section.

2.3.1 CROSS SECTIONS

The probability of a process of any initial state i to a final state j, $i \to j$, is defined by the cross section

$$\sigma = \int \mathrm{d}t \, |f(s,t)|^2 , \tag{11}$$

where $f(s,t)$ is the forward scattering amplitude. The scattering amplitude is used to express a particle by its wave function $e^{ikr_z} + f(s,t)\, e^{ik|\vec{r}|}/|\vec{r}|$, with the wave number k and

the direction in which the particle is scattered $\vec{r}$. When the cross section of a process is known, the number of interactions for a beam with an instantaneous luminosity of $\mathcal{L}_{\text{inst}}$ is

$$N = \sigma \mathcal{L} \,. \tag{12}$$

For hadronic interactions the total cross section is split into three contributions from elastic, production, and quasi-elastic scattering

$$\sigma_{\text{tot}} = \sigma_{\text{el}} + \sigma_{\text{prod}} + \sigma_{\text{qe}} \,. \tag{13}$$

For elastic processes, the projectile and target keep their quantum numbers; the only changing property is the momentum of the particles. In inelastic collisions, particles are produced either via non-diffractive processes, which typically result in a large number of produced particles, or via diffractive processes, for which a gap is present in the rapidities among all secondaries. The former also results in large longitudinal momentum losses of the beam particles and therefore shows a small or no rapidity gap at all. One part of the inelastic cross section is determined by quasi-elastic (qe) collisions. They can occur for colliding nuclei, for which at least one nucleus is excited, breaks up, and leads to detectable hadrons. Defacto, no new particles are created in such a collision. The inelastic cross section is given by the sum of the production and quasi-elastic cross section.

The elastic cross section decreases with rising absolute value of the momentum exchange squared, $|t|$. It follows $\mathrm{d}\sigma/\mathrm{d}t = \exp(-B(t=0)\,t)$, with the elastic slope parameter B, which is with good approximation t-independent at small t. At a certain squared momentum transfer, $|t_{\text{dip}}|$, B becomes smaller, and a dip is visible (Fig. 14 left). The position of the dip is dependent on $\sqrt{s}$. For antiparticles, the dip is not as clearly pronounced.

Data from the TOTEM experiment that were taken in proton-proton collisions at $\sqrt{s} = 7\,\text{TeV}$ recently confirmed that $|t_{\text{dip}}|$ moves to even smaller values for very high energies. The reason for the dip is, however, largely debated, and many phenomenological models can describe the structure. Explanations reach from (a) tapping into three different regions of the proton, namely quark condensate for very elastic, a shell of charges of the valence quarks and the quarks themself at larger $|t|$ [Car13], over (b) a superposition of an exponential dropping component and an exponentially dampened oscillating component in the imaginary part of the scattering amplitude [DN12] to (c) interference terms [Don+02]. The latter explanation suggests that on the left side of the dip the C-parity of the dominant exchange is given by $C = +1$ and on the right side $C = -1$. This results in an mostly real on the left side and mostly imaginary phase of the scattering amplitude on the right. Destructive interference between the parts of the amplitude then lead to a dip in the elastic cross section. Unless structures, like a peak from a second oscillation, are experimentally observed, none of the theories is favoured. Useful is, nonetheless, an analogy to optics, in particular to interference structures of waves. With this crude but suitable analogy one can relate the slope at small-$|t|$ to the size of the target object on which the projectile diffractively scatters. It follows that $B = R^2/4$, where we can express the size $R = 1/m_\pi$ for a typical strong interaction where a pion is exchanged. We obtain $B = 12.5\,\text{GeV}$, which is a

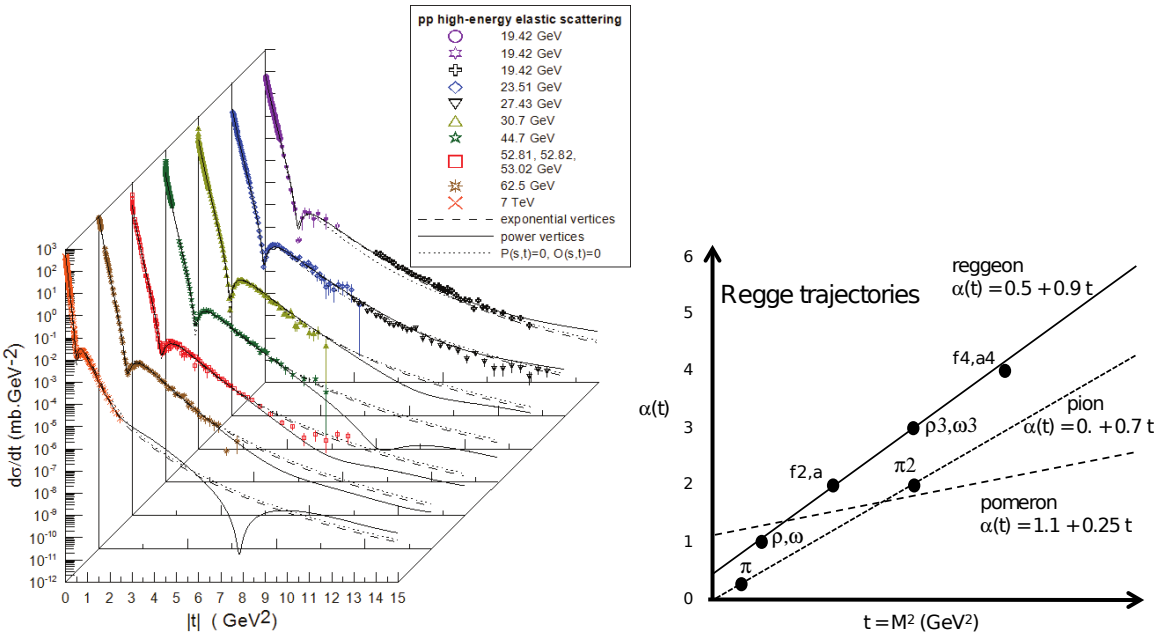


Figure 14: Left panel: The differential elastic cross section for different values of $\sqrt{s}$ that decrease moving further into the visible plane. Each measurement is fitted with several theoretical predictions. The data closest to the front were taken at the LHC. (from [Mar13])
Right panel: Two reggeon trajectories (solid lines) in the spin-M^2 plane for two types of mesons. A typical pomeron trajectory is drawn as a dashed line. (modified from [Car13])

good approximation and we also see that for increasing B at higher energies, the size of the interaction region must grow [Gou83].

Even more relevant is the analogy to the optical theorem. This theorem is used to calculate the total cross section from the measurement of the $|t|$-dependence of the the elastic cross section by, e. g., the luminosity-independent cross section measurement by the TOTEM collaboration. In the limit of high energies, the imaginary part of the scattering amplitude relates to the total cross section as

$$\sigma_{\text{tot}} = \frac{1}{s}\text{Im}\left(f(t=0)\right) . \tag{14}$$

Here, only the forward part of the amplitude is needed, which corresponds to no momentum transfer between the beam particles or likewise a scattering angle of $\theta = 0$. This can be exploited to measure the total cross section. For such a measurement, the elastic cross section at small-$|t|$ has to be determined accurately.

With a Fourier transformation, the scattering amplitude can be transformed from a function of (s,t) into a function of (s,b). Here, $b = |\vec{b}|$ is the absolute value of the impact

parameter and describes the separation of the particles perpendicular to their longitudinal direction. The two cross sections are

$$\sigma_{\text{tot}} = 8\pi \int \mathrm{d}b \, \mathrm{Im}\left(F\left(s,b\right)\right) b \tag{15}$$

$$\sigma_{\text{el}} = 8\pi \int \mathrm{d}b \, \left|F\left(s,b\right)\right|^2 b \,. \tag{16}$$

The relation $\mathrm{d}^2\vec{b} = b\mathrm{d}b\,\mathrm{d}\theta$ was used. The cross sections, especially in proton-proton collisions, have been measured by experiments up to LHC energies. The values of both decrease with centre-of-mass energy up to a few tens of GeV above which they increase again. This can be modelled by contribution from two opposing terms as we will see later. Parametrisations of the total, inelastic, and elastic cross sections are based on this concept. Two exemplary fits are given by the COMPETE Collaboration [Cud+02] and by a more recent publication that can be found in [MS13b]. Such a fit is shown in Fig. 15.

Antiparticles have a different behaviour at small $\sqrt{s}$. In the previous chapter we stated that sea quarks and gluons become important at high energies. Here, the reverse is implied: particle and antiparticle differ in the valence quark content resulting in a difference in their cross sections. With few assumptions, it can be shown that the cross section for particles and antiparticles is identical above a certain energy, which is known as the *Pomeranchuk theorem* (review about particle and antiparticle collisions can be found in [Mic69]). Additionally, at high energies for $s > s_0$ it cannot rise faster than $\propto \ln^2(s)$ [Fro61], which is known as the *Froissart bound*. Whereas it is unclear from the theoretical calculation how large s_0 is, experimental data at energies already available to us from current colliders show that the cross section for particle and antiparticle are similar.

The profile function $F\left(s,b\right)$ of Eq. (15) is a measure of the opaqueness of the interacting particles. In classical physics, interactions of two particles (denoted by 1 and 2) happen up to impact parameters of $b = R_1 + R_2$. In the case of scattered wave-like particles, this would correspond to $F\left(s,b\right) = 1$ if $0 < b < \left(R_1 + R_2\right)$ and 0 otherwise. Such an extreme scenario for colliding quantum mechanical particles is referred to as the *black disk* limit but it is unknown if this limit is reached at very high energies. As a demonstration, we use the charge radius of a proton at rest of about 1 fm as the black disk radius and calculate the total cross section. Integrating the black disk amplitude inserted in Eq. (15) leads to

$$\sigma_{\text{tot}} = 4\pi(R_1 + R_2)^2 \tag{17}$$

$$\Rightarrow \sigma_{\text{tot}} = 1.25 \times 10^{-31}\,\text{m} = 1.25\,\text{mb}\,. \tag{18}$$

Since such cross sections are surpassed even at very low energies, we can infer that the proton grows in size when colliding at larger $\sqrt{s}$. One should keep in mind that measurements at low energies show that the absolute value of the $F\left(s,b\right)$ is of Gaussian shape with a maximum <1 at the centre of the proton and extends further than the black disk radius.

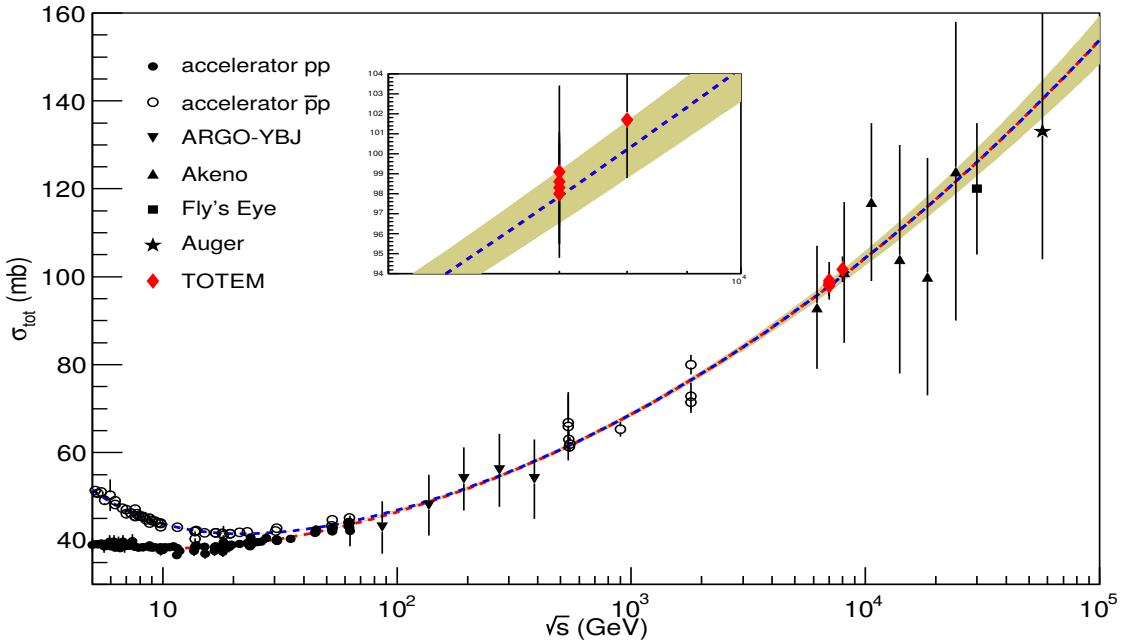


Figure 15: Parametrisation of the proton-proton and proton-antiproton total cross section measurements. (from [MS13b])

2.3.2 PARTICLE PRODUCTION IN REGGE THEORY

The cross section determines the probability of an interaction but details regarding momentum, number of particles, quantum number exchanges, and the type of produced particles must be determined otherwise. The usual approach is perturbative-QCD calculations but this is not possible for air showers and forward physics. The soft processes ($Q^2 < Q_0^2$) of interest cannot be calculated from first-principles. Equivalently, one can say phenomenology or fits to data are required in the calculations. Regge theory and its extension by Gribov, which use a phenomenological ansatz to describe the experimental data are used to calculate these properties. The extension by Gribov is a quantum mechanical description in the language of field theory referred to as Gribov-Regge Theory (GRT) [Gri68].

Hard processes with momentum exchanges above Q_0 can be calculated with perturbation theory of QCD. An observable, f, can be written as a series expansion in order of the strong coupling constant α_s: $f = \alpha_s f_0 + \alpha_s^2 f_1 + \alpha_s^3 f_2 + \cdots$. Often, high-order terms are neglected, and only the leading order is calculated. If so, a requirement is that higher terms vanish, which is true for $\alpha_s \ll 1$. The caveat are ultra-violet divergences that have to be absorbed by renormalising the coupling constant. The renormalisation introduces a dependence on Q as shown in Fig. 16. For small Q, the running coupling has a Landau pole and diverges. This behaviour makes it impossible to use perturbative QCD for soft interactions where distances are large and Q is small. Typical values for the limit are $Q_0^2 \simeq (1\,\mathrm{GeV/c})^2$.

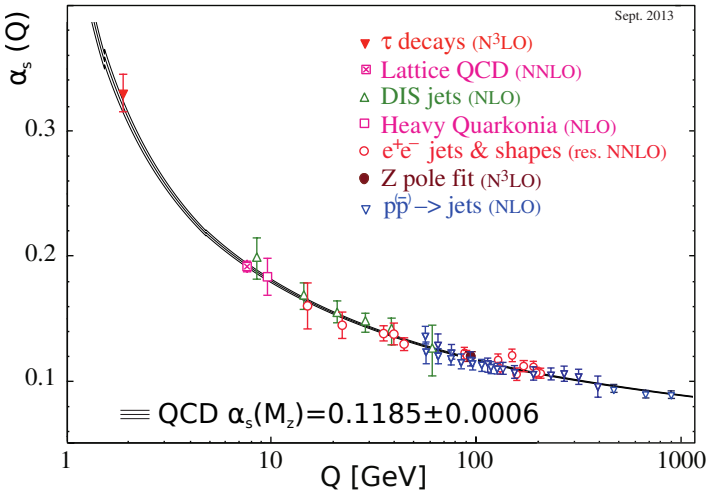


Figure 16: Dependence of the running strong coupling constant on Q. The best fit given at the bottom of the plot is determined by the measurements shown in the legend. (from Particle Data Group 2014)

One possibility is to calculate soft processes numerically on a lattice, a procedure going back to [Wil74] and others. Lattice QCD calculations, however, require long processing times not suitable for our purposes. Currently, the only viable option for air shower physics simulation programs is to use Regge theory with treating nuclear effects within the GRT.

To understand Regge theory, we will first look at the process $\pi^+ + p \rightarrow \Delta^{++} + X$. In fact, this process bears relevance for air showers[1]. As shown in Fig. 17, there is more than one possible exchange mode. To create a Δ^{++} from particles with $Z = 1$, charge exchange is required, which can be understood as an exchange particle with charge $Z = 1$ and quark content $\overline{u}d$. The two Feynman diagrams show a π^+ or a ρ^+ as the exchange particle. Both have the same I_z, but their total angular momenta are $J = 1$ and $J = 0$, respectively, producing as final state (X) either a ρ^0 or π^0 particle.

More generally, interactions can be described not by two particles but by a composite exchange particle called Reggeon, which includes all states on a reggeon trajectory. The trajectories have been discovered in the correlation of the spin of the particle and its mass. In this case a linear dependence is visible for pseudoscalar mesons ($J = 0$: $\pi, \eta, \ldots$), vector mesons ($J = 1$: $\rho, \omega, \ldots$) and so on. The trajectory can also be observed for baryons. The correlation is shown in Fig. 14 (right) where the x-axis shows the squared mass of the exchanged particle or likewise the momentum transferred in the interaction. With reasonable accuracy one can find a linear function describing the particles of each group (lines). These parametrisations $\alpha(t) = \alpha_0 + \alpha' t$ are referred to as Regge trajectories. A

1 This influences the muon number in air showers due to the quick decay with $c\tau \simeq 5{\times}10^{-24}$ s for the process $\rho^0 \rightarrow \pi^+ + \pi^-$ with a branching ratio of over 99 %. Including the process shown in the Feynman diagram on the right side of Fig. 17 in hadronic interaction models increases their predicted muon content. The process has only recently been included in an effort to conquer the problem of muon deficit in air shower simulations (Section 2.1.5).

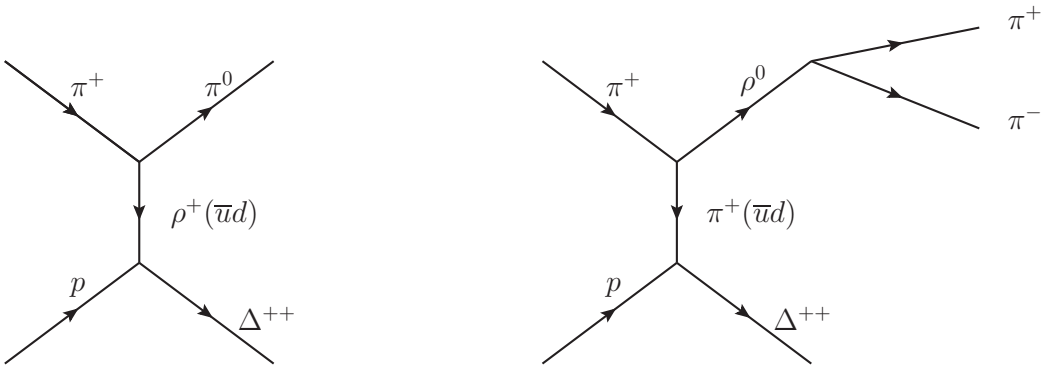


Figure 17: Pion-proton interaction with two different particles exchanged.

property of all trajectories is an intercept of $\alpha_0 < 1/2$ at $t = 0$. In Regge theory, one finds that the imaginary part of the scattering amplitude is given by

$$f(s,t) \propto i\, s^{\alpha(t)}\,. \tag{19}$$

With the optical theorem expressed by Eq. (14) it can be written as

$$\sigma_{\text{tot}} \propto \frac{1}{s}\, s^{\alpha(t=0)} \propto s^{\alpha_0 - 1}\,. \tag{20}$$

We immediately see that even with the largest $\alpha_0 = 1/2$ of the reggeon trajectories the cross section would diminish at high energies. In measurements, the total cross section decreases only up to roughly 30 GeV and then slowly rises. Another contribution that does not include any real particle states on its trajectory is theoretically proposed to explain this behaviour: the trajectory is attributed to the *pomeron* and usually assumed to also have a linear dependence with the additional requirement of $\alpha_0 \geqslant 1$. Most theories include a supercritical ($\alpha_0 > 1$) pomeron of the form $\alpha(t) = 1.08 + (0.25c^2/\text{GeV}^2)t$. The pomeron object has to have the quantum numbers of the vacuum and could be realised by a complex gluon-quark structure, e. g., gluon ladders where the rungs are made of reggeons. With an intercept of $\alpha_0 = 1.08$, the pomeron can explain the rising cross section when inserted into Eq. (20). The total cross section can now be expressed by a term coming from the reggeon contribution (r.h.s.) and one from pomerons (l.h.s.)

$$\sigma_{\text{tot}}(s) = X\, s^{\epsilon} + Y\, s^{-\eta}\,, \tag{21}$$

where the quantities X, ϵ, Y, η can be fitted to data. This form is used in many parametrisation of the cross section. The parametrisations describe the cross section well up to LHC energies. Sometimes an additional term corresponding to a hard pomeron is added to change the high energy behaviour.

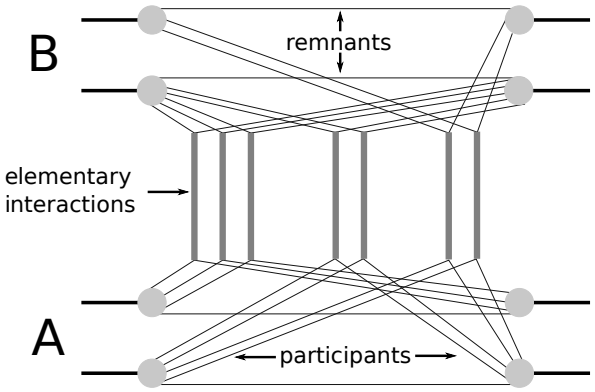


Figure 18: Illustration of an interaction between nucleons or nuclei A and B. The elementary interactions are exchanged pomerons between the participating partons. (from [Dre+01])

2.3.3 GRIBOV-REGGE THEORY: HADRON AND NUCLEUS SCATTERING

Exchanging one pomeron with $\alpha_0 > 1$ violates unitarity (the S matrix has to fulfil $SS^\dagger = \mathbb{I}$) at higher energies. This becomes clear in the eikonal[2] picture, where each of the n possible exchanged pomerons is expressed by the eikonal $\chi(s, b)$. The value of the eikonal increases with energy even to values above 1 for large s and small b. It is also defined to fulfil

$$F(s, b) = 1 - e^{i\chi(s,b)} . \tag{22}$$

By expressing the Fourier transformation of the profile function as its Taylor expansion, it becomes clear that the fit contains contributions from single and all possible multi-pomeron exchanges

$$F(s, b) = 1 - \sum_{n=0}^{\infty} \frac{(-i\chi)^n}{n!} = -\sum_{n=1}^{\infty} \frac{(-i\chi)^n}{n!} . \tag{23}$$

In models using GRT, the contributions to the total cross section from high-n terms rapidly vanishes. For example, the weight for an exchanged pomeron to contribute to particle production in the VENUS model for $n = 10$ are 1×10^{-4}, 3×10^{-4}, 1×10^{-3} for the respective collision energies of 19.4 GeV, 200 GeV, 6.3 TeV [Wer93], respectively. For all energies, about 50 % of the collisions are described by only one pomeron exchange ($n = 1$). Since the high-n fractions are small, it is possible to express multi-pomeron exchange by only a few eikonals, reducing the amount of necessary calculations.

The scattering of two nuclei, A and B, can also be expressed in the GRT approach. Similar to hadron scattering, the individual contributions are summed up, but instead they each express scattering of a nucleon of A with nucleons of B [Dre+01]. The schematic shown in Fig. 18 illustrates how an elementary interaction, described by multi-pomeron exchange between the *participant* nucleons can look like. *Spectator* nucleons exist and are not scattered. Lines denoting the remnant are drawn horizontally indicating that they are undisturbed by the interaction. For nuclei, remnants are lighter ions made up of the spectator nucleons and for hadrons, they are partons.

2 The word eikonal comes from the Greek *εικών*, which means image.

For nuclei, the cross section can be determined with the nuclear thickness function $T_A(b) = \int \mathrm{d}r\, \rho(b^2 + r^2)$, where the nuclear density ρ is Gaussian for nuclei with small masses and Wood-Saxon-like in radial distance away from the centre, r, for large masses. The nuclear density is typically a parametrisation stemming from data. The impact parameter difference between each of the N_{pair} interacting nucleon pairs Δb_i and the difference of the projectile and the impact parameter of the whole nucleus b (sometimes written as $\vec{b}$ to express the x-y-components) is $b_i = b + \Delta b_i$ with $i = 1 \dots N_{\text{pair}}$. Written in the form of a vector it is $\vec{b}_{\text{pair}} = \{b_1, \dots, b_{N_{\text{pair}}}\}$. The cross section can be expressed as

$$\sigma_{\text{inel}}^{AB}(s) = \int \mathrm{d}b \int \prod_{i,j} \mathrm{d}b_i\, \mathrm{d}b_j\, T_A\, T_B\, \gamma_{AB}\,, \tag{24}$$

with quantities of i correspond to the projectile A and quantities of j to the target B. The nuclear profile density function, γ_{AB}, contains the scattering amplitudes and is calculated in the parton-based GRT approach. The formalism is consistent between nucleon-nucleon and nucleus-nucleus interactions [Dre+01]. The profile function is calculated according to the framework of well-defined rules set by Abramovsky, Gribov and Kancheli (AGK) [AGK73].

An illustration of how event generators that use GRT sample the impact parameter distribution in a p-A collisions and ultimately produce final state particles is shown in Fig. 19. One can see that the geometrical parameters are very similar in all shown models but large differences exist for the final-state products in such a collision.

2.3.4 NUCLEAR EFFECTS: A GLAUBER CALCULATION

The nucleus-nucleus cross section can also be obtained using the Glauber model [GM70]. In this approach, the nucleus density profile, ρ, which describes the probability for the nucleons to be at a certain impact parameter is convolved with the scattering amplitude of the partaking hadron-hadron collisions.

We first look at hadron-nucleus scattering and derive the total and elastic cross sections by applying the optical theorem. The Glauber model assumes the nucleons are elastically scattered, referring to the process for which the final state $|f\rangle$ is identical to the initial state $|f\rangle \equiv |i\rangle$ and is determined by the elastic scattering amplitude $f^{\text{ii}}(s,t)$ defined in Eq. (11). In the following, the indices hA and hh will refer to nucleon-nucleus and nucleon-nucleon, respectively, and as before, the products and sums over j run over all nucleons $1 \dots N_{\text{nucleons}}$

$$F_{hh}\left(s, \vec{b}_j\right) = \frac{1}{2\pi i k} \int \mathrm{d}^2\vec{q}\, e^{-i\vec{b}\vec{q}} F_{hh}(s, t = q^2)\,. \tag{25}$$

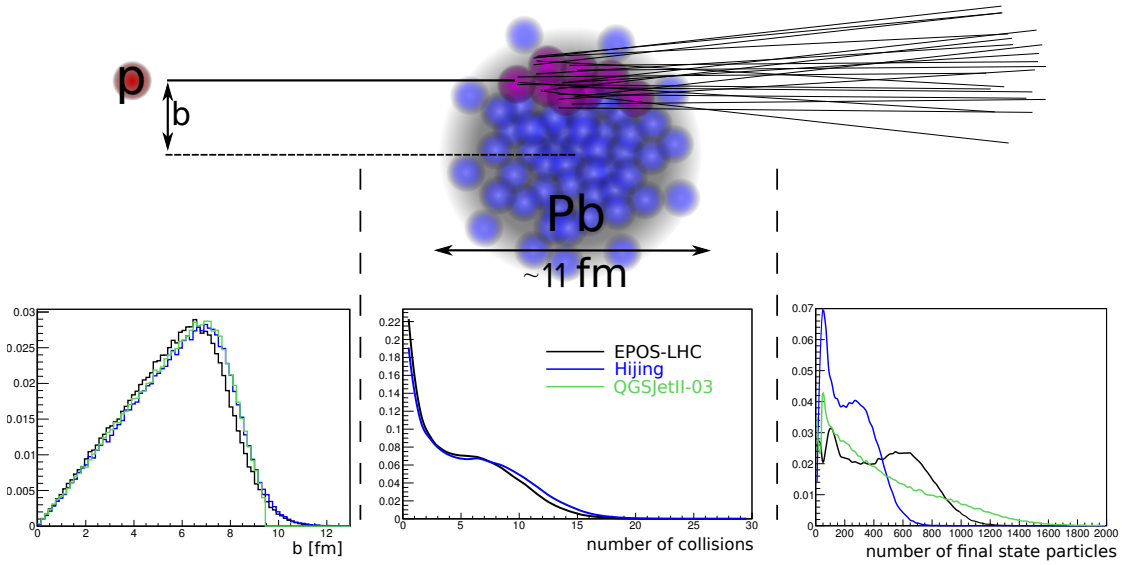


Figure 19: An example for a proton-lead collision (bottom). The proton (red) hits the lead ion at a certain impact parameter, b, away from the centre. The number of wounded nucleons (purple), which produce multiple secondaries shown as (black lines). In the centre-of-mass frame, these would most likely also be produced in the backward direction. Shown below the drawing are distributions for impact parameter, number of binary collisions, and number of final-state particles generated at $\sqrt{s_{NN}} = 5\,\text{TeV}$ for three different interaction models (Section 3.3).

Here, the impact parameter of each nucleon $\vec{b}_j$ is related to the global impact parameter and to the position of the nucleon $\vec{b}_j = \vec{b} - \vec{r}_{j\perp}$. An approximation that Glauber introduced is that multiple scatterings can be expressed as a product of the eikonals by using Eq. (22)

$$F_{hA} = 1 - e^{i\sum_j \chi_j(s,\vec{b})} \tag{26}$$

$$= 1 - \prod_j \left[1 - F_{hh}(s,\vec{b}_j)\right] . \tag{27}$$

To express F_{hA} with the global impact parameter $\vec{b}$, it has to be folded with the nucleus wave function

$$F_{hA}(s,\vec{b}) = \langle i|\, F_{hA}(s,\vec{b}_j)\, |i\rangle . \tag{28}$$

For now, we neglect correlations between nucleons that feel a repulsive force at small scales and perform a dedicated study of the impact of this approximation later. The nuclear density, ρ_j, of each nucleon j is then independent of the other nucleons and

$$\langle i|i\rangle = \prod_j \rho_j . \tag{29}$$

Equation (28) simplifies to

$$F_{hA}(s,\vec{b}) = 1 - \prod_j \left[1 - \int \mathrm{d}^3\vec{r}_j\, F_{hh}(s,\vec{b}_j)\, \rho_j(\vec{r}_j)\right] . \tag{30}$$

Applying the optical theorem Eq. (15) yields the total and the elastic cross sections

$$\sigma_{\text{tot}}^{hA} = 8\pi \int \mathrm{d}b \, \mathrm{Im}\left(F_{hA}(s,b)\right) \, b \tag{31}$$

$$\sigma_{\text{el}}^{hA} = 8\pi \int \mathrm{d}b \, |F_{hA}(s,b)|^2 \, b \, . \tag{32}$$

Some caveats about the Glauber model exist. In fact, the theoretical relevance of these effects was already known to Glauber and Matthiae in 1970 but they need to be treated specifically:

- Short-range correlations: In Eq. (29) the nucleons were assumed to be independent. The forces that nucleons experience between each other can be expressed for light nuclei ($A \lesssim 18$) by the harmonic oscillator potential well [GM70]. In case of the—to us interesting—heavy lead nucleus, the nucleus density is described by a Wood-Saxon potential, which needs additional corrections.
- Inelastic screening: This effect does not arise due to approximations made by the Glauber model but from diffraction for which the quarks inside a nucleus can be excited to larger angular momenta. If such state only exists for a short amount of time and falls back to its lowest-energy state before the collision is completed, it "screens" a part of the production cross section.

Both effects are studied with experimental data in this work.

2.3.5 HIGHER-ORDER CORRECTIONS TO THE GLAUBER CALCULATION: SCREENING AND SHORT-RANGE CORRELATIONS

There is experimental evidence that nucleons in the core feel a repulsive force. At low incident electron energies of 4.6 GeV compared to LHC, this is seen in $e^- + {}^{12}C$ collision data for which the ratio of producing a scattered electron and either one or two protons is measured [Shn+07]. The study reveals that roughly 10 % of the electrons interact with a proton pair instead of just one proton. From theory [Fra+93] we expect that at even higher energies this effect is apparent and leads to an increased nuclear thickness function, consequently making the medium more opaque.

The correction is inferred from modifying Eq. (29) to accommodate two-body correlations to leading order [Cio+11]

$$\langle i|i\rangle = \prod_j \rho_1(\vec{r}_j) + \sum_{i<j} \Delta(\vec{r}_i,\vec{r}_j) \prod_{k\neq(i,j)} \rho_1(\vec{r}_k) + \cdots \, , \tag{33}$$

where higher-order terms of $\Delta(\vec{r}_i,\vec{r}_j)$ are not shown. The used two-body contraction is defined as

$$\Delta(\vec{r}_i,\vec{r}_j) = \rho_2(r_j,r_k) - \rho_1(r_j)\rho_1(r_k) \, , \tag{34}$$

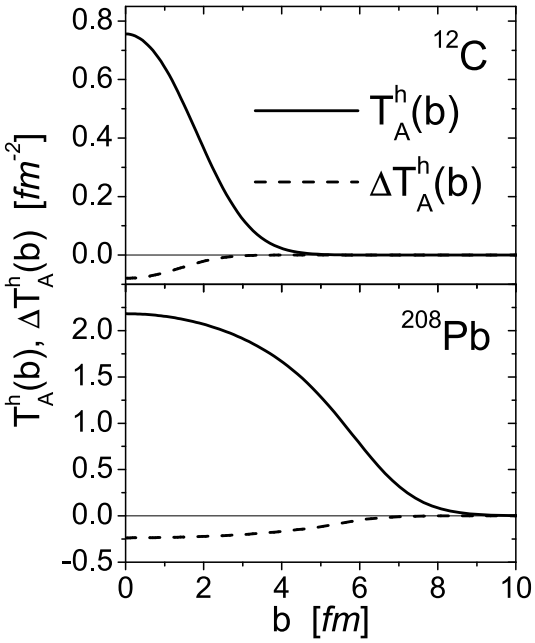


Figure 20: Effect of short range correlations on the thickness profile. The contribution ΔT_A^h is negative and shown by the dashed line. It is subtracted from the nuclear thickness function T_A^h. (from [Alv+10])

where ρ_1 and ρ_2 are the one-body and two-body density matrices. With some calculations, it can be shown that that hadron-nucleus thickness function is modified by the correction $T_A^h(b) \Rightarrow T_A^h(b) = \widetilde{T}_A^h(b) - \Delta T_A^h(b)$. The impact of the modification for carbon and lead nuclei can be seen in Fig. 20. At LHC energies, the authors of [Alv+10; Cio+11] find that for ^{12}C and ^{208}Pb the total cross section is increased by 2.5 % and 0.9 %, respectively. The elastic cross section, on the other hand, feels a stronger effect. It is increased by 5.3 % and 1.5 %, respectively. An experimental measurement does not exist so far.

An effect that introduces a negative correction on the cross section and, therefore, might compensate the effect of short-range correlations is *screening* or *shadowing* as it is sometimes called. It reduces the value of the cross section because there is a non-zero probability for a nucleon to be in a diffractive intermediate state during the collision process. The size of the screening effect can be estimated by describing the diffractive state by low-mass diffraction since no particle production takes place. Low-mass diffraction can be calculated with the Good-Walker approach [GW60]. However, two calculations for screening effects at UHECR energies [EU12; Lip12] use a phenomenological extrapolation of the measured low-mass diffraction cross section, $\sigma_{\text{SD,p-p,low-mass}}$. In [EU12] the extrapolation includes the result of the ALICE Collaboration at 7 TeV. An additional assumption that the $|t|$-dependences of the elastic and diffractive cross section behave the same is used to express the modification by the factor

$$\lambda^2 = \frac{\sigma_{\text{SD,p-p,low-mass}}(s)}{2\sigma_{\text{el,p-p}}(s)}\,. \tag{35}$$

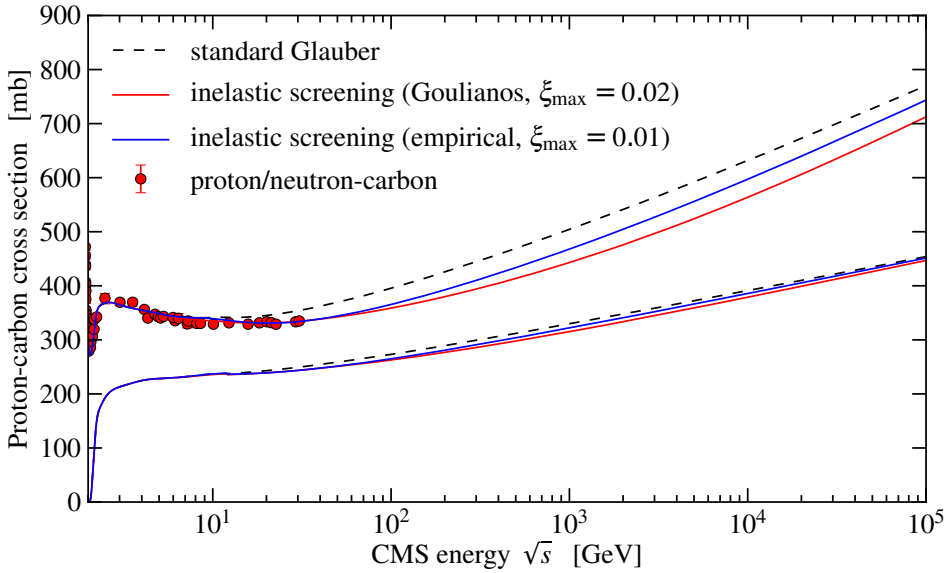


Figure 21: Total (upper lines) and production cross section (lower lines) in proton-carbon collisions. Different theoretical predictions are compared to low-energy measurements. The Glauber cross section (dashed) line lies above the values derived from two different parametrisation of the diffractive cross section (solid lines). (from [EU12] and references therein)

The profile function including screening effects can then be rewritten as

$$F(s,b) = 1 - \frac{1}{2}\prod_{j=1}^{N_{\text{scat}}}\left[1-(1+\lambda)F_{hN}(b,s_j)\right] - \frac{1}{2}\prod_{j=1}^{N_{\text{scat}}}\left[1-(1+\lambda)F_{hN}(b,s_j)\right] . \quad (36)$$

Compared to Eq. (25) the products run over the profiles of the nucleons multiplied by $(1+\lambda)$ or $(1-\lambda)$. The cross section can be calculated by inserting Eq. (25) in Eq. (31). The result is dependent on the parametrisation of the single-diffractive cross section and lies between 5 % and 10 % for proton-carbon at $\sqrt{s_{NN}} = 5.02\,\text{TeV}$ (cf. Fig. 21). For heavier nuclei like lead ions, the diffractive cross section is expected to rise slower than the elastic cross section. The modification factor λ is therefore expected to be lower. Theoretical studies [Alv+13; Cio+11] found the corrections to be of the order of 1 % for proton-lead collisions at LHC energies. Again, the experimental measurements have so far not been performed.

3

EXPERIMENTAL METHODS

In this chapter the tools used in this thesis to improve the understanding of air shower physics are described. Data from collisions at the Large Hadron Collider (LHC) [EB08] at CERN are extremely useful for this. The connection between the collisions at the LHC and the physics discussed in the previous chapter is not immediately evident but is studied with the help of Monte Carlo simulations.

The LHC offers advanced beam monitor tools to determine the luminosity of its high energy particle beams. The luminosity is in important ingredient in many analyses of collider data. Observables of these monitoring tools are studied in this work to calibrate and improve the determination of the luminosity. Even more important are the detectors that record the properties of particle collisions. Of the four big experiments at the LHC, the Compact Muon Solenoid (CMS) experiment is used to improve the description of air shower physics. To achieve an effective comparison between collider and air shower physics, the simulation code of the offline software of the CMS experiment (CMSSW) has to modified extensively. This involves the incorporation of hadronic interaction models used for cosmic ray interactions within the detector simulation of the CMS experiment.

3.1 THE LARGE HADRON COLLIDER (LHC)

The LHC [Brü+04] was built making use of the tunnel of the former Large Electron-Positron collider (LEP), which is close to Geneva at the border of Switzerland and France. With a design energy of $\sqrt{s} = 14\,\text{TeV}$, it can accelerate particles to the currently highest energy achievable by a man-made accelerator. The collider ring has a circumference of 26.7 km and the particle trajectories are bent by 1232 superconducting dipole magnets with a field strength of 8.3 T. The particles revolve with a frequency of 11 245 kHz, and up to 2808 bunches can be filled with particles to maximise the instantaneous luminosity, $\mathcal{L}_{\text{inst}}$. Luminosity, which derives from Latin lumen for light, is the quantity that describes how many events two colliding beams can produce. It is dependent on beam parameters such as the density profile and the number of particles contained within a beam. Its relation to the collisions rate R is given by $\mathcal{L}_{\text{inst}} = R\,\sigma$, where σ is the cross section of the process under study. A peak luminosity of $\mathcal{L}_{\text{inst}} \simeq 10^{34}\,\text{cm}^{-2}s^{-1}$ has been reached. Such a high value is tremendously important in the search for rare processes like Higgs boson production or supersymmetry. When the two antiparallel beams are adjusted to cross and particles collide, the luminosity can be kept for about 30 h whereas the mean beam lifetime is around 100 h when no beam-beam collisions take place. The requirement on the beam lifetime is closely related to the design of the cooled magnets. The main source of losses are beam-gas collisions, which is residual ^{2}H gas (and other elements in smaller amounts) remaining in

the pipes. Even though the gas density is as low as 10^{13} molecules/m^3 in the interaction region and about a factor 100 higher in the beam propagation region, the energy loss is about 30 mW/m for each beam. The helium-cooled magnets that operate at a temperature of 1.9 K should not be heated up too much by the beam-gas collisions as already other processes like synchrotron radiation (0.2 W/m per beam) and image current that is induced by the beam in the walls of the beam pipe (0.2 W/m per beam) heat up the system.

3.1.1 BEAM MONITORING

The bunches in the beam can be focused and defocused by 392 quadrupole magnets, which are distributed over the ring. In this way you can counteract oscillating beam width changes and other beam related effects. They are also used to focus the beams in order to achieve small beam widths at the interactions points and to adjust the crossing angles. Extensive beam monitoring is needed to achieve a stable operation but also to derive quantities that are directly related and needed for analyses concerning physical quantities such as the cross section. One observable is the charge of the bunches which can be monitored via the Coulomb field without directly interfering with the beam. The charge Q is directly related to the number of particles in the bunch via $Q = e\,N_{\text{bunch}}$ with the elementary charge e. With the Direct Current Current Transformers (DCCT [OLT09]) the integrated charge over all bunches of one beam can be measured with a high dynamic range from μA for a single bunch to hundreds of mA for the highest beam intensities with many bunches. Inside a magnetic shielding made of mainly nickel and iron to reduce external effects, three nanocrystalline alloy cores serve as a fluxgate magnetometer. Each ring of the LHC has two DCCT units and another two Fast Beam Current Transformers (FBCT [Bel+07]) units. They measure the relative charges in the bunches with high accuracy. The FBCT devices feature a similar shielding design as the DCCT devices but the sensor must be able to cope with the Radio Frequency (RF) of the beam acceleration. Therefore a high-frequency (low-droop Bergoz transformer covering 1 GHz to 400 Hz) and a low frequency core (Vitroperm 500F going down to 22 Hz) is used. Both device types are installed 160 m to the right of Interaction Point (IP) number 4 and only their combined data gives an accurate description of the beam intensities. Without the low-bandwidth DCCT devices there is no way to calibrate the FBCT devices.

3.1.2 LHC OPERATION

The assembly period of the LHC took place between 1994 and 2008. At the initial planned start-up in 2008 an accident occurred before stable beams were achieved. A faulty connection between two magnets resulted in mechanical deformations and subsequent release of a large amount of helium. As a result repairs and safety checks of the magnets were necessary and only after 14 months in Nov 2009, collisions with lower energy (900 GeV) than originally planned took place. The energy was increased to 2.76 TeV, reaching 7 TeV

in Mar 2010 and finally 8 TeV in 2012. This unintended "energy scan" delivers valuable information about the evolution of processes with energy. After the long shutdown no. 1 (LS1) in 2015, multiple detectors will have received upgrades and the LHC will provide collisions at 13 TeV. This energy is still shy of the design energy by 1 TeV. Notable are several data acquisition periods in 2010 and 2011 when heavy ion collisions of lead-lead (Pb-Pb) could be recorded with $\sqrt{s_{NN}} = 2.76\,\text{TeV}$. In 2013 proton was colliding with lead ions at $\sqrt{s_{NN}} = 5.02\,\text{TeV}$.

Data of the p-Pb acquisition period will be analysed in this thesis and in the following more details about this run will be provided. Table 2 summarises the three run periods of 2013 that were obtained with $\beta^* = 0.8\,\text{m}$ beam optics (see Section 4.1.2.1). During the first period protons were colliding with lead ions and the direction of the particles was reversed for the second period. For the p-Pb direction the proton beam travels clockwise (cf. Fig. 22) around the LHC. The direction is of importance for asymmetric detectors and therefore also for the CASTOR detector at the CMS experiment. Furthermore, the two beams do not have the same energy. The LHC uses dipole magnets with the same field strength to circulate the beams. The mass of the lead ion is $A_{\text{Pb}} \simeq 208 A_{\text{p}}$ whereas its charge is $q_{\text{Pb}} = 82 q_{\text{p}}$. Using the same Larmor radius for both species in Eq. (2), the momentum for the lead nucleus is $|\vec{p}_{\text{Pb,nuc}}| = |\vec{p}_{\text{p}}| q_{\text{p}}/q_{\text{Pb}} = 82|\vec{p}_{\text{p}}|$. For one of the lead nucleon the momentum is $|\vec{p}_{\text{Pb}}| = 82/208|\vec{p}_{\text{p}}|$. The protons are accelerated to $|\vec{p}_{\text{p}}| = 4000\,\text{GeV}/c$ which results in $|\vec{p}_{\text{Pb}}| \simeq 1577\,\text{GeV}/c$ for the nucleons of lead. In the centre-of-mass system the colliding energy is 5.02 TeV and the rapidity boost between the laboratory frame of a detector at rest is the difference of the proton rapidity in the centre-of-mass frame $y_{\text{cm}} = 8.54$ and the rapidity in laboratory frame $y_{\text{lab}} = 9.01$, $\delta y = 0.465$.

Table 2: Data acquisition in 2013. Start date, the number of bunches that were filled, the various LHC fill numbers, and the recorded luminosity by the CMS experiment.

Type	Start date	$\sqrt{s_{NN}}$	Number of bunches	LHC fills	Rec. luminosity
p-Pb	January 20	5.02 TeV	7–296	3474–3510	$20.66\,\text{nb}^{-1}$
Pb-p	February 2	5.02 TeV	72–296	3516–3544	$13.96\,\text{nb}^{-1}$
p-p	February 11	2.76 TeV	46–1278	3555–3564	$5.51\,\text{pb}^{-1}$

The data were accompanied by special beam fills for Van-der-Meer analyses of the luminosity, which will be an essential part of the analysis described in Section 4.1.2. The LHC fill numbers are 3503, 3537, and 3563 for the respective types of beam.

After the p-Pb and Pb-p data acquisition was finished, the LHC delivered proton-proton collisions at $\sqrt{s} = 2.76\,\text{GeV}$ as a reference run to previously recorded heavy ion data. These data allow one to extract nuclear modification factors of measured physics processes, but this will not be analysed in this thesis.

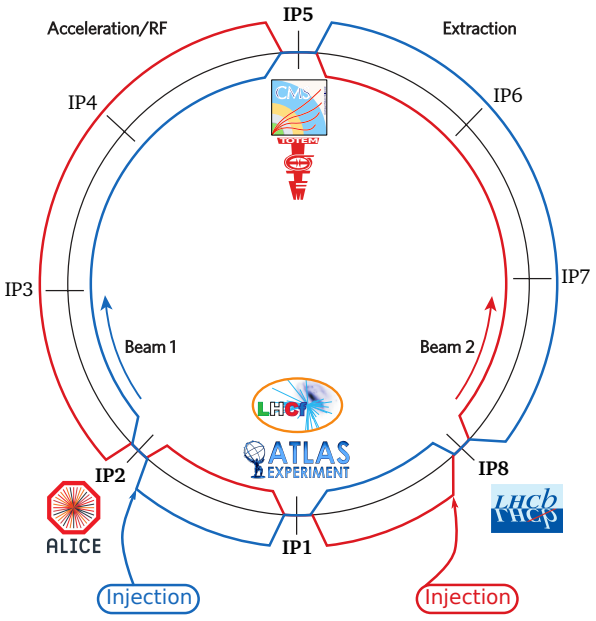


Figure 22: Schematic layout of the LHC. The logos of the experiments have been placed at the corresponding interaction points (IP). Injection from the SPS is indicated by the two arrows at IP2 and IP8.

3.1.3 EXPERIMENTS

The experiments at the LHC are located at four collision points. The four major experiments are in the order of interaction points, A Toroidal LHC Apparatus (ATLAS [Aad+08]) at IP1, A Large Ion Collider Experiment (ALICE [Aam+08]) at IP2, Compact Muon Solenoid (CMS [Cha+08]) at IP5, and Large Hadron Collider beauty (LHCb [Alv+08]) at IP8. A schematic drawing of the layout can be seen in Fig. 22.

The ATLAS and CMS experiments consist of multi-purpose detectors and their physics programme covers a wide range of physic processes. Noteworthy is the discovery of the Higgs boson, which fills an important gap in the description of masses in the standard model.

For air shower physics, the forward phase-space is relevant. We define it via the pseudo-rapidity,

$$\eta = -\ln\left(\tan(\theta/2)\right) \tag{37}$$

which is a measure for the angle between a particle momentum vector and the vector of the beam projectile particle, θ. The quantity η is unitless and differences in pseudorapidity are Lorentz invariant.

Both experiments employ a strong magnet to enable efficient tracking in the central region. Their calorimetry information covers not only the central region but extends into the forward region typically defined as $\eta \gtrsim 3$. ATLAS offers tracking up to $|\eta| < 2.5$ and calorimetry up to $|\eta| < 5$. Both experiments have calorimetric information for neutral particles at a zero degree angle, i. e. along the direction of beam. With the CASTOR calorimeter, the CMS experiment can measure electromagnetic and hadronic energy deposits between $-6.6 < \eta < 5.2$.

At the IP2 and IP5 also detectors of smaller collaborations have been installed that are dedicated to forward physics. The TOTal Elastic and diffractive cross section Measurement (TOTEM) experiment at IP5 aims at measuring the cross section and diffraction with two telescopes (T1 and T2) for tracking information, which cover $3.2 < |\eta| < 6.5$, and beam-pipe-insertion detectors (roman pots), which are brought very close to the beam to pick up elastically scattered protons. The ATLAS experiment incorporates similar roman pot detectors. At IP2 the Large Hadron Collider forward (LHCf) Collaboration installed additional Zero-Degree Calorimeters (ZDC) to measure the energy of neutral particles in two different very high $|\eta|$ regions. Also the CMS experiment has similar detectors. The symbiosis of two experiments at one interaction points makes them the best equipped sites for forward physics studies. The combination of triggers and data of CMS and TOTEM on one hand and the ATLAS and LHCf experiments on the other, is, however, still in development.

The ALICE experiment is also a general-purpose detector with focus on the study of heavy ion collisions addressing many QCD related issues. It consists of an asymmetrical detector design with many subsystems including a tracking detector that has to cope with high multiplicities emerging from the ions. The time of flight can be measured accurately with 6 layers. The central part is also comprised of two electromagnetic detector design with differing granularity. Very good particle identification can be achieved in this way. The forward region is equipped with silicon strip counters and ZDCs on both sides.

The LHCb is built to study CP violation of hadrons consisting of beauty quarks. It has a one-arm design since the decay of B mesons to pions or kaons produce signatures of two particles in the forward region. Therefore only a vertex locator system is on the negative pseudorapidity side and tracking and calorimetry can be found between $1.9 < |\eta| < 4.9$. Due to the fairly large acceptance in the forward region soft QCD processes can be studied with good accuracy.

The details of the detector layouts are covered in the respective design reports referenced above. More detail will only be given on the CMS experiment that is used to analyse proton-lead collisions. Nevertheless, we can sum up the status of the coverage of the different pseudorapidity regions by the experiments as it is also graphically represented in Fig. 23. Here, the regions that can be analysed with different subsystems have been marked with boxes. The central part is covered in the best possible way except at the LHCb experiment due to its asymmetric design to capture hadrons made of beauty and charm quarks. The forward region starting from $|\eta| > 3$ is still covered by calorimetry, sometimes even with particle track information. This continues only up to $|\eta| < 6.5$. For even higher $|\eta|$ only particles counters are available until zero-degree calorimeters with typically $|\eta| > 8.5$ are installed that can only measure the neutral component. This leaves a small but in terms of energy deposit an important gap in the detection. Still the situation at the LHC bears a rather good and global picture of particle compositions with minor drawbacks at very high pseudorapidities.

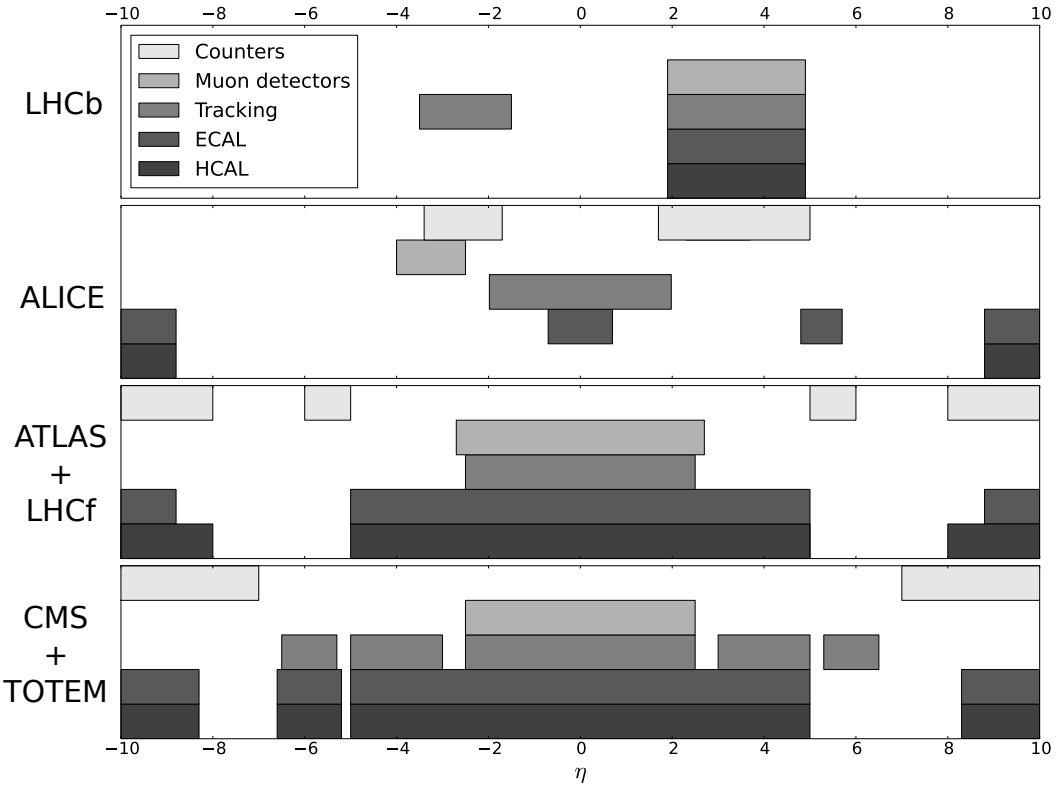


Figure 23: Overview of the coverage by different detection techniques at the four different IP sites. The coverage is divided into Hadronic CALorimetry (HCAL), Electromagnetic CALorimetry (ECAL), tracking, and various counting methods that are based for example on scintillation or pixel detectors.

3.1.4 RELEVANCE TO EXTENSIVE AIR SHOWER PHYSICS

In many aspects, the interaction of the primary cosmic ray is the most important for the development of an air shower. Furthermore, cosmic rays are hadronic particles. At the LHC, collisions of hadrons are currently investigated with protons and heavy ion beams.

The LHC covers a large part of the energy distribution of secondary particles emerging from the first interaction of the primary. In Fig. 24 it is shown that for a proton primary cosmic ray at 10^{18} eV even for the leading particle, which is the most-energetic secondary, the mean of the energy at the second interaction is below the LHC design energy of 14 TeV. Since the inelasticity has only a small dependence on energy, for a 10^{20} eV cosmic ray primary, the distributions of the leading particle energy is shifted towards larger energies by two orders of magnitude. For such primary energy, the LHC only covers leading particles that are produced in subsequent interactions larger than the 5th generation. Extrapolations in energy are unavoidable to describe the leading particles of UHECR at the highest energies.

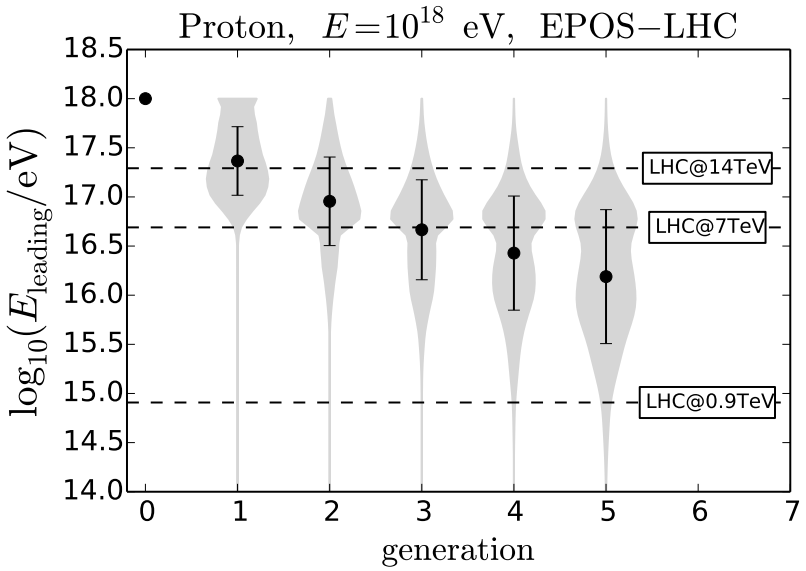


Figure 24: Distribution of the energy of the leading particle after the first interactions. The shaded area in grey shows the probability density function where as the marker and error bars show the mean and standard deviation. The vertical lines show the smallest and the largest energy the LHC can provide and also a typical energy, for which many data have been recorded in 2011. The simulations were run with the air shower Monte Carlo tool CONEX.

In the particle cascade of an air shower, there is also a significant fraction of pions that interact with the atmosphere at LHC energies. Typically, many aspects like diffraction are assumed to be similar between protons, neutrons, and pions in hadronic interaction models. The collisions at the LHC only indirectly constrain all species. Still it is preferable to have pion-proton data to confirm the similarity.

In a study performed with the one-dimensional air shower MC software package CONEX [Ber+07], it can be determined which type of interactions influences the shower observables. The longitudinal profile is separated into individual contributions from sub-showers produced by the secondary particles of the first 100 interactions and of the remaining interactions. One finds that the group of 100 interactions accounts for 50 % of the electrons produced in an air shower but only a small amount of the total muons (Fig. 25). The electrons are mainly produced by decaying neutral pions, whereas charged pions and kaons produce muons when they decay. The decays are governed by an interplay of three ingredients, which depend on the height of the interaction: (1) the density of the atmosphere, (2) the energy distribution of the particles that determines the interaction probability, and (3) the number of particles that are available to possibly decay. Muons can be produced early in the atmosphere. Early produced pions travel for a longer time before they interact than at lower heights even though they have a high γ-factor. The number density of the pions increases at the beginning of the shower. The muon number mainly decreases by decaying into $e^{\pm}$ and two neutrinos with a comparably long decay time of $c\tau = 659\,\text{m}$ and, thus,

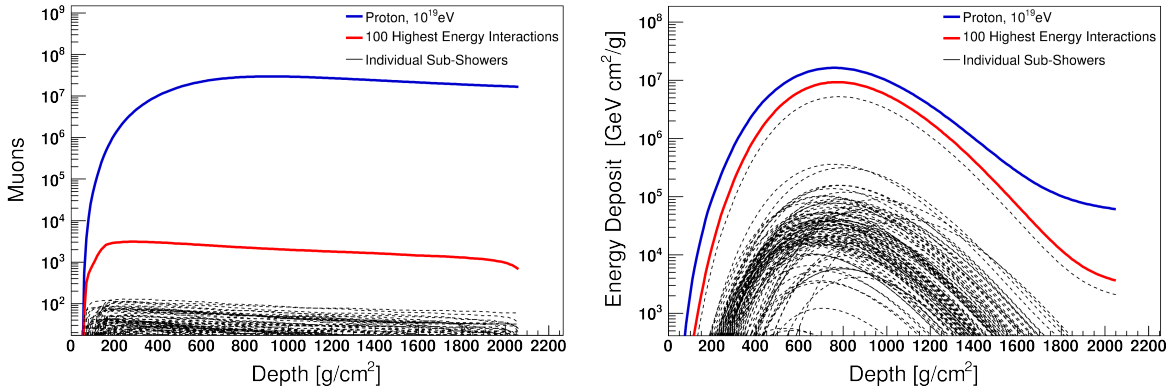


Figure 25: Influence of leading interactions on longitudinal profiles of air showers. One hundred individual subshowers of the highest interactions are drawn in dashed black and added up to draw the red solid line. Left panel: Number of muons per depth. Right panel: Energy deposit per depth. (from [All+13])

the number of muons does not drop as quickly as for electrons. Instead, a plateau in the number of muons over depth is present.

Shower observables are strongly dependent on the most energetic interactions. However, the largest uncertainties arise from the poorly probed forward phase-space. In the centre-of-mass frame, particles are symmetrically distributed over a range in (pseudo)rapidity. Due to stopping (energy losses of secondaries) in hadron collisions most of the particles and all particles of same or heavier masses than the beam particles are produced between rapidities $-y_{\text{beam}}$ and y_{beam}, where y_{beam} is the beam rapidity given by $y = 0.5 \ln \frac{\log(E_{\text{beam}} - p_z)}{\log(E_{\text{beam}} + p_z)}$. For a proton-proton (p-p) collision at $\sqrt{s} = 7\,\text{TeV}\,(14\,\text{TeV})$, the beam rapidity is $y_{\text{beam}} = 8.87\,(9.57)$. The mean distribution of such a collision is shown in the left panel of Fig. 26. Secondary particles are mostly centred at the central part of a detector at LHC with about 5 secondaries per unit of pseudorapidity detectable and decrease towards the forward region with only half of the secondaries being produced at $|\eta| = 6$. Despite the decreasing number, the amount of energy transported into the forward regions increases with $|\eta|$ up to $\simeq$beam rapidity. How far the centre of gravity is away from the beam rapidity is also measured within the scope of this thesis for Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76\,\text{TeV}$. Figure 27 shows this result together with previous measurements that were performed at lower energies. It is found that the centre of gravity is about two units in (pseudo)rapidity smaller than the beam rapidity for LHC energies. This shows that the region at $|\eta| \simeq 6$ is where most energy is deposited. For air showers, this region determines the development of the shower.

Quantitative studies [BUE15] show that the shower development is effectively determined by the few forward-going particles, which carry a large part of the primary energy. The part of the longitudinal and lateral profiles that has its origin in the forward region can be traced through all interactions of an extensive air shower until the cascade reaches the ground. To do this, the secondaries of the first collision are grouped into 6 pseudorapidity regions. Simulations with the 3D air shower simulation program CORSIKA [Hec+98] allow

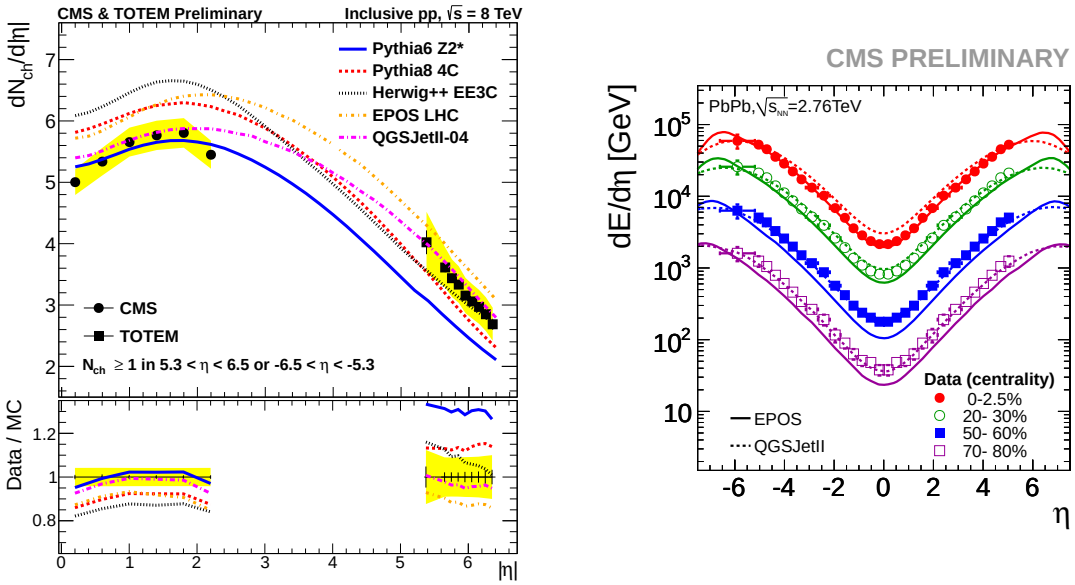


Figure 26: Left panel: Particle production in p-p collisions as a function of pseudorapidity. Using the symmetry of the the x-axis it is shown as an absolute value in the left panel. Right panel: Energy flow in Pb-Pb collisions as a function of pseudorapidity. (from [Cha+14b; CMS12])

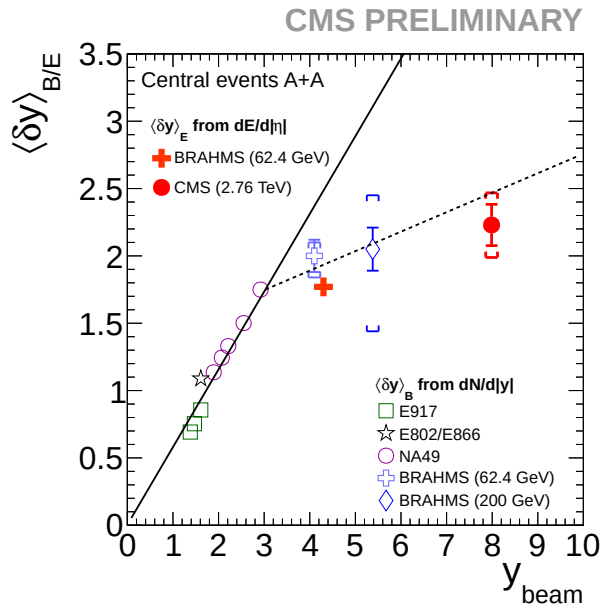


Figure 27: Comparison of measurements of the stopping power in the most central Pb-Pb collisions of various experiments at lower energies with the energy-weighted average pseudorapidity derived with CMS. The solid (dashed) line is a linear fit to AGS and SPS (SPS and RHIC) data. Here $\langle \delta y \rangle_E$ refers to measurements without tracking information in contrast to $\langle \delta y \rangle_B$, which is derived from true rapidity. A linear fit (solid line) without good agreement to high-energy data is shown together with a break in the extrapolation (dashed). The horizontal square brackets for BRAHMS and CMS data indicate the extreme extrapolation scenarios described in the reference. The CMS and BRAHMS data are for the 10 %, NA49 data for the 5 %, and the AGS data for the 4 % most central collision. (from [CMS12] and references therein)

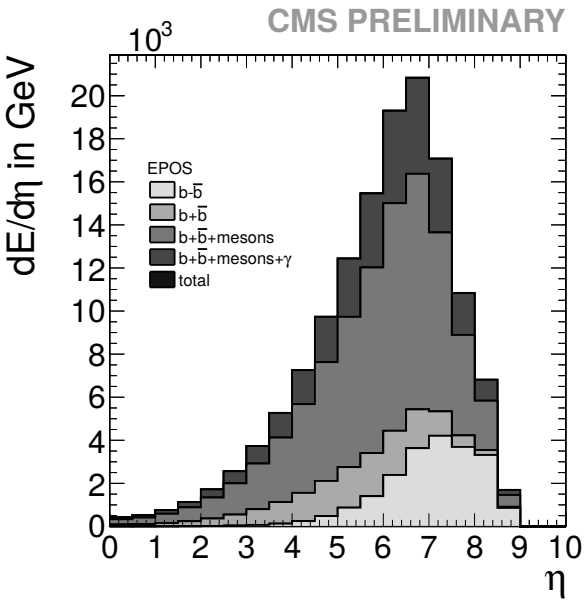
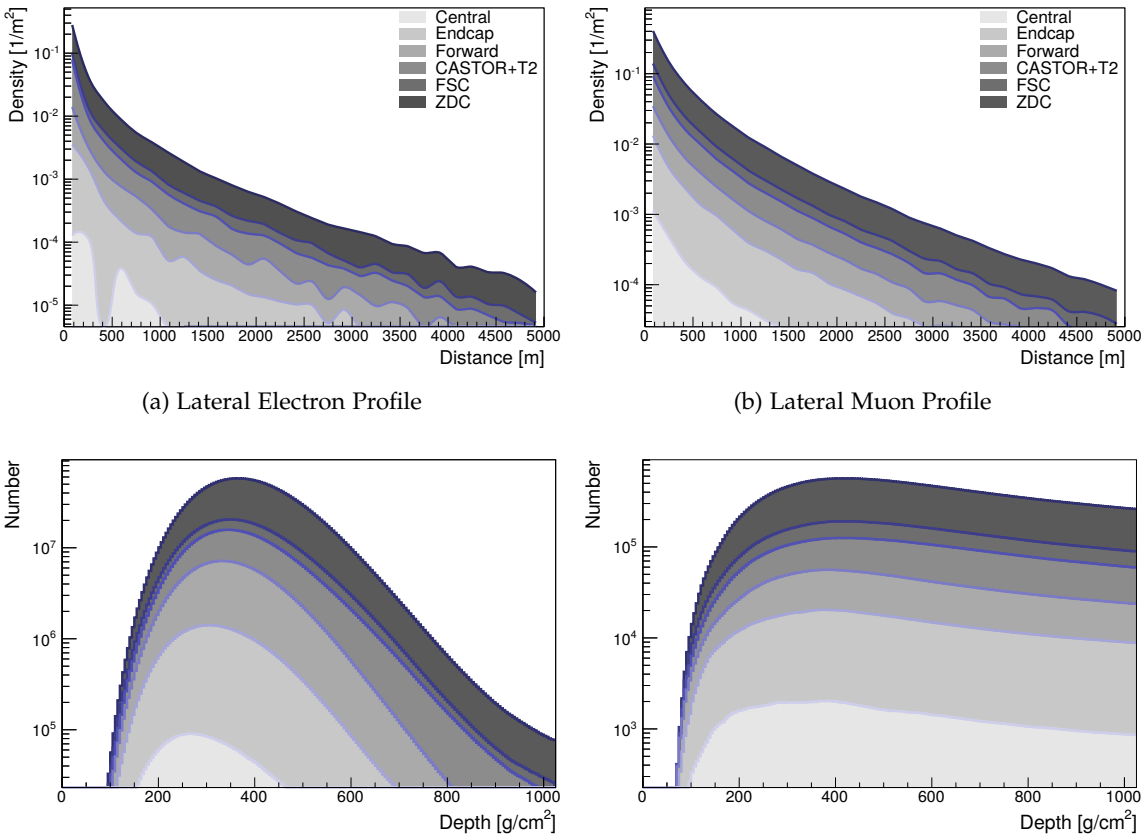


Figure 28: Energy of secondary particles vs pseudorapidity shown for different species of particles as stacked distributions in Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02\,\text{TeV}$. The species are baryons excluding $(b - \bar{b})$ and including $(b + \bar{b})$ anti-baryons, mesons and photons (γ). The results are obtained from the EPOS1.99 MC generator program.

for the evaluation of the lateral electron and muon densities at ground level for each group (Fig. 29 top panels) in addition to the number of electrons and muons as a function of depth (bottom panels). One finds that particles from the central and endcap region are only responsible for less than 10 % of both profiles of the shower. It is noteworthy that most of the particles at all distances and depths originate from the highest pseudorapidity region, $|\eta| > 8$, of the first collision. The pseudorapidity region of the CASTOR detector and the more or less sparsely covered region up to the highest pseudorapidities also contribute significantly. Unfortunately, hadronic interactions models have foremost been tuned to the less important central region so far. It can be concluded that a large part of the observables of extensive air showers is generated by low-energy interactions but their parent particles are nonetheless produced in the forward region of the first interaction.

To improve upon the description of air showers with hadronic interaction models, accurate measurements in the forward region at LHC energies have to be performed. The standard repertoire of inclusive measurements at the LHC in the forward region comprises energy ($\mathrm{d}E/\mathrm{d}\eta$) [Cha+11; CMS12; Aad+12a; Aai+13] and charged particles ($\mathrm{d}N_{\mathrm{ch}}/\mathrm{d}\eta$) [Cha+14b; Aai+14] per pseudorapidity measurements and neutral (π^0, N) particle production cross section [Adr+12; Kaw14] measurements. Also available are jet studies [CMS14c; Aad+12c] and diffractive [Aad+12b; Kha+15; Abe+13] cross section measurements. This list is not exhaustive, however, they all have in common that the underlying process has a large cross section and consequently needs only small integrated luminosities to be studied. It is difficult to single out analyses that bear the largest influence on air showers. Large deviations of predictions of the hadronic interaction models are a good indication for a poorly understood process but not necessarily the only valid indication.



(a) Lateral Electron Profile

(b) Lateral Muon Profile

(c) Longitudinal Electron Profile

(d) Longitudinal Muon Profile

Figure 29: Influence of different pseudorapidity regions grouped at the first interaction (see text) of a shower created by a proton primary with $E = 10^{17}$ eV. Protons can be accelerated to this energy at the LHC. The threshold for the density is 0.3 GeV for muons and hadrons and 3 MeV for electromagnetic particles. (modified from [BUE15])

More comprehensive and accurate measurements in the forward-region are still required to fine-tune the interaction models.

We can conclude that the highest energetic collisions are important and especially the particles produced in the forward region. We also conclude that measuring collisions in p-p interactions gives some understanding but cannot constrain nuclear effects described by GRT. The atmosphere is composed of about 78 % $^{14}N_2$ and 21 % $^{16}O_2$ molecules. Since the two nuclei of each molecule are too far apart to interfere, p-N or p-O are the actual collisions one needs to investigate. Nuclear effects are treated separately from nucleon-nucleon processes but they still introduce a large systematic uncertainty (in case of the total p-p cross section measured by the Pierre Auger Collaboration [Abr+12] the nuclear effects introduce a 16 mb uncertainty compared to 20 mb arising from all other systematic uncertainty). At LHC energies, the available data on ion collisions comes from Pb-Pb and

p-Pb collisions. Proton-oxygen collisions are needed to be studied and are technically feasible at the LHC, according to internal studies [Man13]. The support gas for Pb acceleration at the SPS is oxygen. The injected gas is turned into an oxygen plasma by microwave radiation and ionises the Pb gas. Thus, oxygen is already used in SPS to accelerate ions and is therefore easier to accelerate than nitrogen. Yet, no specific plans exist at the moment to have such collisions at the LHC.

The cosmic ray community is actively involved in shaping the future of the measurements. The experimentalists at the LHC experiments regularly discuss the outstanding problems with theorists and model builders of the community. Another approach is having experimentalists of the cosmic ray community work within the collaborations of the LHC experiments. The *Yellow Report* (a compendium about forward physics at the LHC [LHC15]) discusses all active forward detector upgrades and plans for the 2015 data acquisition. The document supports the requests for specially configured beams favourable to take low-luminosity data useful for forward physics analyses. The authors also argue to consider taking p-O data in the future.

3.2 THE CMS EXPERIMENT

The CMS experiment consists of a multi-purpose detector and is constructed to test standard model physics to the highest precision. Very high luminosities are necessary to study processes with small cross section and the CMS detector is designed to handle very high luminosities of at least $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1} = 10\,\mathrm{nb}^{-1}s^{-1}$ that can occur when particle bunches collide every 25 ns. This requires excellent particle tracking capabilities to identify tenth of overlapping primary vertices (pileup) that can occur in a single collision of bunches that contain many protons.

The heart of the experiment is a superconducting solenoid creating a field strength of 3.8 T and made of windings of stabilised reinforced NbTi conductors. It is 6 m wide and 12.5 m long thereby surrounding the silicon pixel and strip tracker that uses the magnetic field to see curvature of charged particles. The tracker comprises two different designs. The silicon pixels are arranged in three layers between 4.4 cm and 10.2 cm. At larger radii the longer strips follow with 10 layers extending to 1.1 m in radius. With additional detectors at the sides of the inner barrel, i. e. the endcaps the tracker covers $|\eta| < 2.5$. The next detector in the onion-layered design of the CMS experiment is a lead-tungsten scintillating-crystals Electromagnetic CALorimeter (ECAL). The polished crystals have good optical light transmission properties tied with radiation hardness. The endcaps of the ECAL sit directly behind the Preshower, a tungsten-silicon sampling calorimeter, built to measure neutral pions and photons with a good separation power. Including the endcaps the pseudorapidity coverage is $|\eta| < 3$. A brass-scintillator sampling calorimeter with the same pseudorapidity coverage extends to the inner diameter of the solenoid. This Hadron CALorimeter (HCAL) has a larger (angular dependent) thickness of 5.8–10.6 interaction lengths compared to 1.1 interaction lengths of the ECAL. An additional tail catcher is installed outside

of the solenoid. The outermost layer consists of the muon detector that plays a central role in the detector design of CMS. To cope with different luminosities and magnetic field uniformities at higher $|\eta|$, three different techniques are used: drift tubes (DTs), cathode strip chambers (CSCs), and fast restive plate chambers (RPCs). All three detectors combined cover the pseudorapidity region $|\eta| < 2.4$ without any gaps in acceptance.

The overall dimensions of the CMS detector measure 21.6 m in length, and a diameter of 14.6 m with a total weight of 12 500 t. Further technical information can be found in [Cha+08]. Important is the CMS coordinate system, which is oriented such that the x-axis points south to the centre of the LHC ring, the y-axis points vertically upward and the z-axis is in the direction of the beam to the west.

Special forward detectors starting at 11 m away from the interaction point complement the central detectors.

3.2.1 FORWARD INSTRUMENTATION OF THE CMS DETECTOR

Multiple detectors based on different detector designs are in place. The CMS coverage is focused on calorimetric information while the TOTEM experiments provides tracking information in a similar pseudorapidity region. A schematic overview of all forward detectors at the IP5 is shown in Fig. 30. Detector from both the TOTEM experiment and the CMS experiment are included. For the CMS experiment, the hadronic forward calorimeters and the CASTOR calorimeter are best established detectors in the forward region. Both have acquired excellent data in the past data acquisition periods and their software basis is well advanced within the CMS framework, CMSSW [ISS01].

Additional detectors include the Forward Shower Counters (FSC) that comprise 3 stations currently at 59, 85, and 114 m away from the interaction point on each side. The stations can be moved freely so their pseudorapidity coverage is about $7 < |\eta| < 11$. Their physics motivation are diffractive event studies by detecting showers produced when the diffractively created particles collide with the beam pipe. In scintillator panels with an area about $0.06\,\mathrm{m}^2$ showers create scintillation and this is read-out with PMTs. Thus, no information about the energy or the direction of the particle is possible but they can be used for rapidity gap detection. This makes them promising detectors but in this work, the FSC detectors are not be employed due to problems in the 2013 data acquisition and the currently missing detector simulation.

Of all detectors of the CMS experiment, the ZDC detectors are located the farthest distance away from the interaction point on both sides. They are placed inside the casing of the neutral absorber (TAN) that absorbs neutral particles in order to protect other equipment like magnets and pumps from radiation damage. The detectors are 140 m away from the interaction point and just behind the dipole beam separation magnet, where the charged particles of the beams are split into two beam pipes in accordance with their direction of circulation. The neutral particles are, however, not influenced by the magnetic field and will hit the ZDC detectors. The ZDC detectors use tungsten plates as radiators and quartz-

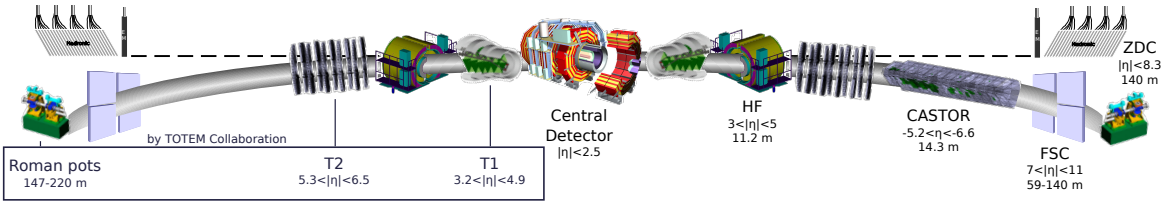


Figure 30: Schematic view of the central and forward detectors at the IP5. The individual distances from the IP along the z-axis are given along with the pseudorapidity regions they cover. Detectors that do not belong to the CMS but to the TOTEM Collaboration are grouped in a box.

fibre as the active medium. They have a electromagnetic section ($19\,X_0$) that is followed by a hadronic section ($5.6\lambda_i$). Both sections cover the pseudorapidity region larger than 8.3 . ZDC detectors play an important for heavy ion physics, for which the centrality of the collisions is correlated to the observed response but also ultra-peripheral collisions can be studied.

The roman pots are part of the detectors of the TOTEM experiment [Ane+08]. Recently, many accomplishments in a common data acquisition between the two collaborations were achieved [Alb+14]. The TOTEM experiment comprises two charged-particle telescopes, T1 and T2, as well as the roman pot detectors. The CSC-based T1 telescope is placed on both sides about 9 m away from IP5. T2 employs Gas Electron-Multiplier (GEM) chambers and they are located at around 13.5 m from IP5. On the negative pseudorapidity side, the telescope is located directly in front of the CASTOR detector of the CMS experiment. Silicon detectors are used in roman pot stations at distances of ± 147 m and ± 220 m from IP5. Similarly to the collimators of the LHC, the pots can be driven into the beam in steps of 5 µm and thus detect elastically or quasi-elastically scattered protons.

In general, particle tracks cannot be attached to identified collision vertices in the forward region due to large distances to the interaction point and the resulting small angles. One consequence is that pileup collisions cannot be resolved and are counterproductive for many processes such as gap detection of diffractive events. Very low instantaneous luminosities are preferable and are often combined with special optics (large β^*, cf. Section 4.1.2.1) to make use of roman pot detectors that measure few µm away from the beam itself.

In this thesis mainly the hadronic forward and the CASTOR calorimeters are employed.

3.2.2 HADRONIC FORWARD CALORIMETER

The Hadronic Forward calorimeter (HF) is built to withstand large amounts of radiation that accumulates when running at high instantaneous luminosities in the forward region. It is estimated that in its life time (after 5×10^2 fb^{-1} of luminosity) the dose in the parts closest to the beam pipe will experience $\simeq$ MGy. About 1/20 of the dose was collected during the 2012 data acquisition and it is constantly monitored by neutron detectors close to HF. The detector is therefore made out of steel absorbers and quartz fibres that combined result

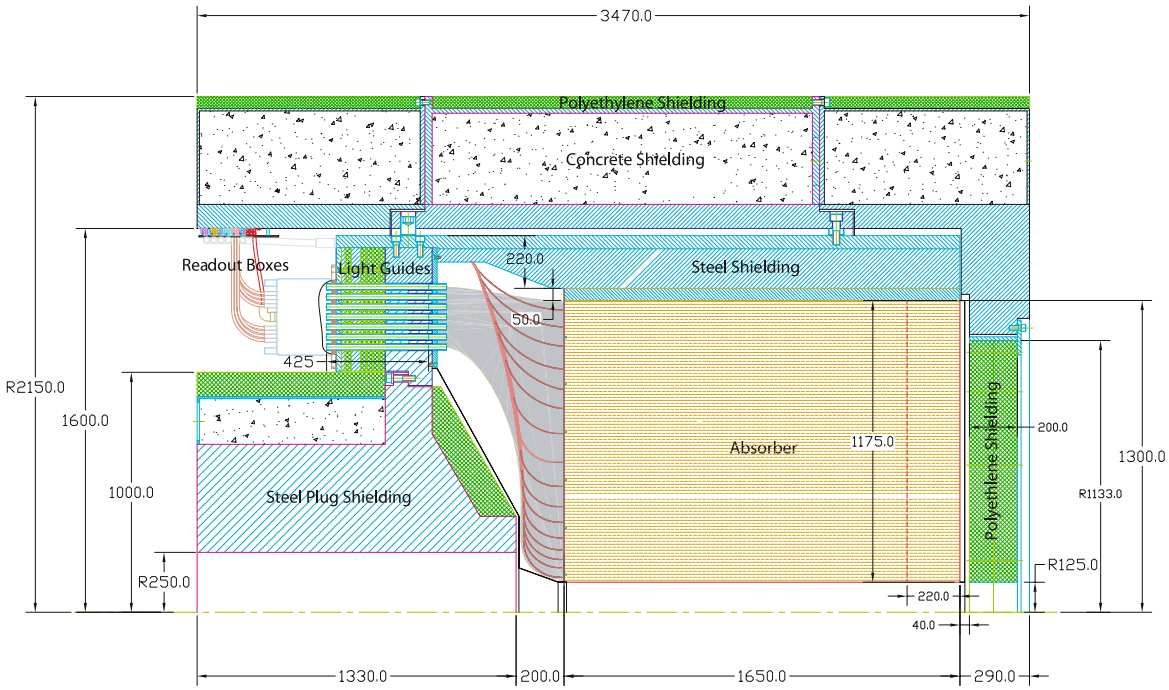


Figure 31: The cross section through the HF calorimeter. The interaction point is 11 m to the right. All dimensions are in mm.(from [Bay+08])

in 108 t of weight. They are mounted on a special table capable of supporting twice its weight and lifted by four transporter platform that can hydraulically lift HF with millimetre precision. The half of the quartz fibres run over the full depth of the absorber (165 cm $\simeq$ 10 λ_I) alternating with shorter fibres that start at a depth of 22 cm from the front of the detector. The fibres reside in grooves that are present in the cylindrical steel absorber surrounding the beam pipe. A cross section of the calorimeter is shown in Fig. 31.

The detection method utilises Cherenkov light produced by charged particles going through the fibres. When the light hits the silica polymer-cladded cores of the fibres at large angles the light is collected and via light guides transported to the PhotoMultiplier Tubes (PMTs) in special readout boxes that are shielded against radiation. The PMTs are of type Hamamatsu R7525HA. Since most particles originate from the interaction point the requirement on the angle is fulfilled most of the time and photon collection efficiency is above 90 % however about 50 % of the light intensity is lost in the light guides. The light is finally converted into an analogue charge signal and digitised by QIEs (charge, integration, and encoding devices) on Application-Specific Integrated Circuit (ASIC) chips. In this way, it is possible to detect electrons as low as 0.2 MeV. Even particle identification is possible to a certain degree because long fibres see the full electron and hadron signal but the short fibres, which skip the first 22 cm, see only 30 % of the signal for an electron but 70 % for hadronic particles compared to long fibres.

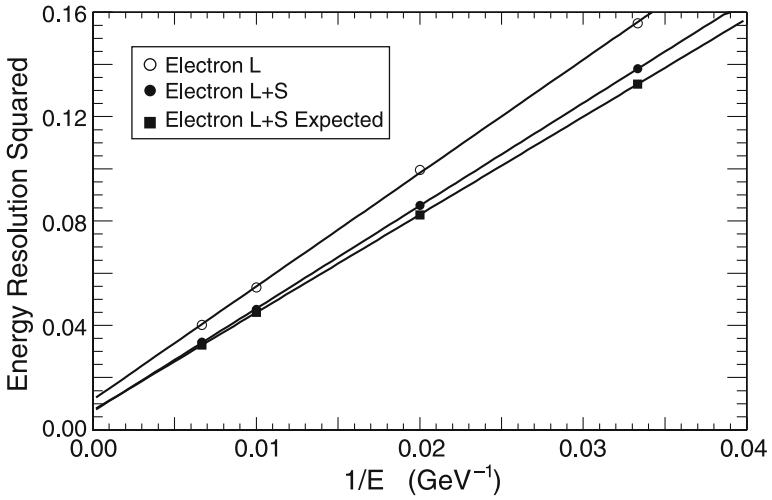


Figure 32: The resolution squared for an electron beam over the inverse of the beam energy. L refers to long fibres, S to short fibres and L+S to the sum of both. (from [Bay+08])

Energy Calibration and Detector Performance

To calibrate the detector, various measurements have been performed and are reported on in [Bay+08]. The most detailed ab initio study comes from beam studies with the CERN SPS beam providing electrons and pions at different energies. With a movable table multiple Φ-segments of the calorimeter are tested for non-uniformity, energy resolution, and non-compensation. Figure 33 shows such a measurement of the detector response for different fibre lengths. The response is normalised to electrons measured by long fibres. The right panel shows the non-compensating behaviour of the detector where 60 % to 80 % of the pion beam energy are visible to the detectors with the long fibres and about 45 % to 65 % in the short fibres. The values depend on the beam energy and increase from 30 GeV to 300 GeV for pions. The response for electrons in the short fibres also increases from 20 % to 35 % between 30 GeV and 160 GeV.

The result of the energy resolution is shown in Fig. 32. A non-zero resolution comes mainly from the photoelectron statistics. It is slightly lower for electrons and in general better if both, long and short fibres, are combined as their sum since a larger part of the signal is seen. Sampling fluctuations and a constant term due to non-uniformities also contribute to the resolution. For a 100 GeV electron it is about 20 % and increases proportionally to $1/\sqrt{E}$ apart from the smaller constant term. Non-uniformities arise due to the geometry of the fibre inlets in the steel. A particle hitting right in between two fibres will see a larger response since R_M is typically larger than the distance between two fibres.

The absolute calibration is determined with a radioactive ^{60}Co source. The source is driven at speeds of about 10 cm/s through thin conduits in the detector, which are specially designed for this purpose. The source has a half time of 5.27 years and emits γ-radiation creating Compton electrons that create Cherenkov photons in the detector. Every 6.55 ms

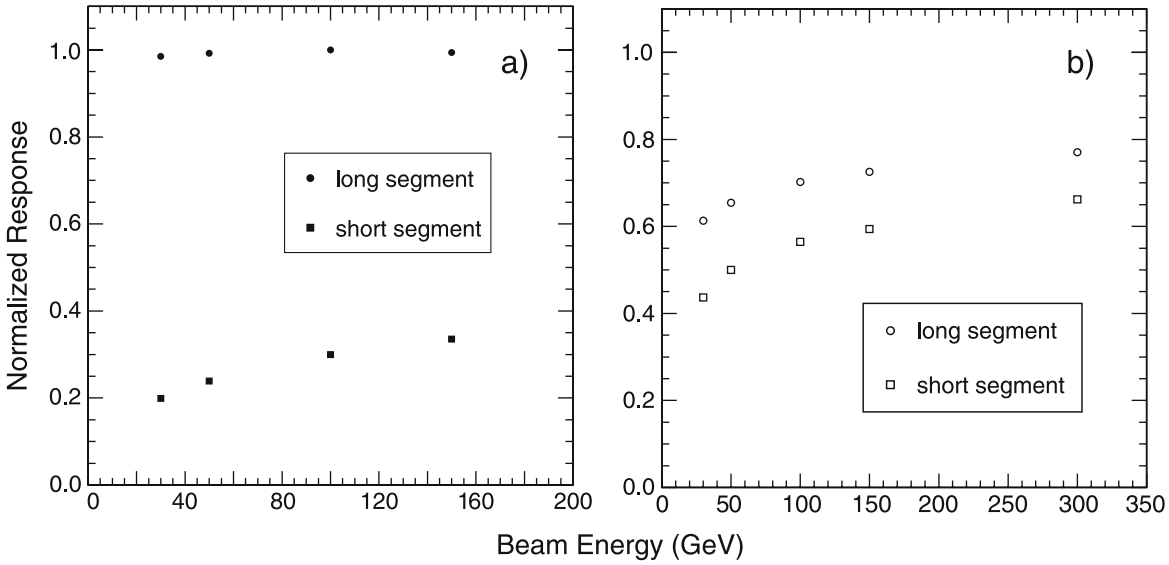


Figure 33: The normalised response for both long and short fibres for a) an electron and b) a pion beam. (from [Bay+08])

the distribution of 4×10^{12} accumulated signals is read out. The most promising method that results in 10 % uncertainty is applied by calculating the mean of the distribution with and without source. The latter is subtracted as background and one obtains $\langle Q_j \rangle$. Now one can use the relation

$$a\epsilon = \frac{C_j}{R_j}\langle Q_j \rangle \,, \tag{38}$$

where a is the activity of the source, ϵ is the energy registered for a source (calculated from previously discussed beam measurements), R is a geometric correction (obtained by a MC method) and C is the correction factor that one wants to determine with the radioactive source method. In Eq. (38) both sides correspond to energy registered in the tower j. The correction factors can now be calculated.

Increasing delivered luminosity leads to a lowered response due to radiation damage. In later analyses [MH11] a method based on the known invariant mass of the Z-boson in a process $Z \to e^- + e^+$ has been employed to calibrate the absolute energy scale. Radiation damage is higher at the towers closer to the beam pipe and, thus, the correction is η-dependent. It is required that one $e^\pm$ with $p_\mathrm{T} > 15\,\mathrm{GeV}$ falls into the acceptance of ECAL and another with $p_\mathrm{T} > 12\,\mathrm{GeV}$ into the HF acceptance region. From the two-body decay of the Z-boson one expects the energy in HF to be

$$E_\mathrm{pred} = \frac{m_\mathrm{z}^2 \cosh(\eta_\mathrm{ECAL})\cosh(\eta_\mathrm{HF})}{2E_\mathrm{ECAL}\cosh(\eta_\mathrm{ECAL}-\eta_\mathrm{HF}) - \cos(\phi_\mathrm{ECAL}-\phi_\mathrm{HF})} \,, \tag{39}$$

with the mass of the Z-boson, m_z, and the direction of momentum of the electron in ECAL and HF detectors, $\eta_\mathrm{ECAL/HF}$ and $\phi_\mathrm{ECAL/HF}$. The predicted energy can be compared to the

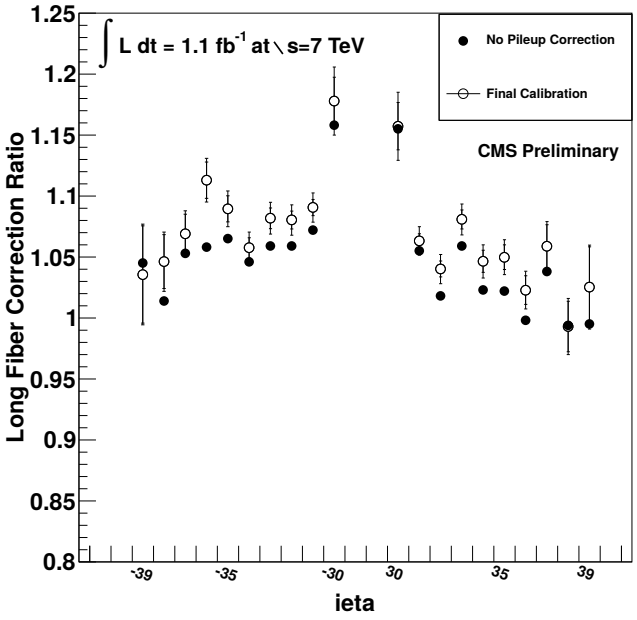


Figure 34: Correction factors for radiation damage as a function of $i\eta$ (see text) for HF measured with the $Z \to e^- + e^+$ method. The inner error bars are statistical uncertainties on the calibration factors while the outer error bars are the combination of statistical and systematic uncertainties. The values shown as "No Pileup Correction" visualise the effect of the vertex correction. The dead material correction is applied in both cases. (from [MH11])

reconstructed energy in HF. Two effects have to be taken into account: (1) Dead material in front of HF attenuates the electron signal. To counteract this effect, the reconstructed energy is weighted by the simulated response, whose value takes dead material into account. (2) Photons from neutral pion decays artificially increase the signal if pileup collisions occur. The response is fitted as a function of pileup collisions and is extrapolated to 0 pileup collisions. The results for the 2011 corrections for the long fibres can be seen in Fig. 34.

Also the short fibres, which see only $\simeq$30 % of the electron response of the long fibres are calibrated with this measurement. During the data acquisition other measurements of the radiation damage correction include LED and laser measurements that can cross-check the light guide and PMT responses.

Furthermore, fake signals of (138 ± 38) GeV are induced by muons that directly hit the PMT glass [Bay+08]. These signals are reduced by requiring a correlation between long and short fibres.

Trigger and Reconstruction

At the QIE boxes the channel responses are corrected for the particle travel time from the interaction point to reach different parts of the detector and different optical path lengths within the design of the ASIC chips. The readout takes place at 40 MHz and the pulse length of the HF detector is about 10 ns in width. Therefore 90 % of the pulse is contained in one time bin. This results in negligible out-of-time pileup, which is important high-luminosity runs at 25 ns bunch spacing. The signal of all HCAL detectors are transferred via a gigabit connection as 32-bit data words to the data acquisition (DAQ).

Multiple triggers are based on the signal of the HF calorimeter. A coincidence of signals above a threshold in both pseudorapidity hemispheres is used to select an interaction as inelastic. A *minimum bias* trigger is a trigger influencing a set of possible interactions as little as possible by its selection. Such a trigger can be based on the central detector but one possibility is to employ the HF calorimeter, which can detect events with diffractive masses down to $\xi = 6\times10^{-5}$. This has a considerable advantage over triggering on signal in the central detectors as they cannot detect diffractive events with large rapidity gaps. The efficiency to select an inelastic event is about 92 % to 95 % depending on the centre-of-mass energy. For higher energies the ratio of diffractive to non-diffractive events is smaller. Another method to trigger minimum bias events is to select a signal in the beam scintillation counters (BSCs) or to combine the HF signal with a reconstructed track in the central tracker. The latter reduces any misidentification due to electromagnetic noise in the calorimeter. Further triggers based on HF include selecting dijet events with a central and a forward jet above a p_{T} threshold or selection of beam gas events.

Once an event is recorded, the digitised responses in each of the 0.175×0.175 ($\Delta\eta \times \Delta\phi$) towers are combined to a tower object as the sum of the responses from long and short fibres during the reconstruction process in CMSSW. The reconstructed object in the software framework is available from the "CaloTower" derived C++ class for all HCAL detectors and offers information on the position of the tower, the status of saturation, the saturation recovered signals, and the energy separated into hadronic and electromagnetic part. The latter can be constructed with weights but for HF, the weights of the long and short fibres are set to 1. The threshold that determines whether to include the signal of fibres is 0.5 (0.85) GeV for long (short) fibres. The electromagnetic part is the difference between long and short fibre signal and the hadronic is twice the short fibre signal; their sum results in the summed signal of the two fibre types.

The signal in the towers are grouped in pseudorapidity rings of the cylindrical detector. For a definition of the rings see Table 3.

Simulation

The detector simulation is based on the GEANT4 [BCG94] framework. For hadronic interactions the physics list "QGSP_FTFP_BERT_EML" is used. The full geometry of the ϕ-segments made of steel and quartz fibres are described in the simulation. Cherenkov photons are produced relative to there distance travelled by charged particles in the quartz fibres with an angle $\theta_c = \arccos(1/(\beta n))$, where n is the refractive index of 1.4–1.5 for the different materials in the fibre. The number of the photons in each step of depth Δx in [m] is calculated by the Frank-Tamm formula,

$$N_\gamma = \Delta x \left[2\pi\alpha \left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right) \sin^2(\theta_c) \right] , \tag{40}$$

with the fine-structure constant α and the wavelengths $\lambda_2 = 280\mathrm{nm}$ and $\lambda_1 = 700\mathrm{nm}$ that define the sensitive region for the optical readout.

Table 3: Properties of the rings in pseudorapidity of the hadron forward calorimeter in numbers of $i\eta$ that start at 0 for the central detector and 41 (-41) corresponds to the ring closest to the beam pipe of the HF detector on the plus(minus)-side. They comprise ϕ-segments with given $\Delta\eta$ and $\Delta\phi$. The number of fibres in each ring is given as well.

$i\eta$	$\langle\eta\rangle$	$\Delta\phi$ [°]	$\Delta\eta$	Number of fibres
±41	±5.04	20	0.300	45
±40	±4.80	20	0.175	42
±39	±4.63	10	0.175	30
±38	±4.45	10	0.175	41
±37	±4.28	10	0.175	59
±36	±4.10	10	0.175	85
±35	±3.93	10	0.175	120
±34	±3.75	10	0.175	171
±33	±3.58	10	0.175	242
±32	±3.40	10	0.175	346
±31	±3.23	10	0.175	491
±30	±3.05	10	0.175	696
±29	±2.91	10	0.111	594

There exist two concepts to enhance the speed of the simulation. One uses tabulated and pre-simulated electron and pion showers, the so-called *ShowerLibrary*, and the other uses a parametrisation of the electron and pion energy deposits within the *GFlash* framework [RK12]. The former method employs the simulation of single particles in discrete energy bins in units of GeV (10, 20, 30, 40, 50, ...), ϕ- and η-bins to obtain information about the number of photons produced by interacting with the quartz fibre material, their momentum and the timing information. When in one step a particle moves into the volume of the HF calorimeter it will be replaced by the shower stored in the library. The average computing time per event increases as the library increases in size. Replacing the particles is one of the disadvantages of the method since information about leakage of the shower is lost.

The GFlash package on the other hand parametrises electron and positron showers in homogeneous calorimeters based on a parametrisation described in [GP93]. The binning of the ShowerLibrary does not exist here, instead three probability density function are used in GFlash: a gamma-function for the longitudinal profile in units of radiation length, assumed ϕ-symmetry, and a distribution of the form

$$p\frac{2rR_C^2}{(r^2+R_C^2)^2}+(1-p)\frac{2rR_T^2}{(r^2+R_T^2)^2} \tag{41}$$

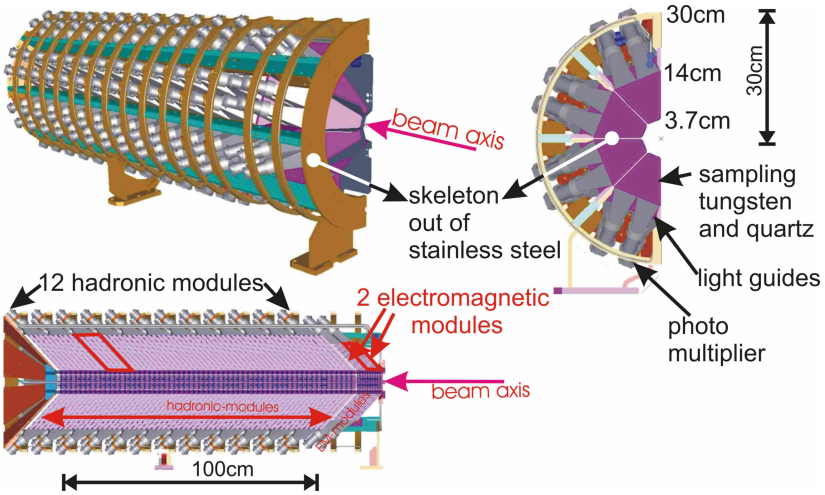


Figure 35: Mechanical structure of the CASTOR detector. A 3D drawing is shown together with slices along and transverse to the beam axis. (from [Got10])

to express the radial influence of the cylindrical detector. Here p is a weight that can be tuned and has a value between 0 and 1; R_C (R_T) give the mean of core and tail distribution in radius r of the detector. The tail is here defined to be larger than $R_M > 1$ in radius. For GFlash a different physics list, "QGSP_BERT_EMV", is employed.

In none of the discussed simulation techniques noise is calculated, which mainly arises from the QIE units.

3.2.3 CASTOR CALORIMETER

The CASTOR detector is a quartz-tungsten sampling calorimeter located at $z = -14.37\,\text{m}$. Its segmentation in ϕ is provided by 16 sectors and together with 14 longitudinal modules the 1.6 m long detector is divided into 224 readout channels. A schematic drawing is shown in Fig. 35. It covers the pseudorapidity region $-6.6 < \eta < -5.2$ without segmentation, where the region is constraint by the radiation shielding at low-$|\eta|$ values and by the beam pipe at high-$|\eta|$ values. In radial coordinates the active volume begins about $r_{\text{min}} \simeq 40\,\text{mm}$ and extends to $r_{\text{max,EM}} \simeq 143\,\text{mm}$ and $r_{\text{max,HAD}} \simeq 176\,\text{mm}$ for the first two and the other modules, respectively. The first two modules represent the electromagnetic (EM) calorimeter. They are half as thick as the 12 hadronic (HAD) modules with $0.39\,\lambda_I$ compared to $0.77\,\lambda_I$. All modules combined amount to material corresponding to about 10 nuclear interaction lengths. The energy deposited by particles when they initiate particle cascades is measured in these modules. The detector is shielded from the magnetic field of 3.8 T from the central CMS solenoid but stray magnetic fields of less than 0.2 T penetrate the shielding and have an impact on the readout of the central modules. Most channels of modules 6 to 7 have more than 90 % decreased signal, however, even for hadronic particles only a limited amount of energy is deposited in the rear modules. In Pb-Pb collisions, it was estimated that including the modules larger than 5 in addition would result in 15 % improvement in the fraction of the observed energy deposit.

The detection technique is based on measuring Cherenkov light produced by charged particles of the shower cascade. To accomplish this, each module is made up of alternating tungsten (W) and fused-quartz-silica SiO_2 (Q) plates that are inclined in a 45° angle. The radiation length of the module is dominated by the dense tungsten (cf. Table 1), whereas the quartz plates are arranged to guide Cherenkov photons via total reflection to light guides attached at the outer edges of the plates. Finally, the fine-mesh PMTs translate the number of photons into a charge signal. Recently, the original PMTs of type Hamamatsu R5505 in the front modules that suffered from significant radiation damage were replaced with new radiation-hard PMTs of type Hamamatsu R7494. The signal is integrated in QIE units and Flash Analog-to-Digital Converters (FADCs) convert the signal in a non-linear way to achieve a dynamic range of 10^4. The information on the energy deposit in the detector is fully integrated into the DAQ of the CMS experiment. This enables to base triggers on energy deposited at high-pseudorapidities, which is valuable to many physics analyses.

The pseudorapidity acceptance covers a region close to the maximum energy deposit of secondaries for most centre-of-mass energies. CASTOR is, thus, sensitive to the fragmentation region. Model predictions are spread widely in this region as can be seen in Fig. 36 for p-p collisions. With a possible separation of signal from electromagnetic particles that dominates in the first two modules and hadronic particles that deposit their energy later, the CASTOR detector is sensitive to study the different energy contribution that mainly comes from charged and neutral pions (combined $\gtrsim 70\,\%$). The physics goals of CASTOR are summarised to be the study of diffractive events [Kha+15], forward jets [CMS14b], the underlying event [Cha+13b], and the forward energy in heavy ion collisions [CMS12]. Measurements related to these are published in the given references. Furthermore, the acronym CASTOR stands for Centauro And STrange Object Research and reflects the possibility to study exotic particle production in high-energy heavy ion collisions [Ang+99].

In these measurements the CASTOR response of each channel is equalised with halo muon data. These muons circulate in a halo around the bunch of the beam when they a created by proton collisions with collimators or beam gas. They traverse the CASTOR detector in the longitudinal direction along a sector. The events, which are recorded by a designated trigger while no collision data is taken by CMS, are characterised by signal above threshold in multiple channels in one sector. The signal is created by minimum ionising muons (Section 2.2.1) that lose a similar amount of momentum in each channel they pass. The charge spectrum recorded in one channel is shown in the left panel of Fig. 37 and contains a contribution above the pedestal from electromagnetic noise. This contribution is normalised over all channels, effectively "intercalibrating" the detector.

The absolute response of the detector was studied with pion and electron beams at the Super Proton Synchrotron (SPS) accelerator at CERN for a prototype of a CASTOR sector [And+10]. The energy resolution was found to be 5 % and 5 % for 100 GeV electrons and pions respectively. The absolute calibration has been proposed but so far not been performed with methods exploiting p_T-balancing of events with two jets or normalising to the invariant mass of the Z-boson from the decays $Z \to e^- + e^+$ as was done for the

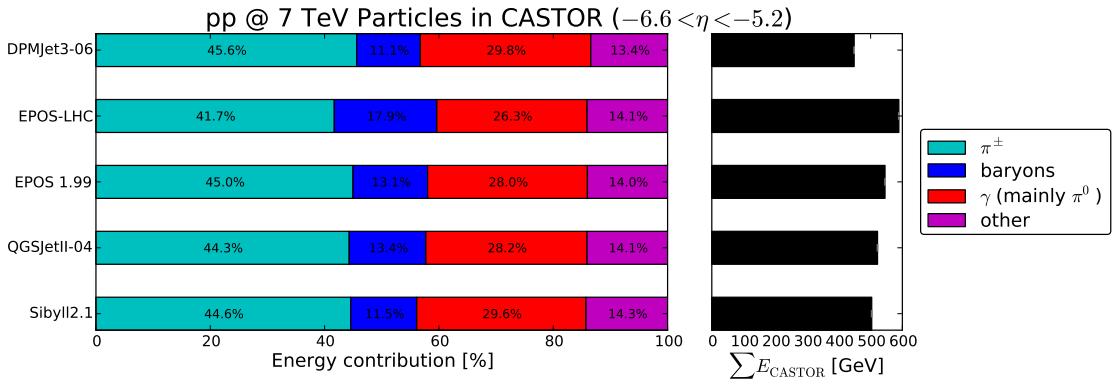


Figure 36: Energy deposit in pseudorapidity region of the CASTOR detector separated as contributions from pions, baryons, photons, and other final-state particles predicted by various hadronic interaction models. The photons in CASTOR are mainly produced in decays of neutral pions. The combined energy deposits are shown on the right in black. The definition of final-state particles is determined by a minimum $c\tau > 1\,\mathrm{cm}$.

HF calorimeters. Both processes require nonexistent large statistics and the energy- and η-resolution of the CASTOR detector introduce a large uncertainty. So far, the adopted solution foresees to normalise the response to the HF energy deposit measured in collisions of a different species of beam particles than the studied data (e. g., p-p collisions when studying Pb-Pb collision). For this purpose, the result from HF data is extrapolated to the CASTOR pseudorapidity region using only the shape of multiple MC generators (Fig. 37 right). The method introduces 22 % uncertainty on the calibration.

Many technical challenges exist for the CASTOR detector due to its position in the high-$|\eta|$ region. The detector is installed in an extremely high radiation environment that requires radiation-robust hardware and good monitoring of the detector response with increasing recorded luminosity. Components like the glass of the PMTs age and lose efficiency. Due to this, recently, the PMTs were partly replaced with new radiation hard ones. Even though fine-meshed PMTs are employed, magnetic fields are problematic and solutions to accommodate both radiation and magnetic fields, like moving the readout devices to a region of smaller radiation cannot be followed up due to space limitations. From the physics motivation, it makes sense to move the detector as close as possible to the beam pipe to cover the physical unexplored region of hadronic collisions. The CASTOR detector is only about 1 cm to 2 cm away from the beam pipe, which requires special monitoring equipment to safely install and uninstall the detector. At the same time, the reverse argument implies that small movements close to the beam pipe translate to large shifts in the measured pseudorapidity region. The position of the detector hidden to outside observers beneath the shielding needs to be known extremely accurately to keep systematic uncertainties low. During ramp-up and ramp-down of the CMS magnet, the detector, which weighs about 2 t and is fixed to a support platform, is displaced on the order of cm.

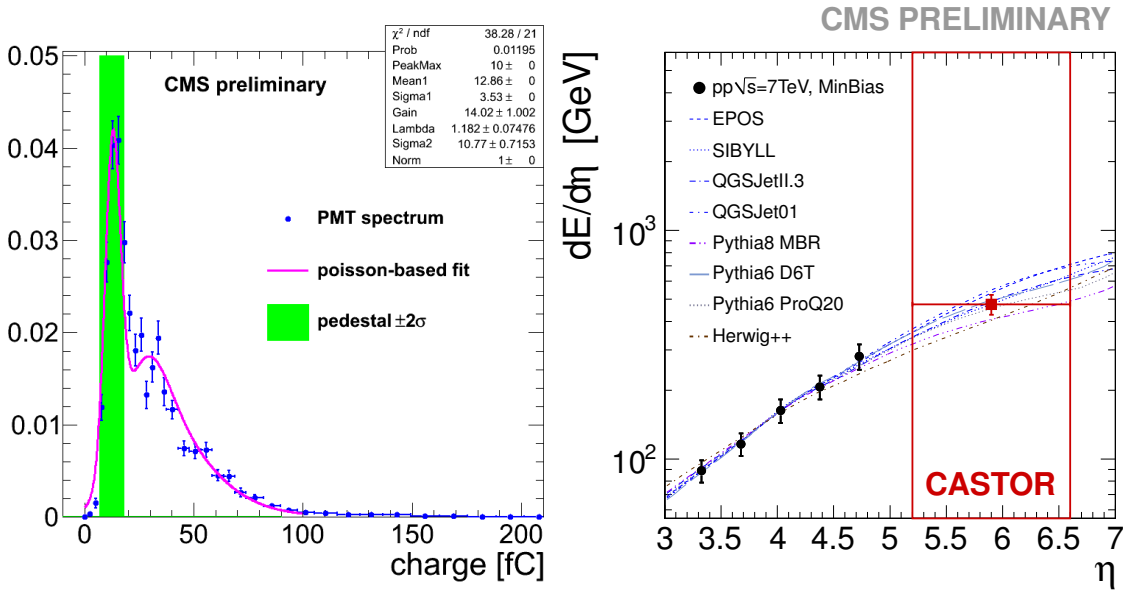


Figure 37: Calibration of the CASTOR detector. Left panel: Distribution of charge integrated over 25 ns during data taking with the halo muon trigger. A Poisson distribution that convolved with a Gaussian distribution describes signal and background is fitted to the spectrum. (from [Kuz12]) Right panel: Extrapolated measurement of energy deposited per pseudorapidity performed with the HF detector. Multiple predictions that are normalised to data in the pseudorapidity region of HF are extrapolated into the region of CASTOR. (from [CMS12])

Within the work of this thesis many improvements in tackling these technical challenges are achieved.

3.2.4 IMPROVEMENTS IN THE CASTOR ALIGNMENT

Mechanically the CASTOR detector consists of two halves. One is covering the positive (negative) hemisphere in x and is called *near* (*far*) side in agreement with the distance to the centre of the LHC. During installation the two halves are individually pushed into a position close to the beam pipe but not so close as to cross a security margin of about 1 cm. The closing procedure can only be observed at the side away from the interaction point (non-IP side) but the clearance at the side closer to the IP (IP-side) is hidden from view. The support platform is shared with the TOTEM T2 telescopes that block the access at the IP-side. Operators check data of attached distance sensors during the installation and can if needed readjust the position. During ramp-up of the CMS magnet the area is inaccessible, the two halves move to a large extent independently from each other, and the monitoring of the procedure relies solely on the distance sensors. The impact of such movement on the measured energy deposit is large. This was estimated for Pb-Pb collision (Fig. 38) to be 16 % for reasonable values of shifts. This displacement also corresponds to changes in the pseudorapidity coverage for individual sectors by up to ~ 0.2 units of pseudorapidity.

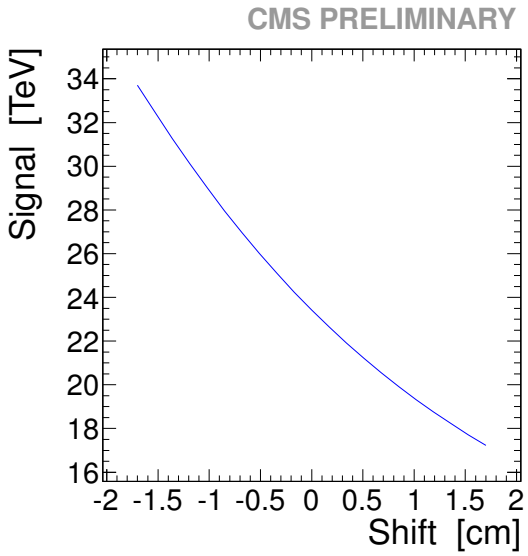


Figure 38: Dependence of total energy observed in CASTOR as a function of radial displacement. Simulated with EPOS for Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02\,\text{TeV}$.

An accurately known position that is also taken into account in the event simulation can reduce the systematic uncertainty greatly.

The position can be derived from multiple position measurements of CASTOR. In 2013, the position of the detector was measured relative to a global reference point known in the coordinate system of CMS by shooting a laser beam at reflecting targets that are attached to the mechanical structure of the detector. The distance to each measurement point is derived from the timing information of the reflected signal. This laser survey was only performed before the magnet ramp-up and after data acquisition when the magnetic field was switched off. To monitor the position during installation, both, contact and non-contact sensors are employed. The former are VTR24F1H infrared (IR) long-distance sensors produced by Excelitas and the latter are potentiometers CLS1300 produced by Active Sensors. The infrared sensors are installed at the IP and non-IP side of CASTOR, while at the non-IP side three sensors are mounted on each half, the IP-side offers only two sensors that are mounted on the support structure for the quartz and tungsten plates at the closest point to the beam pipe. How the distances are measured for the non-IP side is demonstrated in Fig. 39 (left). The sensor has an LED that emits light with infrared wavelength and measures the reflected light with a photo-sensor. Constant light emission from the LED is ensured by separate circuits for the LED and the photosensitive area. The reflective properties are optimised by sticking a matte tape (FARNELL 1394774) to the beam pipe. The calibration of the delivered voltage at a certain distance to the beam pipe is measured and parametrised with a calibration bench where a mock-up of a beam pipe is mounted on [Ant15]. During calibration, the sensor is shifted perpendicular to the beam pipe by a step motor. The calibration curves are fitted with a 8th-order polynomial function and are valid from a few mm to about 60 mm. This calibration works well for a 0° incident angle of

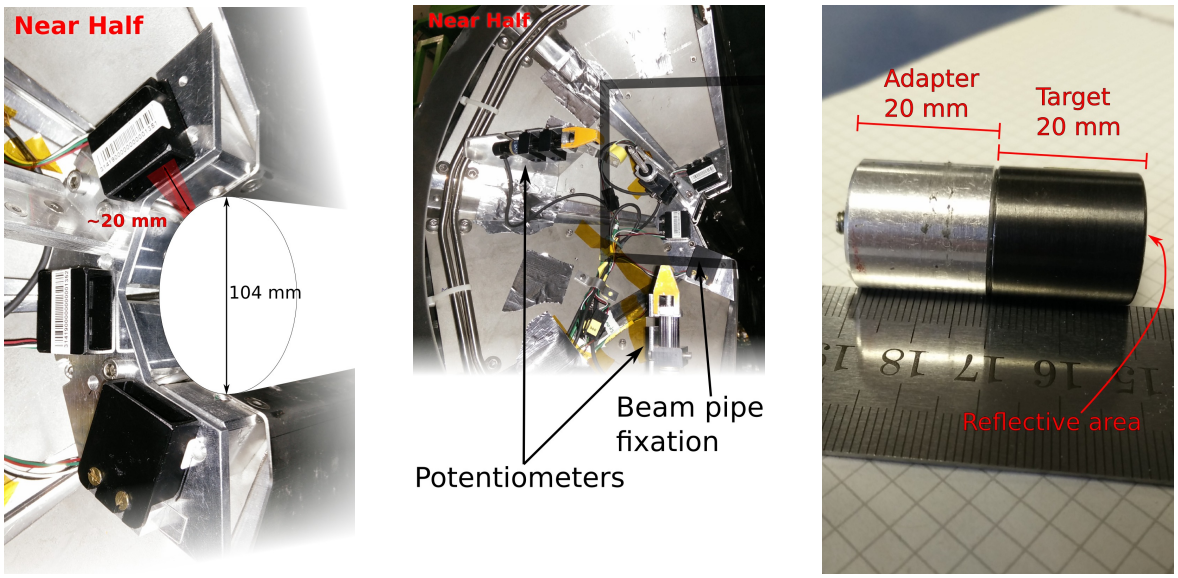


Figure 39: Positional measurement devices for the CASTOR detector. Left panel: Three infrared sensors on the non-IP near side. Middle panel: Potentiometers on the non-IP near side. The beam pipe support is visualised as a grey transparent frame. Right panel: Target and adapters as used in the laser survey measurement in 2013.

the light when it hits the reflective foil, which is sufficient for installation purposes where the knowledge at close distances is relevant where the light hits the tape under small angles. The angular dependence is significant when a detector half is far from its nominal position because the reflected light does not fall back into the photosensitive area, leading to a decreased output voltage compared to the case of negligible incident angle.

In the following, the IR sensors are used outside of their specifications to derive a position for the individual halves of the CASTOR detector on the IP-side. This has never been achieved before and is tremendously important input to estimate the systematic uncertainty on the alignment. The data were taken before and after the ramp-up of the magnets in 2013. An advantage of having both data available is that the shift can be used to correct other measurements. For example, the detector is not accessible to the survey done by laser measurements when the shielding is closed. Infrared sensor measurements are used to account for this. During the run the infrared sensors quickly deteriorate due to radiation exposure and less light or higher distances to the beam pipe are the result. Table 4 summarises the positions and distances to the beam pipe for all sensors in the condition at the beginning of data acquisition. From this positions of the CASTOR halves are induced by moving the each half individually and minimising the measured distance, d_{meas}, with the calculated

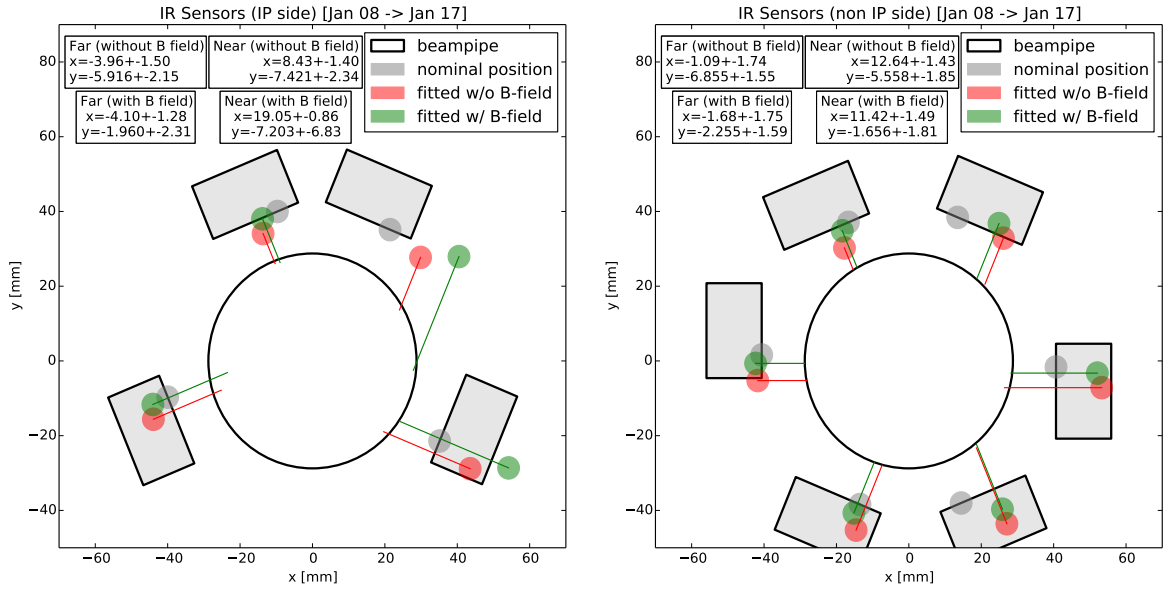


Figure 40: Result of the IR sensor measurement. Left Panel: IP-side. Right Panel: Non-IP side.

distances, $d_{\text{calc}}(x,y)$, of each sensor at x,y of the centre of the half. The procedure is based on minimising χ^2/ndf, where ndf is the number of sensors in each half and

$$\chi^2 = \frac{d_{\text{meas}} - d_{\text{calc}}(x,y)}{\sqrt{\sigma^2_{\text{noise}} + \sigma^2_{\text{sys}}}}, \tag{42}$$

where σ_{noise} is estimated only from the electronic and light noise in the measuring period. Additionally each distance has a $\sigma_{\text{sys}} = 1\,\text{mm}$ systematic uncertainty. The fitted positions are shown in Fig. 40. Notably, the fit of the IP-side positions is constrained by a second sensor type. The potentiometer measure the opening at the top and bottom (for the non-IP side sensors see the middle panel of Fig. 39). The measured opening by the potentiometers is $11.3(20)\,\text{mm}$ and as such about $15\,\text{mm}$ smaller than derived from IR measurement alone. Taking into account both sensor types, the positions during run conditions in (x,y) for both halves are

Near half:	$(x,y)/\text{mm} = 19.05 \pm 0.85,\ -7.20 \pm 6.83$
Far half:	$(x,y)/\text{mm} = -3.96 \pm 1.5,\ -5.92 \pm 2.32$
Opening:	$13.10\,\text{mm}$
χ^2/ndf	7.27

The position non-IP side reveals that the modules are slightly rotated around the y-axis. This tilt that is characterised by a slightly smaller opening on the non-IP side. Due to

Table 4: Properties and distances of infrared sensors as measured for 2013 data after initial ramp-up of the CMS magnet. The photosensitive area is shifted by 6.35 mm relative to the centre of the face of the sensor. If the sensor is "flipped", the shift is clockwise, otherwise it is anticlockwise. The ϕ position of each sensor is given in the CMS coordinate system. The distance is the measured distance with its uncertainty only from the electronic and light noise in the measuring period.

Sensor	ϕ [°]	**Flipped**	**Distance [mm]**	σ_{noise} **[mm]**
IP far top	112.5	no	8.7	0
IP far bottom	-157.5	no	20.2	0.2
IP near top	67.5	no	15.2	0.0
IP near bottom	-22.5	no	25.8	0.8
non-IP far top	112.5	no	6.3	0.1
non-IP far centre	0	yes	13.6	0.1
non-IP far bottom	-157.5	yes	18.7	0.1
non-IP near top	67.5	yes	13.3	0.0
non-IP near centre	180	yes	26.6	0.8
non-IP near bottom	-22.5	no	22	0.1

small influence in actual physical measurements the tilt is not included in the position used for simulation for 2013. The average longitudinal energy deposit in the rear modules is increased by such a tilt. This effect arises from particles that directly hit the—in case of no tilt, hidden—modules.

As a result of this study, an calibration of the angular dependence of each infrared sensor was suggested. As of December 2014, this measurement has been performed by S. Baur and will be implemented for determining the position for the 2015 data acquisition. Under large angles, the distance of an individual sensor can be determined more accurately. However, it cannot be applied in a straightforward way to the present data, since the sensors break in the radiation environment and have to be replaced and recalibrated for each installation. Instead only the average angular calibration can be derived for the 2013 data.

The position of the laser survey measurement is obtained in a similar procedure. On each half are three reflecting targets (Fig. 39 right), one on the IP-side and two at the non-IP side. The three-dimensional position information of each target is included in the fitting procedure to accurately derive the position. The distance of the measured values are minimised again using Eq. (42) but this time, $d_{\text{calc}} = d_{\text{calc}}(x, y, \rho, \theta)$, where θ and ρ denote the rotation angle around the x-axis and y-axis, respectively. The result is shown in Fig. 41. It is, however, only valid for negligible magnetic field. The infrared sensor measurements are used to correct for this by additional shifts of $\Delta x_{\text{near,B}} = 10.62\,\text{mm}$ and $\Delta y_{\text{near,B}} = 0.22\,\text{mm}$ for the near half and $\Delta x_{\text{far,B}} = -0.15\,\text{mm}$ and $\Delta y_{\text{far,B}} = 3.96\,\text{mm}$ for

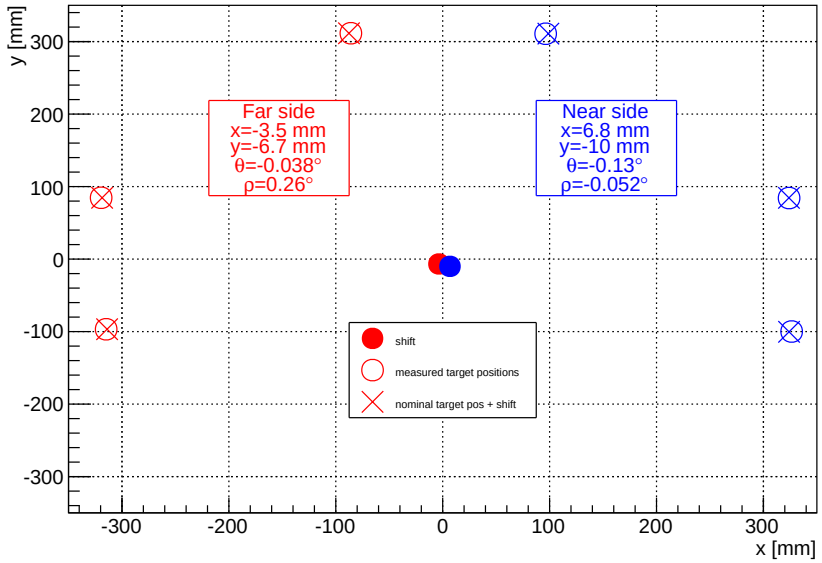


Figure 41: Result of the laser survey measurement projected in two dimensions. The targets at $y \simeq +90\,\text{mm}$ are closer to the IP-side. The measurement is performed for $B = 0\,\text{T}$.

the far half. The systematic uncertainty for this type of measurement is 1 mm but another uncertainty comes from the magnetic field correction. An conservative estimated is, thus, used of 2 mm in x, y, and z. Applying this correction results in the following position valid for conditions with magnetic field:

Near half:	$(x, y)/\text{mm} = 17.43 \pm 2,\ -9.79 \pm 2$
Far half:	$(x, y)/\text{mm} = -3.68 \pm 2,\ -2.76 \pm 2$
Opening:	10.34 mm
χ^2/ndf (near/far)	2.7/1.8

The values agree with the IR measurement on the order of a few mm. The largest disagreement comes from the y-position of the far side and is 3.2 mm.

Finally, the measurements are combined with a study of the position based on charge particle tracks seen in the T2 telescope in front of the CASTOR detector [CMS14a]. TOTEM can be aligned with 2 mm uncertainty by studying a decrease in the track multiplicity that is observed where particles traverse a particular thick piece of the beam pipe at $\eta \simeq -5.5$. In this method, electron tracks are extrapolated to the CASTOR detector and the correlation to the energy deposit they create in the EM sectors is maximised by shifting the individual halves in x- and y-direction.

The final values are derived by taking the average of the IR sensor measurements, the survey measurement corrected for the magnetic field and the position derived with the T2

tracks. The uncertainties of each measurement are added in quadrature. The values given in the following are the ones approved by the CASTOR group. The only change to the here presented analysis is, that the average angular calibration is applied to the infrared sensors, which reduces the systematic uncertainty especially for the y-shift of the near half from 6.83 mm to 2.74 mm. The CASTOR detector alignment is then derived as shifts of the near and far half of the IP side in the x-y-plane.

Detector half	x**[mm]**	y**[mm]**
Near	13.87 ± 2.02	-8.09 ± 3.73
Far	-5.18 ± 2.34	-4.17 ± 3.61

The derived values show that a large separation of the two halves exists of about 19 mm. The detector is also lower in the y-direction than in its nominal position. The y-shift can be partly corrected in upcoming data acquisition periods by adjusting the feet on which the detector is mounted or by placing shims beneath the detector. The separation in x comes mainly from the effect of the magnetic field ramp-up after which no access to the detector is possible. Therefore, the alignment has to be treated with offline methods.

3.2.5 IMPROVEMENTS IN THE CASTOR SIMULATION

The obtained position needs to be described in the detector simulation to reduce the systematic uncertainty from the alignment. The CASTOR simulation [KP06] is implemented in the CMSSW framework to be compatible with the GEANT4 simulation tool. The tool runs within the framework as one of the modular steps of event generation.

The CASTOR detector is a complex geometry that has multiple nested volumes starting with the outermost volume that is filled with air and contains all detector components. The geometry is then separated into the EM and HAD section. The section are separated into octants in ϕ, split into modules in z-direction, and finally the W and Q plates are inserted. The full list of nested volumes is shown in Fig. 43 together with their assigned names. The description does not treat the two detector halves separately but copies each volume describing the sectors (CAHR) 16 times with a ϕ-rotation around the z-axis. Most volumes are tilted by 45°, which is realised with the `G4PolyHedra` class of GEANT4. The polyhedra are separated into 4 z-r-sections that can be understood as looking at a projection of the volume into the z-y plane. This description has the following advantage over tilting the plates by 45° around the x-axis: the plates are flat at the top and bottom and, thus, fit to the surface of the steel frame, which would not be given when they are tilted.

The detector description is verified to work well for describing the energy resolution and the longitudinal response from $e^{\pm}$ hits. It also works reasonably well to describe the longitudinal profile of hadronic particles. The response is modified by an additional correction factor to describe the non-compensation for hadronic particles. The calibration

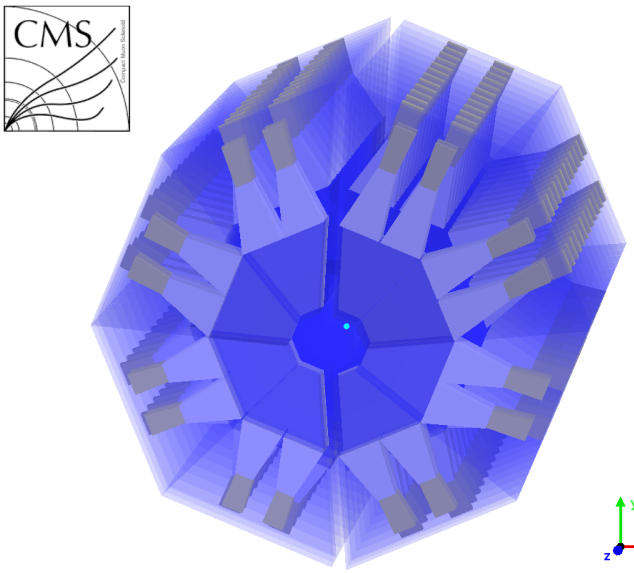


Figure 42: Visualisation of the GEANT4 description of CASTOR detector. The near and far halves are placed at the determined position from Section 3.2.4. An opening along the x-axis between the two halves is visible. The image is produced with the Fireworks event display.

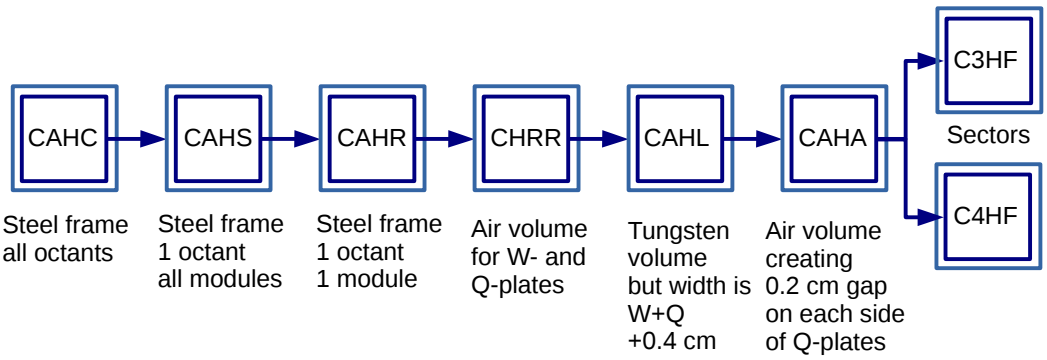


Figure 43: Volume names of the CASTOR geometry. Here the nested volumes that are contained in the mother volume "CAST" are shown. Until the last step, the volumes still describe octants and only then they are divided into volumes of angular size $2\pi/16$. The CAHA volume is inserted to simulate a Tyvek layer that is wrapped around the Q-plates in order to increase uniformity by light scattering. The difference in the response for Tyvek and air is negligible and is not simulated.

at the SPS accelerator to the ratio between the measured energy of pion and electron beams, E_π / E_{e^-}, is obtained by downscaling the hadronic signal by a factor of 0.85 in simulations.

A problem of the geometric description is that tilts and the separation of the detector halves cannot be described. To improve this, the CAST volume is split in two entities and the possibility to individually shift and rotate the halves is added. The procedure is straightforward with only few difficulties of taking symmetries into account when applying the ϕ-rotations to the CAHR volumes and assigning the correct identifiers in the framework for each sector. Photoelectrons created in the GEANT4 simulation need to produce charge in the correct PMT. This is handled in other parts of the framework. Previous approaches to estimate the systematic uncertainties from an opening between the halves required a twofold simulation. Once it was performed where the entire detector was shifted so that the near half was at the right position and once where the right half is correctly positioned. The signals were then combined. With the halves separated in simulation, this procedure is faster, but also effects at the sector borders close to the opening are simulated properly. When the two halves are separated, particles can produce hits directly in the rear module. With these improvements, the position determined from the distance sensor measurements can be included in the simulation. For now, accurate data for tilts are not present but with the proposed improvements on the sensor measurements future alignment analyses are likely to determine the rotation of the halves. The simulated geometry with the position derived by the sensor measurements is shown in Fig. 42.

Along with the geometry description that treats the two halves separately, many minor mistakes are corrected. Previously present overlapping volumes in are, thus, completely removed but the simulated response is only negligibly affected. However, further studies following up on the improvements to the geometry description are conducted within the work of obtaining a bachelor degree by M. Eliachevitch [Eli15]. The corners of the W- and Q-plates in an octant have cut-off corners at high-$|\eta|$, i. e. close to the beam pipe, to allow their full insertion into the steel frame. At the bottom of the octant the length is reduced by about 20 % resulting in an pseudorapidity-dependent decrease of the response compared to the original geometry description. The effect of the more accurate description of the corners becomes evident in Fig. 44. The electron and the pion responses are decreased by more than 10 % at $|\eta| > 6.5$.

The longitudinal profile is also affected (Fig. 45), for which the largest decrease of the response is visible in the first 4 modules. In later modules the broadening of the shower cascade smears out the local dependence. For high centre-of-mass energies the pseudorapidity acceptance of the CASTOR detectors lies on the rising slope of the $\mathrm{d}E/\mathrm{d}\eta$ distribution, therefore the high-$|\eta|$ region contributes more to the absolute deposited energy in the CASTOR detector than the low-$|\eta|$ region. Moreover, studies of the response with the SPS beam were performed with a centrally radiated area of the W-plates of about 3 mm in radius [And+10]. This is indicated by the pseudorapidity region “TB comparison” in Fig. 44 where the simulated response shows no influence by the more accurate geometry description. Consequently, this effect is hidden from measurements with test beam as previously performed.

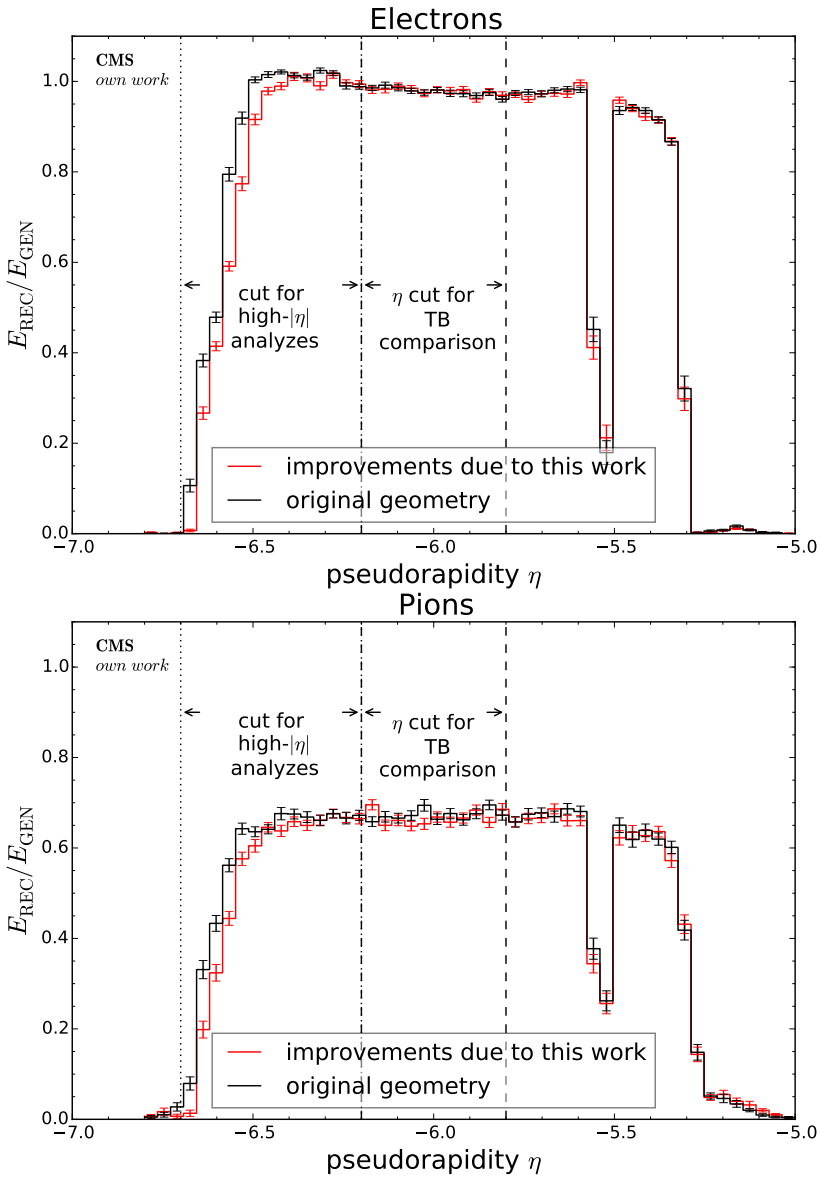


Figure 44: Simulation of the response of the detector for an electron (top) and pion (bottom) beam. The response is divided by the generated particle energy of 50 GeV and is shown for the original geometry description as well as the improved version due to work done in the context of [Eli15]. Two pseudorapidity regions are shown, one for high absolute values and one to which the SPS beam studies are sensitive.

The total change is response is checked in the simulation of minimum bias collisions with the PYTHIA8 generator (Fig. 46). At $\sqrt{s} = 13\,\text{TeV}$, the response is 7% lower compared to the original geometry. Due to a smaller slope in the $\mathrm{d}E/\mathrm{d}\eta$ distribution, the effect is less significant at smaller centre-of-mass energies.

The studies of the alignment of the CASTOR detector, ultimately lead to three improvements: (1) The positions of the two CASTOR halves during the 2013 data acquisition are accurately determined with an uncertainty of 3 mm. (2) The geometry description in simulation is updated to allow the implementation of such shifts and even rotations. (3) An improved description has been found that introduces changes in the response of up to 7 %.

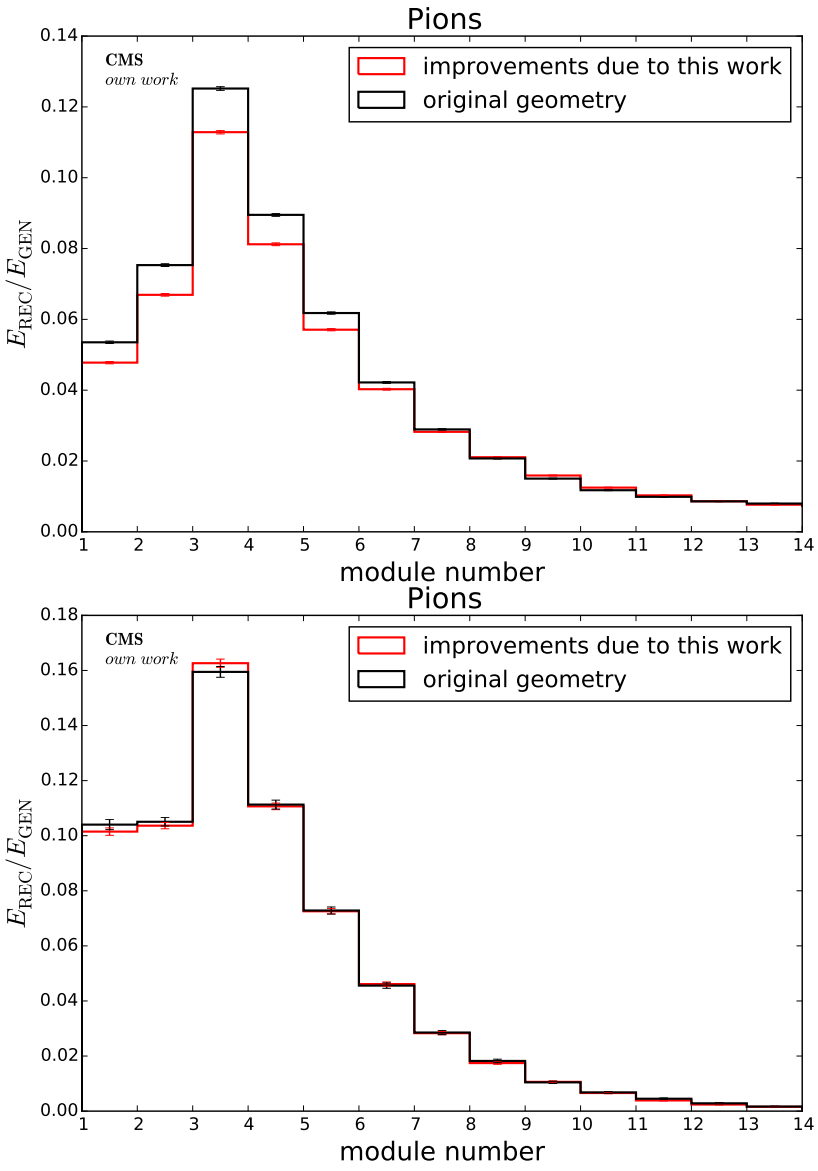


Figure 45: Energy response per module for the same simulation conditions as in Fig. 44. The response is shown for the region closer to the beam pipe (top) and the region for which the SPS beam studies are sensitive to (bottom). The improved geometry description of [Eli15] is shown in red and compared to the original geometry shown in black.

The findings of (2) and (3) are already propagated to the CMS framework. In this way, simulations of new events will take the updated position into account.

3.3 HADRONIC INTERACTION MODELS DESIGNED FOR AIR SHOWER PHYSICS

Phenomenological models are needed to describe hadronic interactions in air showers created by cosmic rays (Section 2.3). Macroscopic quantities measured in experiments are subject to statistical fluctuations. In air showers this is, e. g., expressed in the standard devi-

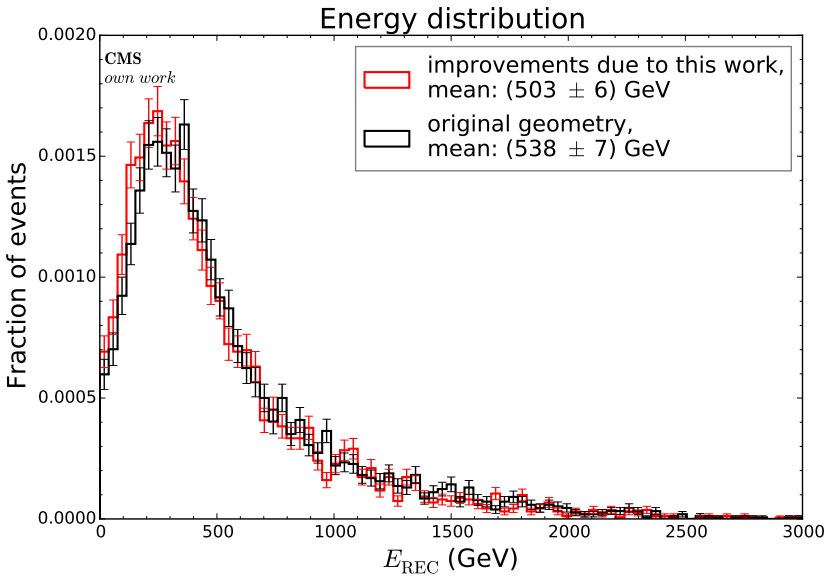


Figure 46: Distribution of reconstructed energy in the CASTOR detector. The minimum bias collisions are simulated with the PYTHIA8 generator (CUEP8M1) at $\sqrt{s} = 13\,\mathrm{TeV}$. The improved geometry description of [Eli15] is shown in red and compared to the original geometry shown in black.

ation of X_{max} that arises from shower-to-shower fluctuations. In Monte Carlo programs, the phenomenological models are implemented to simulate individual events. These programs employ random number generators together with probability density functions from theory, to obtain parameters of interaction, e. g., how many partons interact or how many particles are produced. In applications for collider experiments and for extensive air shower experiments, usually more than a thousand events have to be simulated to describe the intrinsic fluctuations.

The phenomenological models have $\mathcal{O}(100)$ free parameters that are tuned to measurements mostly from collider and fixed-target experiments. Collider data are available at higher energies compared to fixed-target measurements but for the latter it is easier to instrument the forward region of the phase-space. Even with tuning of the parameters not all data can be described perfectly. On the other hand only data that influence the air shower development are really necessary. Some models like EPOS strive to give a holistic description where others, e. g., SIBYLL, concentrate on faster and light-weight implementations of the necessary physical processes in air showers. It is common practice to release updated versions of the code every few years in order to implement latest advancements in theoretical understanding and to tune the parameters to newly available measurements. Another intriguing aspect of the tuning is that for each model only very few people know about the sensitivity and usage of the parameters as they are specific to each model. Furthermore, it involves a lot of manual adjustment and eye-by-eye comparisons. We propose a more automated and structured way of doing this in Section 4.3.1.

In the following section an overview of the implementation of MC-based hadronic interaction models is given for: DPMJET, EPOS, QGSJET, and SIBYLL. The main focus are the models used in subsequent studies reported on in this work. The models are complex

with about 5000 for SIBYLL and over 80 000 lines of codes in case of EPOS. All models are based on GRT (Section 2.3.3). Therefore, in the overview, only the most relevant differences between them are presented.

3.3.0.1 *DPMJET*

DPMJET [Bop+08] adds nuclear treatment to nucleon-nucleon interaction calculated within the PHOJET framework. To calculate hadron-hadron collisions, PHOJET uses a pomeron that has a hard and soft part and can describe the results from perturbative QCD for hard collisions, here defined with a cut-off $p_T > 3\,\text{GeV}$. PHOJET makes no distinct differentiation between diffractive and non-diffractive events while it employs such a pomeron. This combined approach is extended by assuming similar Feynman graphs for the diffractive and non-diffractive parts that can be of higher orders in the case of diffraction (e. g. triple pomeron graphs). Non-diffractive collisions are described by multiple-parton exchanges between target and projectile. PHOJET specifically allows for this to happen in diffractive collisions as well. Shadowing for hadron-hadron collisions implicitly comes from unitarising the sum of all enhanced graphs and the standard one-pomeron-exchange graph. The latter therefore loses relevance compared to the case where no diffractive graphs exist and is therefore shadowed by them. Even photon-hadron interactions via the Dual Parton Model and perturbative QCD to first order are included [ER96].

DPMJET also incorporates the Dual Parton Model but allows for nuclear collisions via the Gribov-Regge and Glauber theories. With the enhanced graphs, it has a good description of diffractive processes but within this work it was discovered that no coherent scattering is included for nuclear interactions. This can constrain the ability to simulate diffractive events significantly since the contribution of coherent scattering can be as large as 50 %. Apart from this, realistic nuclear densities and radii are used for light nuclei and Woods-Saxon densities for heavier nuclei. From its predecessor DTUNUC (where DTUJET played the role of PHOJET) it inherited the description of the Fermi motion and Pauli's exclusion principle for nuclei collisions, which distinguishes it from the other nuclear MC models. For both PHOJET and DPMJET, hadronisation is strongly intertwined with the PYTHIA framework and uses its Lund string fragmentation [Sjo88].

3.3.0.2 *EPOS*

EPOS [WLP06] is based on parton-level GRT and its acronym stands for

- Energy conserving quantum mechanical multiple scattering, based on:
- Partons, and parton ladders
- Off-shell remnants
- Splitting of parton ladders.

The acronym also highlights the features that separate this model from the others. From the list it is clear that energy sharing is explicitly included, connecting each of the multiple scatterings in an interaction and not considering them as independent. For nuclei, we already saw earlier that one of the fundamental flaws of a pure Glauber calculation is to treat each interaction elastically and therefore independently. EPOS resolves this problem already on the level of partons. This was introduced in the predating NEXUS code, which combined the QGSJET model for semi-hard and the VENUS model for soft collisions. In 2009 EPOS was released as version 1.99, which was not yet tuned to any LHC data, but received an update in 2014 referred to as EPOS-LHC. Apart from those versions, EPOS2 exists (not publicly available), which focuses on the hydrodynamic flow description in 3D but cannot be used to produce high statistics due to the long execution time per event. We concentrate on describing EPOS-LHC in the following.

For collisions of two heavy ions, EPOS-LHC implements a simplified version of EPOS2, where a full quark-gluon plasma with all stages until the freeze-out is simulated (Fig. 47). In spatial coordinates there are defined areas with high numbers of strings and therefore also energy densities that build up shortly after the interaction. In the program they are treated as a fluid core. The core is evaluated at a point in time, τ_0, before the hadronisation begins and strings are still present. The result only slightly depends on the choice of τ_0 and this dependence is mainly visible in the transition region of gaining high enough densities to form a core. In this core, particles are formed to clusters and can hadronise as a whole. Even in p-p collisions, a core can be produced if the number of charged particles exceeds $N_{\mathrm{ch}} \gtrsim 25$. With a reduced number of particles produced at very central events both p-p and heavy ion data can be successfully described. For air showers, the fluid core does not have any influence on the shower distribution and is by default switched off to save time.

Apart from the fluid core, standard parton ladders that follow the DGLAP evolution describe scattering processes. Hadronisation is implemented similarly to the Lund algorithm. Diffractive masses follow a $\mathrm{d}\sigma/\mathrm{d}M \propto 1/M^{2\alpha_{\mathrm{diff}}}$ distribution, with $\alpha_{\mathrm{diff}} = 1.05$, which is similar to the other models. In the previous version (1.99), α_{diff} is set to 0.4 generating a flatter distribution. The updated value leads to a significant change in the muon production depth in air showers, which is in tension with the measured result. For other observables the updated version that includes tuning to the LHC data [Pie+13] is preferred and used in this work.

3.3.0.3 *QGSJet*

The QGSJET framework [Ost11] is based on the Quark-Gluon String model and the version QGSJETII-04 extends its preceding versions by including corrections for pomeron loops to all orders [Ost10] as will be discussed in the following.

The eikonal of the pomeron needed for the cross section calculation is made up of an eikonal of a soft and a semi-hard part of the pomeron:

$$\chi^{P}(s,b) = \chi^{P_{\mathrm{soft}}}(s,b) + \chi^{P_{\mathrm{semi\text{-}hard}}}(s,b)\,. \tag{43}$$

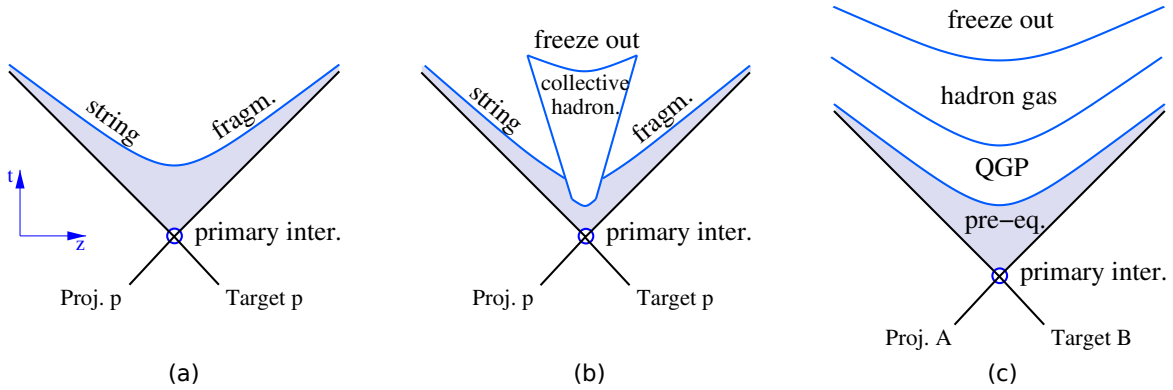


Figure 47: Simplified pictures of hadronisation in standard high-energy interaction models: EPOS1.99 (a), EPOS-LHC (b), and EPOS2 (c). All show the space-time evolution in time (t) and z-coordinate where the beam particle with speeds close to c in the z and $-z$ direction lie on the black lines. String fragmentation for produced partons happens once the freeze-out temperature is reached. The drawing in the middle shows that collective hadronisation in clusters can happen on local scales that have high energy densities. The right-hand side shows the full complexity involving a Quark-Gluon Plasma (QGP) assumed to occur in heavy ion collisions. (modified from [Pie+13])

The semi-hard part is modelled as emission of two soft pomerons from each projectile and target and in between the two pomerons a QCD ladder made from sea quarks and gluons forms. For the ladder DGLAP ordering is used. The term of the two-component pomeron can be used in a non-trivial way to recursively enter equations for higher-order loop corrections of the pomeron interactions. In this way shadowing and saturation effects are included implicitly and self-consistently in QGSJET. This is an advantage over other models that employ phenomenological methods to treat these effects.

Nevertheless, high mass diffraction is implemented in a purely phenomenological way containing a single elementary production process. The mass distribution follows as usual a $1/M^2$ distribution. Low mass diffraction is implemented via Good-Walker-like multi-component treatment [GW60].

3.3.0.4 *Sibyll*

SIBYLL2.1 [Ahn+09] concentrates on describing basic elements of various theories concerning hadronic interaction. This includes string fragmentation, minijet production, multiple scattering, and diffraction for hadron collisions and Glauber theory for collisions of nuclei. With this approach very good results are achieved for air showers and inclusive observables measured by collider experiments even up to LHC energies. This is remarkable since the model was introduced in the early 1990s and updated in 1999, long before the LHC delivered data.

The SIBYLL model describes string connections of projectile and target in the DPM as is used by PHOJET and DPMJET. In this picture of a nucleon-nucleon collision two strings are formed between a diquark and a quark of the target and the projectile. The Lund string

fragmentation is used to create mesons and baryons from the strings. The energy in the string is proportional to the sum of the quark and diquark fractions of momentum, x. For the quark, x is sampled randomly from a distribution

$$f(x) = \frac{(1-x)^3}{(x^2 + 0.35^2\,\mathrm{GeV}^2/s)^{1/4}}\,, \tag{44}$$

and the diquark receives the remaining momentum so that their sum gives unity. The particles produced in the fragmentation process obtain a p_T sampled from a Gaussian distribution with an energy dependent mean. In this basic description strings are connected between partons without specific treatment of a remnant resulting in an inevitable formation of a string between the constituents of the beam particles. The other partons are subsequently connected matching their colour and flavour.

SIBYLL also implements the minijet model to describe hard interactions. Above a $p_\mathrm{T}^\mathrm{min}$ that slowly rises with energy from 2 GeV at $\sqrt{s} = 10\,\mathrm{GeV}$ to 3 GeV at LHC energies, minijets can be produced for each scattering. The minijet cross section rises quickly with energy and becomes important at LHC energies. The number of interactions in each collision is given by $n_\mathrm{hard/soft} = A_\mathrm{hard/soft}(b)\sigma_\mathrm{minijet/soft}(s)$ for the respective hard and soft collisions. Whereas the hard part is derived by the minijet production, the soft cross section is given by a parametrisation closely linked to GRT. The profile function, $A(b)$, is given by a Gaussian that describes the fuzzy interaction region in impact parameter space that grows with smaller p_T of the partons (Pauli's principle).

Nuclear interactions are implemented via the Glauber model. In hadron-nucleus collisions, which are important later in this work, it leads to an intuitive expression to calculate the number of wounded nucleons (nucleons that are scattered), N_W,

$$\langle N_W \rangle = \frac{1}{\sigma^{hA}_\mathrm{prod}} \sum_{N_w} N_W \sigma_{N_W} = \frac{A\sigma^{hp}_\mathrm{prod}}{\sigma^{hA}_\mathrm{prod}}\,, \tag{45}$$

where $\sigma^{hA}_\mathrm{prod}$ is calculated from $\sigma^{hp}_\mathrm{prod}$.

Low-mass diffraction is implemented via the Good-Walker approach for longitudinal momentum loss $\xi_X < 0.1$. High-mass diffraction only takes place if the excited diffractive state has a mass $M_X > 10\,\mathrm{GeV}$. In SIBYLL it is described by a pion exchange. The resulting mass distribution follows $1/M_X^2$.

Currently the authors work on a retuned version to data from LHC experiments and to additionally include heavy particles made of charm quarks useful for studying backgrounds to the astrophysical neutrino spectrum measured by IceCube.

3.3.0.5 *Photo-Nuclear Collisions*

The STARLIGHT [KN99; But09] program generates electromagnetic processes. The calculations are based on two photons emerging from the electromagnetic beam particles. They can also be generated from the charged nuclei involved in heavy ion collisions and to

a lesser extent in (anti)proton collisions. In these Ultra-Peripheral Collisions (UPC), the electromagnetic field of the moving charges is described as equivalent photons within the Weizsäcker-Williams-Approach. The energy density of photons with energy ω at impact parameter b is described as

$$n(\omega, b) = \frac{Z^2\,\alpha\,\omega}{\pi^2\,\gamma^2}\left[K\left(\frac{\omega\,b}{\gamma}\right)\right]^2, \tag{46}$$

with the nuclear charge Z, the Lorentz factor γ, the fine structure constant α, and a modified Bessel function, as defined in [Wei34]. The differential cross section folded with $n(\omega, b)$ results in the total cross section as

$$\sigma_{\mathrm{EM}}\,(\mathrm{A\text{-}A} \to \mathrm{X}) = 2\int \mathrm{d}\omega\,\sigma_{\gamma\mathrm{A}\to\mathrm{X}}(\omega)\,n(\omega)\,, \tag{47}$$

where "A" denotes a nucleus. Ultimately a vector meson ("VA") or diffractive system ("XA") can be produced. The factor of two can be used for symmetric nuclei and would be close to unity for proton-nucleus collisions. The photon-nucleon cross section for a diffractive system $\sigma_{\gamma\mathrm{p}\to\mathrm{XA}}$ must be calculated by external models whereas $\sigma_{\gamma\mathrm{p}\to\mathrm{VA}}$ is given by an exponential fit to data given by the function

$$\left.\frac{\mathrm{d}\sigma_{\gamma\mathrm{p}\to\mathrm{VA}}}{\mathrm{d}t}\right|_{t=0} = b\left(X\,(s)^{\frac{\epsilon}{2}} + Y\,(s)^{-\frac{\eta}{2}}\right), \tag{48}$$

where s denotes the Mandelstam variable of the γp system and all other variables on the right-hand side are fit parameters and the dependence on the transverse momentum, t, of the cross section is approximated by an exponential distribution $\propto e^{-b|t|}$. The corresponding nucleus cross section needs to be calculated with the Glauber model.

In summary, photons from the electromagnetic field of each nucleon can collide and produce vector mesons or resonances. Both of processes are generated by STARLIGHT under the assumption that an external hadronic interaction model is linked to it.

3.4 INTERFACING HADRONIC INTERACTION MODELS

Interaction models used in particle physics are based on standards to enable straightforward comparison of output of different models. Cosmic ray models, however, have not adopted these standards. Dedicated tools like the previously mentioned air shower MC package CORSIKA are necessary to access event output and simulate air showers with hadronic interaction models. None of the tools are able to pass events in a format needed by particle physics experiments and proprietary solutions are used to link the generators to the detector simulation. In the following, the requirements and the implementation of a solution will be outlined. Its implementation has been realised in the frame of this work and has been released as the code package CRMC[1].

1 Its current release is version 1.5c that can be downloaded from https://web.ikp.kit.edu/rulrich/crmc.html (accessed 2015-03-10).

Comparisons between models and data are often straightforward but technical limitations do not allow for this in an easy and timely manner. A number of tools have been developed not only to standardise the model interaction information (LHE [Alw+07] (LesHouches) and the HepMC [DH01] formats) but also to automate the comparison between theory and data with, e. g., the RIVET framework [Buc+13][2]. To utilise RIVET, hadronic interaction models need to be compatible with such output formats. In particle physics the formats are widely used by general-purpose generators. Matrix-element-level generators can use the LHE format to transfer parton-based event information to another generator for hadronisation. Models used in the cosmic ray community do not support this. Instead each cosmic ray code uses its own specific interface to the models. Now with the rise of LHC analyses and enough computer power it is imperative to have a common interface that allows, on the one hand, ease of access to the models itself and, on the other, to provide a format that theorists and experimentalists can use at the same time. The interface developed here fulfils these requirements and is distributed under the name Cosmic Ray Monte Carlo interface, CRMC.

3.4.1 COSMIC RAY MONTE CARLO PROGRAM (CRMC)

CRMC is designed to give experimental physicists easy access to many different hadronic interaction models used in air shower simulations (see Fig. 48). A simple command allows the simulation of events directly after installation[3]. The package includes the latest code of each linked hadronic interaction model and gives the user the ability to quickly switch between them ("-m" command line option). The standard setup is compatible with many uses in particle physics. The main CRMC options are changed via the command line to change basic properties of the collisions like the collision energy or beam momenta ("-p" and "-P" options), beam type ("-i" and "-I" options) and additional settings connected to the written output file ("-o" and "-f" options for output type and filename). The number of events can be adjusted via the "-n" option. Batch jobs are the recommended way of parallelising the event generation.

Designated functions take care of calculations for Lorentz-boosting particles into the chosen reference frame, decaying unstable particles, or creating the event vertex structure. Such routines have to be run for each simulated event and are very important part of CRMC.

Once the program is started the compiled dynamic library of the selected model is linked and the model is initialised to the set collision parameters. Every model offers a Fortran routine to start the generation of one event, which is repeatedly called until the desired amount of event collisions has been written to a file.

2 A website to compare analyses currently available in the RIVET format can be found at http://mcplots.cern.ch/ (accessed 2015-03-10).

3 This can actually be done by running the CRMC executable in the bin/ folder, which will generate 100 p-p collisions at $\sqrt{s} = 7$ TeV and store them in a file with an automatically generated filename.

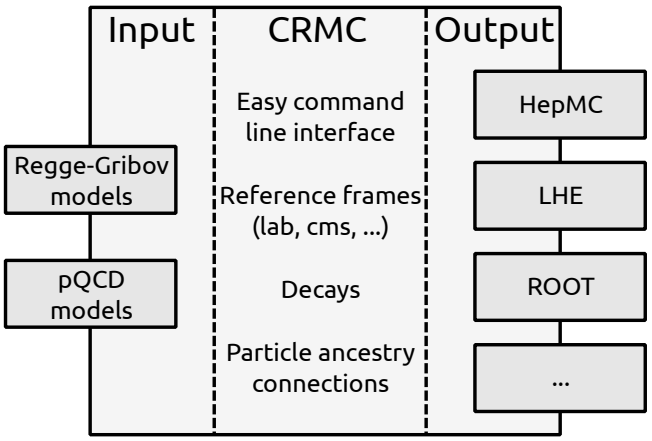


Figure 48: Schematic overview of functionality of CRMC. The program takes various input formats of various hadronic interaction models (left column) and outputs the generated events to formats often used in particle physics (right column). In between (middle column) many steps have to be generalised since different implementations exist for the individual models. These items are replaced or extended in the models without changing their physics.

3.4.2 AVAILABLE INTERACTION MODELS

The hadronic interaction models used to simulate air showers are linked to CRMC. The models DPMJET, EPOS, QGSJET, and SIBYLL were discussed in the previous sections. Apart from those, CRMC also contains additional models used in particle physics. The model HIJING1.383 uses a Glauber model to simulate nuclear effects [WG91]. Similar to DPMJET, HIJING uses an external generator, in this case JETSET, to calculate the nucleon-nucleon collisions. The JETSET model [SB87] is the predecessor of PYTHIA6 [SMS08]. PYTHIA6 is a general-purpose generator and the industry standard when it comes to high-energy collider physics and is also integrated into CRMC. Furthermore PHOJET can be run within CRMC.

Not all models provide the same functionality. Whereas all models are able to simulate proton-proton collisions, collisions involving heavy ions or mesons can only be obtained from specific models. A summary of the features of the individual models is listed in Table 5.

3.4.3 REFERENCE FRAMES OF RELATIVISTIC COLLISIONS

Particle collisions within the framework of each model described above are calculated in the centre-of-mass frame. The program CRMC specifically sets individual options of the models to work in this way. However, sometimes it can be advantageous to simulate events in the laboratory frame, for example when one wants to compare to air showers, fixed-target experiments but also to circular collider experiments with asymmetric beam conditions. CRMC therefore has the option to input the two absolute values of the beam momenta, p_1 and p_2, separately. In fixed-target experiments p_1 is often referred to as projectile momentum and the target has $p_2 = 0$. Together with the input of the particle IDs, which determine the particle masses that are read from a table, the four momenta $P_{1,2}$ for both beam particles are constructed.

Table 5: Features of models as included in CRMC. In some cases they might differ from the stand-alone versions of the models.

Model	Version	p-p	p-A & A-A	γ p	π, K_L^0	$\sqrt{s}$ Range [eV]
DPMJET	3.06	✓	✓	✓	✓	5×10^9 to 443×10^{12}
EPOS	1.99 and LHC	✓	✓	✗	✓	6×10^9 to 443×10^{12}
QGSJETII	03 and 04	✓	✓	✗	✓	3×10^9 to 443×10^{12}
SIBYLL	2.1	✓	A<56 [4]	✗	✓	10×10^9 to 443×10^{12}
HIJING	1.35	✓	✓	✗	✗	$< 14 \times 10^{12}$
PHOJET	1.12	✓	✗	✓	✓	5×10^9 to 443×10^{12}
PYTHIA	6.1 and 6.4	✓	✗	✗	✓	$< 50 \times 10^{12}$

The centre-of-mass energy is calculated by

$$\sqrt{s_{NN}} = (P_1 + P_2)^2 \,, \tag{49}$$

where the index NN refers to the nucleon-nucleon system in case of ion collisions and to the hadron-hadron system otherwise. The hadrons can, for example, be protons or mesons. The centre-of-mass energy is used as input for the models to calculate the collision. Afterwards a Lorentz boost along the z-axis is applied to the results returned by the model. It is applied to both the vertex position and particle momenta. Since both primary particles are assumed to travel along the z-axis, the rapidity to boost as defined in Eq. (6) can be expressed by

$$y_{\text{boost}} = \frac{1}{2} \ln \left(\frac{(E_1 + E_2) + (p_1 + p_2)}{(E_1 + E_2) - (p_1 + p_2)} \right) \,, \tag{50}$$

where y is the rapidity of the laboratory system. Using Eq. (7), CRMC boosts the particles into the laboratory frame.

3.4.4 PARTICLE DECAYS

Final-state particles with short lifetimes are likely to decay before they reach the detector of an experiment. Thus, it is common to assume a minimum value for $c\tau$ above which particles are regarded as stable by the generator. The product is the distance a particle with speed of light, c, travels its mean decay time, τ. In the frame of the observer a particle with the Lorentz factor γ will on average travel $\gamma c\tau$ before it decays. Setting the value without the Lorentz factor, makes the option robust against Lorentz transformations. In the `HepMC` format, a status code of 2 is assigned to all unstable particles whereas the status code for stable particles, including the decay products of unstable particles, is 1.

The option for the minimum decay time can be set in the file *crmc.cfg* with a default value of $c\tau = 1\,\text{cm}$. This setting classifies $\pi^{\pm}$, kaons including the K_S^0, and so on as final-state particles but induces the decay of short-lived particles such as neutral pions.

Decays in CRMC are either handled by the decay routine of the selected model itself or are piped through the routine of EPOS. The latter method is applied to models that do not integrate the ancestry information of all decay vertices. Particle decay is a standard procedure and in general implementations differ only for rare decays, such as decay modes with unknown ratios. Using the routine of EPOS does not influence the output of a selected model but depending on the minimum decay time can add more diverse particle types to the event list than originally foreseen by a particular model. The decay routine of EPOS reads in a list of possible decays for each type[5] with its associated ratio to occur.

The distribution of the mass of an unstable particle has a finite width. This Breit-Wigner distribution is important for short-lived resonances. The width w in units of energy [GeV] of the Breit-Wigner distribution is related to $c\tau$ as

$$w = h\,\nu = 2\pi\hbar\,\frac{c}{2\pi\,c\tau} = \frac{0.197}{\gamma c\tau}\,, \tag{51}$$

with the Planck constant, $\hbar$. In EPOS-LHC (similarly to PYTHIA) the masses of the particle types are actually fixed to the mean value[6] but the mass used in decays is sampled from a Breit-Wigner distribution. To conserve energy, the energies of the decay products are rescaled compensating the modification. The mass used in decays is sampled within w_{atmax} around the mean mass. The method to sample the Breit-Wigner distribution is based on inverse transform sampling and is explained in Appendix A.1. The modified mass is given by

$$m' = m + \frac{w}{2}\tan(x)\,, \tag{52}$$

where $x = [-\arctan(w_{\text{atmax}})\,, \arctan(w_{\text{atmax}})]$. The latter defines the maximum broadening of the peak and is set to 20 by default.

Having particle decays turned on, often gives a more realistic picture of the simulated event when compared to a real collision seen in a detector at a collider experiment. Most short-lived particles decay before they reach even the first tracking layer and only their decay products can be detected. This can have an impact in the forward region where detectors are used that are more than a hundred meters away from the interaction point. Typically the Lorentz factor of the particles in the forward region are larger by a few dozen times than in the central detector. Figure 49 shows that the energy in the forward region is stable within less than a percent for a minimum τ lower than $10^{-8}\,\text{s}$. This includes the default stable particle definition. Only at values larger than $10^{-8}\,\text{s}$ the energy of the particles increases by roughly 3 %. The number of particles, however, goes up by 100 %. The definition for stable particles must, therefore, be chosen carefully.

5 The list can be found in src/epos/epos_dky.f.

6 A special branch of CRMC is currently under development where the particles are already created with a mass according to the Breit-Wigner mass distribution.

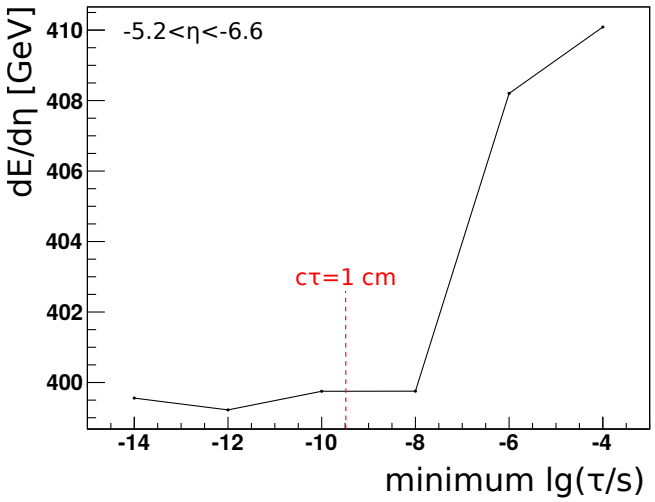


Figure 49: For simulations, the definition of the final state is of importance. Here, the energy over pseudorapidity measurable within the region of the CASTOR detector ($-6.6 < \eta < -5.2$) is shown depending on the minimum decay time τ selected to define a stable particle.

3.4.4.1 *Output Formats*

One of the advantages of CRMC is the free choice of the output format for every model. Three of the most common ones used by the software frameworks of particle physics experiments are `HepMC` [DH01], `LHE` [Alw+07], and ROOT [BR97]. All three are supported by CRMC and an empty template provided within the code allows the user to easily add and link other output formats (see src/OutputPolicyNone.cc).

`HepMC` is written in the programming language `C++` but offers an interface to `Fortran` written programs, which is useful since most interaction models are still based on `Fortran`. The event record stores information about beam particles, produced particles, decay vertices, and additional sets of variables that can be added optionally. CRMC makes use of additional heavy ion-related variables and cross section-related variables. One highlight of the `HepMC` format is the saved ancestry information in vertices, whose parent and daughter particles can be accessed in a straightforward way. Having information about time and position of the vertices allows for, e. g., analyses that require knowledge about prompt particle decays. In CRMC, the `HepMC` format can be selected by option "-o hepmc" and automatic compression with the `gzip` algorithm is available by selecting the "-o hepmcgz". When enabled the file size is reduced by a factor of about 4.

`HepMC` focuses mainly on final-state particles. The `LHE` format, however, has its focus on parton-level information. At this stage partons (quarks, gluons, ...) have not yet hadronised. In contrast to cosmic ray physics, models that describe higher-order loop diagrams or other rare processes can make use of this by only simulating the physics at parton level and on top using a standard program such as PYTHIA to hadronise the particles. The information is best passed on with the `LHE` event format. Like `HepMC` a `Fortran` common block has to be filled which is written to files by calling Fortan routines. To select `LHE` one can use option "-o lhe". Both `HepMC` and `LHE` write human-readable files.

The ROOT output of CRMC contains only basic information about the produced particles but sufficient information to construct their Lorentz vectors. Information about the event is

kept as well, e. g., the cross section, impact parameter, and random seed. The information is saved in a ROOT tree and is therefore easy accessible. It is also the format offering best compression.

3.4.4.2 *Validation*

Prior to release, output of CRMC is compared to the raw output of models. In most cases basic inclusive variables can tell whether the code runs as it is supposed to. With large enough statistics even small deviations can be found in comparing $\mathrm{d}N/\mathrm{d}\eta$ and $\mathrm{d}E/\mathrm{d}\eta$ distributions. Examples are shown in the plots of Fig. 50. Both the distribution of energy and number of particles agree well within the statistical uncertainties and we can conclude that CRMC does not change the output of the models.

Continuous validation over the time of development of the interface is also of great importance. Due to constant development a "test mode" was introduced that provides a quick check of basic variables to confirm that particle production is not influenced by the interface. Changes involving decays or directly the particle tree like reordering or connecting mother and daughter statuses need to be monitored closely. CRMC can be run with the "-T" option, for which 1000 events (or as many as specified with the "-n" option) are simulated and the number of produced particle, the total energy of the final-state secondaries, the mean plateau height of the $\mathrm{d}N/\mathrm{d}\eta$ distribution, and the cross sections of the event type are given.

3.4.4.3 *Code Structure and Build System*

Installation of CRMC is handled via the open-source build manager `CMake`, which checks dependencies on the obligatory packages: a `C++` compiler, a `Fortran` compiler, `HepMC`, and `Boost` as well as the optional packages `Root` and `FastJet`. The `LHE` interface does not need any further dependency. While running the "cmake <source path>" command from within the installation path, a `Makefile` will be generated. With all dependencies found, installation is a trivial matter of running "make install".

The procedure described above will compile and install all included and selected[7] models and the actual interface to each of them. The models will be built with the `Fortran` compiler and the direct interface routines such as initialising the `Fortran` common block variables, reading configuration from files are also written in `Fortran`. Every model has to be initialised with the set momenta and types of the beam particles. Afterwards the boost required to transform into the centre-of-mass system is calculated. Furthermore, parameters for paths to various tables of the models, parameters for EPOS, and decay parameters for all models are set[8]. The last step in initialisation contains the calculation of the cross sections.

7 Options can be found in `CMakeLists.txt` in the source folder.

8 The settings are contained within the `crmc.param` file in the installation path. Best practice is to make a copy of that file and use the "-c" option.

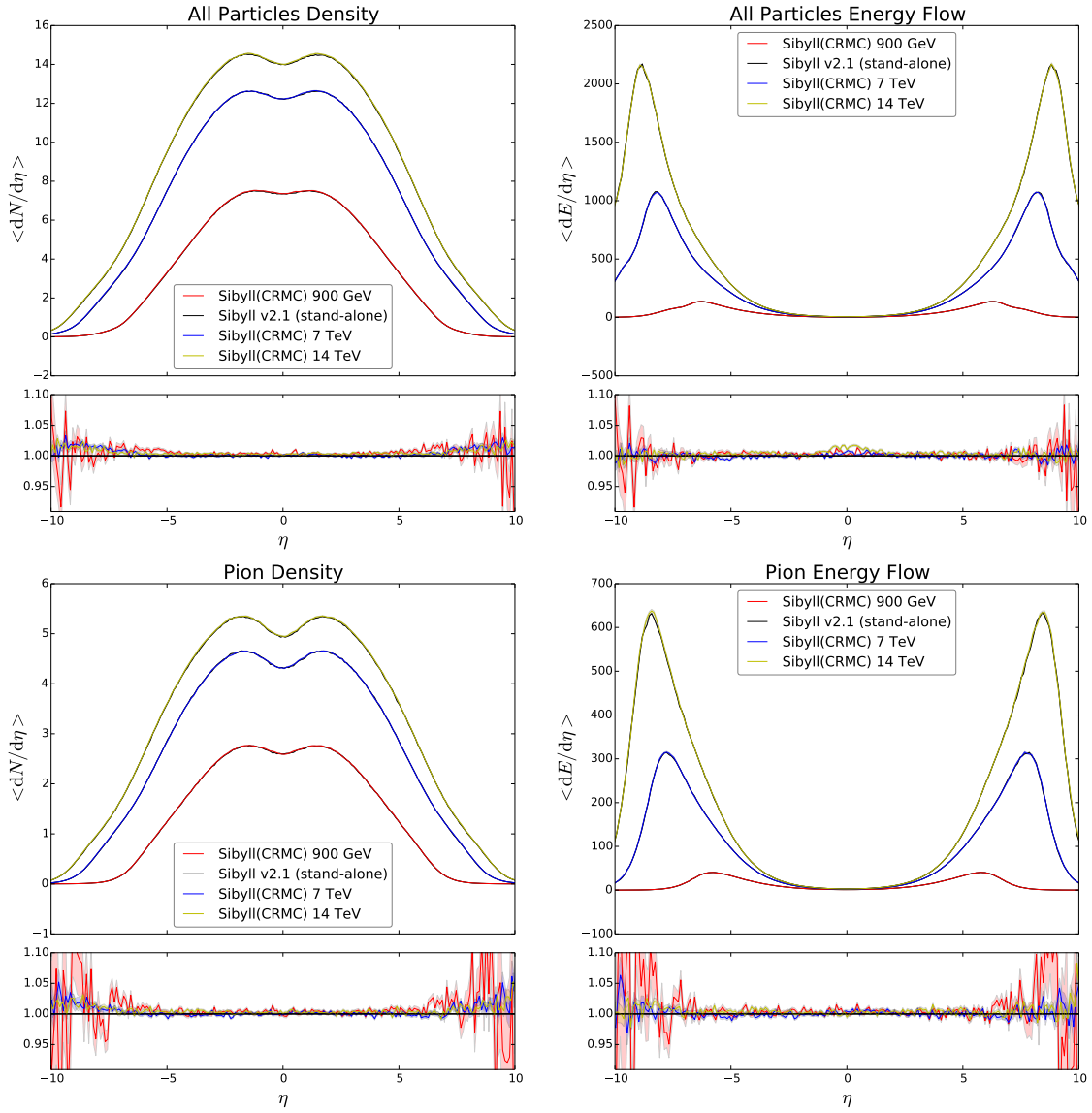


Figure 50: Comparison of running SIBYLL without the CRMC interface and running it with the interface. The top plots show the average particle distribution per event per unit of pseudorapidity (left panel) and the energy deposited per event per unit of pseudorapidity by all final state particles (right panel). The lower plots show the same for pions only. This test bypasses effects introduced by decays.

The code of the models is separated into sub-folders and therefore different versions can quickly be interchanged. With the exception of QGSJETII and EPOS, whose previous versions are still in use, only one version of the model can be selected in CRMC. The same is true for PYTHIA, where usually several parameter sets (tunes) are used, only the version 6.4.28 is enabled by default. However, in CRMC the focus lies on supporting models used in air shower simulation whereas PYTHIA is only included as a form of validation.

As an additional feature for QGSJETII the large tables, that require around 220 MB, are provided in a compressed versions and are unpacked during runtime with the `zlib` library by the interface. With the chosen `LZMA` compression only around 50 MB of disk space is needed for each version of QGSJETII. For slower drives like network storage this can lead to a significantly lower start-up time for this model.

All high-level functions such as the command line interface, the output handling, and the control over initial model parameters are developed in `C++`. The command line options are passed by `boost/program_options`. The loop to simulated multiple events can therefore also be found in the `C++` part. Typically `extern` functions are called to directly interact with the `Fortran` routines. These are linked as dynamic libraries and only the library of the model needed for the event generation is loaded. This saves a significant amount of memory if many models are linked. It also makes an exchange of model versions possible. An option[9] to use static instead of dynamic libraries that can be useful for `MacOS` users is available.

Lastly the output handlers all derive from a class called `OutputPolicy` and can access the same information about an event. The derived class for each output format only has to handle the file writing process. For `HepMC` a common block specified by its protocol is transferred by reading an `extern` memory block. For the other output formats the variables get transferred via a function and can be fed to the formatting process. In this way event information can be written to the user specified format.

3.4.5 APPLICATION WITHIN THE CMS FRAMEWORK (CMSSW)

Naturally, it is of great interest to link the cosmic ray generators to the detector simulation of CMS. This standardises the usage in large simulation campaigns needed to evaluate physics interests and optimise triggers even before taking data. The framework supports a large amount of models compiled either as external packages or directly integrated into the main code. For the hadronic interaction models used in air shower simulation, CRMC is ported to be used within the framework. The current supported models are: EPOS, QGSJETII, and SIBYLL.

Some problems had to be overcome to make this possible. Due to strict policies about compiler warning messages that require no warning messages for unused or uninitialised variables some of the code of the models has been updated and the changes are also in-

9 Also in `CMakeLists.txt`.

cluded in CRMC. The tables of the models are placed in an external storage folder shared by multiple versions of the CMS framework to save storage space. Additionally, the boost into laboratory frame of CMS had to be switched off, since this happens automatically during the vertex smearing of the generator particles.

The HepMC output format is used to pass event information on to the framework. This is the most natural choice to store the full ancestry information and many other models are connected in the same way to the framework. One important aspect is the passing of the event type identification, that contains information about whether a non-, single-, double-, or central-diffractive collision took place. The process ID is implemented to be conform with process IDs from PYTHIA8. This feature is useful for many forward analyses and is supported by EPOS and QGSJET.

The interface has already been used extensively to simulate collisions with EPOS and QGSJET for proton-lead data taking conditions and all three models have been employed for proton-proton collisions at a centre-of-mass energy of 7 and 8 TeV. Official production campaigns included proton-lead collisions simulated with EPOS and a number of requests for the upcoming 13 TeV data are planned.

4

DATA ANALYSIS

It was an excellent opportunity that in 2013, for the first time, collisions between proton and lead nuclei took place at the LHC at CERN. These data allow to study nuclear effects at the LHC. The collisions are much more similar to cosmic ray interactions in the atmosphere than p-p collisions. It is one of the major achievements of this thesis to perform the first measurement of the production cross section in proton-lead collisions at a centre-of-mass energy of 5.02 TeV per nucleon pair [CMS13b]. This measurement at LHC energies is so far unique and requires new methods to treat, e. g., electromagnetic and photo-hadronic processes, for which secondary particles are produced. So far the p-Pb production cross section, σ_{prod}, has never been measured at similar energies. The relevance of screening and short-range-correlation effects can be determined from these data. The magnitude of the acceptance correction due to diffraction is determined with the CASTOR detector. In addition, the importance of the measurement for air shower physics is evaluated.

To measure σ_{prod} it is first of all necessary to precisely determine the luminosity of the proton-lead beams. Special measurement of the particle beam densities are conducted and all relevant second-order effects like displaced charges in the bunch filling scheme or afterglow in the luminometer are studied in detail. The analysis of the luminosity in proton-lead beams was mainly driven by this production cross section measurement and is entangled with it. Any uncertainty on the luminosity determination enters directly into the systematic uncertainty on the cross section. The analysis is published in a different reference [CMS14d], however, significant contributions have been provided as an integral part of the measurement of σ_{prod}.

Obtaining the production cross section studies the important nuclear effects, yet, this is only one aspect of hadronic interaction models. In addition, a global approach is taken to evaluate multiple parameters of hadronic interaction models at once. There are multiple measurements at LHC that constrain these parameters and these are combined with measurements from air showers to evaluate how sensitive predictions of all measurements are to varying such parameters.

4.1 MEASUREMENT OF THE PRODUCTION CROSS SECTION OF PROTON-LEAD COLLISIONS AT LHC ENERGIES

The production cross section in proton-lead collisions is measured with the CMS experiment at $\sqrt{s_{NN}} = 5.02\,\text{TeV}$. The terminology requires some clarification because often no separation is made between the inelastic and the production cross section. In inelastic collisions, the beam particles and/or the absolute values of their momenta are not conserved. In

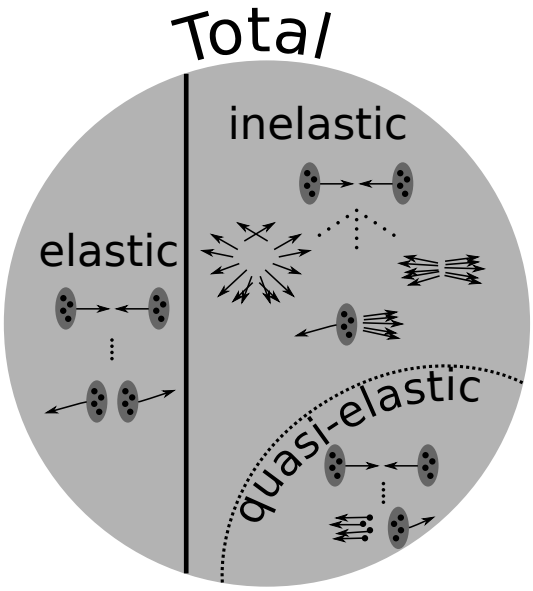


Figure 51: Simplified drawing of the components of the total cross section. In elastic collisions, the nuclei stay intact but are scattered. For inelastic collisions either new particles can be produced or at least one of nuclei breaks up. The latter is categorised as quasi-elastic since no other particles are produced. For the other part, three processes are shown: non-diffractive, single-diffractive and double-diffractive (left-to-right).

collisions involving nuclei, the production cross section refers to the part of the hadronic inelastic cross section that produces particles, thus, excluding quasi-elastic events. Figure 51 symbolises the three main components of the total cross section. The total cross section is consistently defined as the sum of elastic, inelastic, and quasi-elastic cross sections The cross section excluding the quasi-elastic component is referred to as the production cross section, σ_{prod}, in the following [Eng+98]. At lower energy the amount of quasi-elastic cross section is estimated to be smaller than 5 % of the inelastic cross section [Car+79]. At LHC energies this fraction becomes even smaller. In nucleon-nucleon collisions, the concept of quasi-elastic scattering is not required.

The production cross section determines the probability for a projectile particle to interact and produce secondary particles with a given target particle. It is a useful quantity for collider experiments like the LHC but also for extensive air showers. The measurement of σ_{prod} is crucial since the cross section cannot be calculated from first-principles by the theory of the strong interaction.

However, nuclear cross sections can be derived from phenomenological calculations with input from proton-proton cross section measurements. The necessary measurements of the proton-proton cross sections were performed by different experiments in a wide range of energies up to $\sqrt{s} = 8\,\text{TeV}$. The proton-proton cross section is independently measured by the ATLAS Collaboration [Zen12; Aad+14], the CMS Collaboration [Cha+13a], and the TOTEM Collaboration [Ant+13c; Ant+13a]. Additionally proton-antiproton measurements at lower centre of mass energies, e. g., by the CDF Collaboration [Abe+94] help to understand the energy dependence at even higher energies as needed in comic ray physics. To a good accuracy, one can parametrise the inelastic cross section by physics-motivated power law fits. One of the more sophisticated approaches is proposed by the COMPETE Col-

laboration [Cud+02]. Another more recent parametrisation that contains the LHC proton-proton results is given in [MS13a]. An important drawback is that these parametrisations contain only limited predictive power towards higher energies and the uncertainty from extrapolation can rapidly increase [BUE15].

Furthermore, nuclear interaction models need the knowledge of the nuclear cross section to have access to quantities such as the nuclear overlap function, the number of nucleon-nucleon collisions, and the number of participant nucleons, each of which are required for many physics applications [dEn03; Mil+07]. The same quantities are crucial for experimental comparisons between nuclear and more elementary nucleon-nucleon collisions. One important application of these models is their use in cosmic ray air shower physics at the highest energies where inelastic cross sections must be extrapolated over a wide range of energies and to a variety of collision systems. Measurements at high energy reduce the resulting extrapolation uncertainty. In fact, the inelastic cross section introduces one of the largest uncertainties to air shower simulations [UEU11; Par+11]. Recently, the sensitivity of air showers to the inelastic cross section was exploited by the Pierre Auger Collaboration by measuring the proton-air cross section at $\sqrt{s_{\mathrm{NN}}} = 57\,\mathrm{TeV}$ [Abr+12]. Also the p-p cross section was determined in this context; however, the precision of the extended Glauber calculation with screening corrections gives rise to a large model dependence of the result.

Nuclear cross sections are in general derived with the Glauber model or Gribov-Regge Theory (GRT) calculations. This leads to known shortcomings with two dominant effects, short-range correlations and inelastic screening (Section 2.3.5). These corrections can have a significant impact on the cross section, as is measured for low energy proton-carbon data [Der+00; Bel66; Mur+75; Eng+70; Bab+74]. For centre-of-mass energies as seen at LHC or beyond, such measurements do so far not exist, yet the size of the corrections is predicted to be a few percent for the proton-lead data at hand [Cio+11; Alv+13]. Another study suggests that it can be up to 10 % [Abr+12] for proton-air collisions. Such a large effect can be detected by a measurement at the LHC, but one should bear in mind that the corrections have opposite sign and might cancel out in the cross section measurement.

Previously, the proton-lead cross section was measured only at much lower energies [Den+73; Car+79; Ava+86]. Now, with data from the LHC, it can be measured at a centre-of-mass energy of 5.02 TeV per nucleon pair. For this, an accurate determination of the luminosity is of the utmost importance. With the CMS detector the most straightforward method is to employ its forward detectors to count inelastic events, whose collision products deposit energy in the calorimeters. When the number of all events with hadronic particle production, N_{prod}, and the integrated luminosity, $\mathcal{L}$, are determined, the cross section is calculated as

$$\sigma_{\mathrm{prod}} = \frac{N_{\mathrm{prod}}}{\mathcal{L}}\,. \tag{53}$$

The acceptance to detect inelastic particle production is much larger in the forward region. Additionally, high- to medium-mass diffractive events can easily be observed. In principle, the central detector can also be used; this has the advantage of a low noise level and

allows the primary vertices to be reconstructed. The distribution of the number of primary vertices in an event is related to the cross section. However, the experimental acceptance is much worse for vertex counting as all collisions that do not produce a sufficient number of charged-particle tracks to form a vertex are lost. This includes almost all diffractive events. For the p-p cross section the Hadronic Forward (HF) detector is used to measure events with diffractive masses as low as $\xi > 6\times10^{-5}$. The acceptance of different detectors in p-Pb collisions is explored in the scope of this measurement.

4.1.1 PROTON-LEAD COLLISION DATA OF 2013

In 2013, the first proton-lead collisions at the LHC took place. The first collisions under stable conditions were taken on 2013-01-20 at 13:22:23. The data were collected with the proton beam injected both clockwise and counter-clockwise. The beam momentum is 4.00 TeV/c for the protons and 1.58 TeV/c per nucleon for the lead ions. The momenta differ due to the charge and mass difference of the beam particles and the requirement that both beams go around the LHC with the same radius as they are bent by the same magnetic field strength. With the charge of lead ions $Z_{\text{Pb-ion}} = 82$ and their mass $A_{\text{Pb-ion}} = 208$, one can calculate the momentum of the ion as a function of the proton momentum using the Larmor radius

$$|\vec{p}_{\text{Pb}}| = \frac{|\vec{p}_{\text{Pb-ion}}|}{208} = |\vec{p}_{\text{p}}|\frac{|Z_{\text{Pb-ion}}|}{208|Z_{\text{p}}|}\,, \tag{54}$$

where we differentiate between the momentum of the whole nucleus $\vec{p}_{\text{Pb-ion}}$ and the momentum of one nucleon, $\vec{p}_{\text{Pb}}$. The latter quantity is of interest to us in later calculations. Under these conditions, the laboratory frame of the CMS detector is not the same as the centre-of-mass frame of the projectile and target. The difference in rapidity between the two frames is $y_{\text{frame}} = 0.47$. When transforming from the centre-of-mass frame into the laboratory frame, all produced particles and the beams are boosted in the direction of the proton.

The beam configuration with (counter-)clockwise proton beams, is called (p-Pb) Pb-p. For p-Pb, the protons move towards the negative z-direction in CMS and for p-Pb to the positive z-direction. The CMS detector is asymmetric due to the CASTOR calorimeter, which is installed only on the negative pseudorapidity side from the interaction point. For this reason, CASTOR is not used for the main cross section analysis but for assisting studies about diffraction.

In CMS, a "run" denotes a session during which data are continuously recorded and written to the storage elements. They often coincide with LHC fills, whose start is the particle injection procedure into bunch slots. As more and more collisions take place, the beam particles are drained and used up. This limits the maximum amount of time an LHC fill can last. For proton-lead collisions, this corresponds to about 1000 luminosity sections, where one luminosity section is defined by the time interval of 23.31 s. Combining all LHC fills, a total luminosity of 35.50 nb^{-1} was collected (see left side of Fig. 52). After recording 21 nb^{-1} the beam direction was reversed. Each fill has slightly different beam

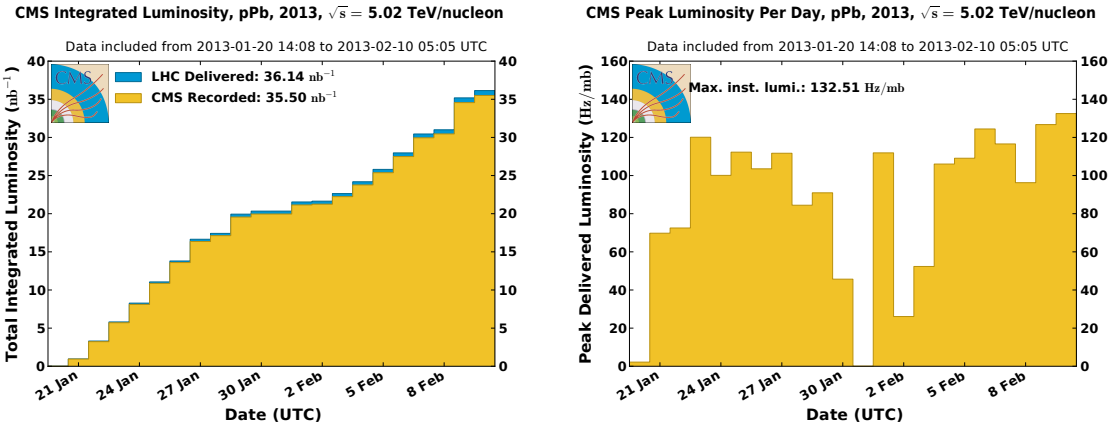


Figure 52: Recorded and delivered luminosity at the CMS interaction point for the p-Pb collisions. The left panel shows the integrated luminosity and the right panel the peak luminosity per day.

conditions and different numbers of particles filled into the 3564 available bunch slots of LHC. This leads to a varying peak instantaneous luminosity from fill to fill (see right side of Fig. 52), which influences the amount of multiple collisions ("pileup") that can occur in an event. Pileup collisions can happen independently from the first collision and are, thus, Poisson-distributed. The pileup collisions depend on the cross section and the interaction probability, which is

$$\lambda = \mathcal{L}_{\text{inst}} \sigma_{\text{inel}} / (\nu N_{\text{bunches}}) \,, \tag{55}$$

where $\mathcal{L}_{\text{inst}}$ is the instantaneous luminosity, $\nu = 11\,245.6\,\text{Hz}$ is the orbit frequency of the LHC, and N_{bunches} is the number of observed colliding bunches. The distribution for i simultaneous interactions to occur follows a Poisson distribution with mean λ,

$$\text{Pois}(i;\lambda) = \frac{\lambda^i}{i!} e^{-\lambda} \,. \tag{56}$$

Towards the end of a fill the pileup collisions become less frequent. The peak pileup varied between 2 % and 8 % for the data analysed in this study. Not all data of the fills are taken into account since measuring σ_{prod} is not limited by statistical but rather by systematic uncertainties. Table 6 lists the integrated luminosities and peak pileup values for each of the included runs. Altogether, the data used in this analysis comprise an integrated luminosity of $\mathcal{L} = (12.6 \pm 0.4)\ \text{nb}^{-1}$ where $(4.95 \pm 0.17)\ \text{nb}^{-1}$ were taken with the proton-lead beam configuration and the rest with lead-proton beam configuration.

Except for the first run in the table, which had $N_{\text{bunches}} = 206$ filled bunch slots, all the others were taken with the number of occupied bunch slots, $N_{\text{bunches}} = 296$.

Table 6: Runs used for event counting in this analysis. Other runs were used for van-der-Meer scans and noise studies that are not listed here.

Run number	Integrated luminosity (nb^{-1})	Peak pileup (%)
210614	1.38	4.42
210885	1.31	6.28
210998	0.568	4.57
211000	0.387	3.03
211001	0.125	2.31
211032	1.18	4.95
	Beam reversal	
211256	1.18	7.11
211371	0.631	6.68
211390	0.952	6.34
211460	1.09	7.83
211532	1.27	7.31
211538	1.34	7.36
211607	1.18	7.63
Total (13)	12.6	-

4.1.2 VAN-DER-MEER SCANS FOR LUMINOSITY DETERMINATION

At the CMS experiment, the luminosity is estimated in real time during the data acquisition. There are two luminometers used for this, where a luminometer is a detector that is sensitive to the instantaneous luminosity and ideally has a linear response function. The strip and pixel trackers can be used to either count hits in the silicon pixels or to count the number of reconstructed vertices. Another method is based on counting hits in the HF detector. For the p-Pb data acquisition the luminosity is estimated using the latter. The visible energy in the calorimeter is positively correlated with the instantaneous luminosity. The relative luminosity in a time interval can be accurately determined with HF and is usually scaled to an absolute luminosity by an estimation of the visible cross section. This was also done during a feasibility study in November 2012. The measured event rate, R, was multiplied by the efficiency to see an event in HF estimated to be $\epsilon = 0.97$, assuming a production cross section of $\sigma_{\text{prod}} = 2.1\,\text{b}$, and subsequently transforming Eq. (53) to obtain $\mathcal{L}_{\text{inst}} = \epsilon\, R/\sigma_{\text{prod}}$. This results in an estimate of the instantaneous luminosity for a given rate. However, to achieve accurate results this scale has to be calibrated offline (after the data acquisition). This can be achieved in multiple ways: An accurately known cross section from theory for an electromagnetic process can serve as a standard candle. In this case,

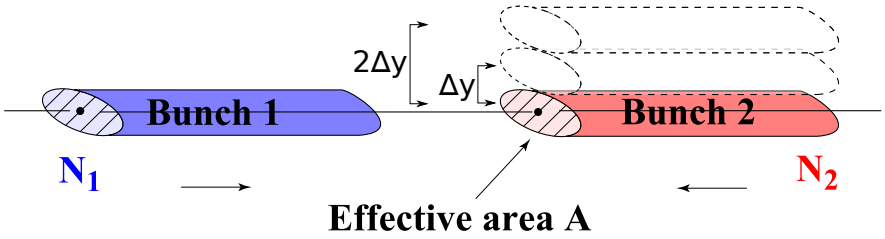


Figure 53: Drawing of the effective area of two colliding particle bunches. The dashed bunches show two possible shifts in y for bunch 2 as applied during the vdM scans. (modified from [WBP10])

one can normalise the visible cross section times the measured event rate to the expectancy from the calculation. Another possibility is to make use of the optical theorem to obtain a cross section via the elastic slope but this is limited to nucleon-nucleon collision and cannot be applied to beams studied here. We present the van-der-Meer scan method (vdM) to achieve an absolute calibration. The method was originally proposed by Simon van der Meer to calibrate the CERN ISR luminosity [Mee68]. The method consists of displacing colliding bunches against each other in the plane transverse to the beam axis (the xy-plane) in order to determine their overlap region. Thus, the beams are displaced with respect to one another in the xy-plane by certain distances, and the corresponding interaction rate is measured as a function of the displacement (these shifts of the beams are indicated in Fig. 53). Concurrently, individual bunch currents are measured by dedicated devices, the Fast Beam Current Transformers (FBCT) (Section 3.1.1).

4.1.2.1 *Overview of Beam Parameters*

At the LHC, beams are divided into 35640 RF-buckets. Each tenth can be filled, resulting in 3564 bunch slots, which are 25 ns long. Only 2808 of the 3564 bunch slots can be filled due to abort and injection gaps. Charges slowly diffuse from one bucket into neighbouring buckets or even bunch slots. These effects are denoted as satellite and ghost charges, respectively. Typically, the ghost charge fraction is very low and satellite charges even lower. A schematic overview of the buckets, bunches, ghosts and satellites is shown in Fig. 54.

Satellite charges contribute towards the luminosity of a bunch when there is no crossing angle between the beams. For the January and February scans when the beams collided under a non-zero crossing angle, the satellite charges need to be subtracted from the measured charge. When obtaining the luminosity scale via vdM scans, the signal due to ghost charges has to be subtracted. The luminosity of a beam depends on the individual charges of the bunches. The charge of all bunches in beam 1 and in beam 2, $B = \{1, 2\}$, can be written as a sum over the charges in every bunch slot, $N_B = \sum_{i=0}^{N_{\text{bunches}}} n_i$. Then, the luminosity is defined as [GK14]

$$\mathcal{L} = \nu\, 2c\, \cos(\alpha)\, N_1\, N_2 \int \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z\, \mathrm{d}t\, \rho_1(x, y, z, t)\, \rho_2(x, y, z, t)\,, \tag{57}$$

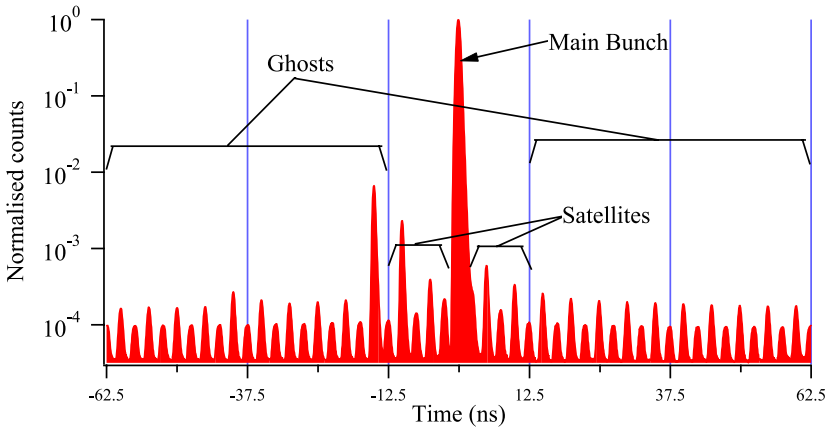


Figure 54: Classification of ghost and satellite charges as they appear separated in time from the actual filled bunch slot. (from [Ali+12])

with the LHC orbit frequency, ν, particles in bunch from beam 1 and 2, $N_{1,2}$, and time- and position-dependent density functions, $\rho_{1,2}$. The half angle, α, is defined by the two beam trajectories in the xy-plane.

For each beam, one can define the betatron function, $\beta(z)$, the emittance, ϵ, and the transverse beam sizes, $\sigma(z)$. The betatron function is defined as $\beta(z) = \beta^* + (z - z_{\text{IP}})^2/\beta^*$. In this definition, β^* equals the value of the betatron function at the z-value of the interaction point, z_{IP}. The emittance is the volume in the coordinate-angle phase-space that every particle in the beam consumes. In just one dimension (x or y for the longitudinal direction), the emittance is typically given in units of [mm mrad] where the angle is defined between the z-axis and the axis of the respective direction. Liouville's theorem states that the emittance is conserved. This leads to the effect that focusing a beam reduces the effective area of it but the angles will get larger. This way, an analogy between beam optics as implemented using the quadrupole magnets and classical optics is established.

The transverse beam sizes are defined as $\sigma(z) = \sqrt{\epsilon\,\beta(z)}$, where the emittance in general does not depend on z. To maximise luminosity, σ needs to be minimal at the IP, where β is identical to β^* at the interaction point.

For one single beam, the particle density for a Gaussian beam profile can be written as

$$\rho_B(x,y,z,t) = \frac{1}{\sqrt{(2\pi)^3}\sigma_{xB}\sigma_{yB}\sigma_{zB}} \exp\left[-\frac{(x-x_B)^2}{2\sigma_{xB}^2} - \frac{(y-y_B)^2}{2\sigma_{yB}^2} - \frac{(z-ct)^2}{2\sigma_{zB}^2}\right], \tag{58}$$

where $(x - x_B)$ and its equivalent in the y-direction is the distance to the geometric centre of the bunch. Combining Eqs. (57) and (58) and integrating over time leads to the equation for the luminous region, $\mathcal{L}(x,y)$, which only depends on the transverse coordinates.

4.1.2.2 *Mathematical Concept of Van-der-Meer Scans*

During vdM scans, the beam profile, ρ, is measured for both beams simultaneously. The profile is measured step-wise with offsets to the beam centre in the x- and y-directions, Δ_x and Δ_y. The z-direction is of no importance when the crossing-angle of the beam is small

and one integrates over the time of the measurement. More explicitly, for all particles of a bunch, the probability to interact with the other beam is not reduced even when the beam profile is not a δ-function in z. When multiple values of $\Delta_{x,y}$ are sampled, ρ can be fitted with a Gaussian function. To include other influences on the profile, sometimes a double Gaussian or a Gaussian plus a high-order polynomial function is fitted. The additional terms then describe second-order corrections mostly induced by Coulomb forces.

The luminosity dependence on the beam displacements is

$$\mathcal{L}(\Delta_x, \Delta_y) = N_1 N_2 \int \mathrm{d}x\, \mathrm{d}y\, \rho_1(x,y)\, \rho_2(x-\Delta_x, y-\Delta_y)\,. \tag{59}$$

The profile functions are normalised,

$$\int \mathrm{d}x\, \mathrm{d}y\, \rho_{1,2}(x,y) = 1\,. \tag{60}$$

The assumption needed for vdM scans is that the profile function factorises into two independent functions of x and y,

$$\rho(x,y) = \rho_x(x)\, \rho_y(y)\,. \tag{61}$$

where each function is normalised independently

$$\int \mathrm{d}x\, \rho_x(x) = \int \mathrm{d}y\, \rho_y(y) = 1\,. \tag{62}$$

Inserting the factorised profile functions in Eq. (59) leads to a product of two one-dimensional integrals

$$\begin{aligned}\mathcal{L}(\Delta_x, \Delta_y) = \nu N_1 N_2 \left[\int \mathrm{d}x\, \rho_{x1}(x)\, \rho_{x2}(x-\Delta_x)\right] \times \\ \left[\int \mathrm{d}y\, \rho_{y1}(y)\, \rho_{y2}(y-\Delta_y)\right] \equiv k\, R(\Delta_x, \Delta_y)\,.\end{aligned} \tag{63}$$

On the right-hand side, the equation is expressed by the response of the luminometer, R, and a calibration constant, k, that does not depend on $\Delta_{x,y}$. Now the normalisation of ρ can be exploited to write Eq. (63) as two equalities by integrating over Δ_y or Δ_x, respectively,

$$\begin{aligned}\nu N_1 N_2 \left[\int \mathrm{d}x\, \rho_{y1}(y)\rho_{y2}(y-\Delta_y)\right] &= k \int \mathrm{d}\Delta_x\, R(\Delta_x, \Delta_y) \\ \nu N_1 N_2 \left[\int \mathrm{d}y\, \rho_{x1}(x)\rho_{x2}(x-\Delta_x)\right] &= k \int \mathrm{d}\Delta_y\, R(\Delta_x, \Delta_y)\,.\end{aligned} \tag{64}$$

Finally, we can substitute Eq. (64) into Eq. (63) and obtain for the calibration constant

$$k = \frac{\nu N_1 N_2\, R(0,0)}{\left[\int \mathrm{d}\Delta_x\, R(\Delta_x, 0)\right] \times \left[\int \mathrm{d}\Delta_y\, R(0, \Delta_y)\right]}\,. \tag{65}$$

This form enables us to easily derive the absolute calibration from the vdM scans. Since $R(0,0)$ is the maximal rate obtained for no beam displacement with respect to the beam

centroid, one only needs the integral of the beam profile independently in x and y. The integral will be determined either analytically from fitted Gaussian profiles or from a numerical integration over the known rates at the measured displacement values.

To become independent of diminishing particle densities in the beam over the time span of the scans, it is helpful to replace R with $\widetilde{R}$,

$$\widetilde{R} = R/(N_1 N_2)\,. \tag{66}$$

Sometimes the luminosity is also expressed by the geometrical quantity, the effective area, A_{eff}, as

$$\mathcal{L}_{\text{inst}} = \nu N_1 N_2\, R(t)\, A_{\text{eff}}^{-1}\,, \tag{67}$$

with

$$A_{\text{eff}} = \int \mathrm{d}\Delta_x\, d\Delta_y\, R(\Delta_x, \Delta_y) = \frac{\left[\int \mathrm{d}\Delta_x\, R(\Delta_x, 0)\right] \times \left[\int \mathrm{d}\Delta_y\, R(0, \Delta_y)\right]}{R(0,0)} \tag{68}$$

A visualisation of the area is shown in Fig. 53.

4.1.2.3 *Van-der-Meer Measurement of 2013*

Two scans for each beam direction were performed and are referred to as scan 1 and scan 2. In previous LHC vdM scans, the data acquisition with stable beams was stopped and dedicated beam conditions were set up. In this way, for example, afterglow of the detectors can be reduced by decreasing the number of filled bunch slots or the filling of the bunch slots can be adjusted to minimise ghost and satellite charges. For the 2013 data acquisition no such dedicated runs could be afforded due to the required setup time, and the vdM scans were executed during normal data taking conditions. The beams could be displaced at two interaction points at the same time. Consequently, experiments at IP1 and IP5 took calibration data together while IP2 and IP8 took collision data and vice versa. This procedure worked satisfactory and allowed the scans to be completed with minimal loss of collision data. Table 7 shows the two fills during which CMS acquired its vdM data. Additionally, during each fill data for a length scale calibration were taken. For this, both beams were shifted in respect to the nominal position at the same time. The analysis of this data is explained in Appendix A.2.

Table 7: Details about the data acquired for vdM scan fills for proton-lead and lead-proton collisions.

Type	Date	$\sqrt{s_{NN}}$	Number of bunches	LHC fill	CMS run
p-Pb	January 28	5.02 TeV	254	3503	210986
Pb-p	February 7	5.02 TeV	248	3537	211561

One essential ingredient is the number of particles in each beam labelled N_1 and N_2 in previous equations. The quantities are measured with the FBCT and DCCT devices

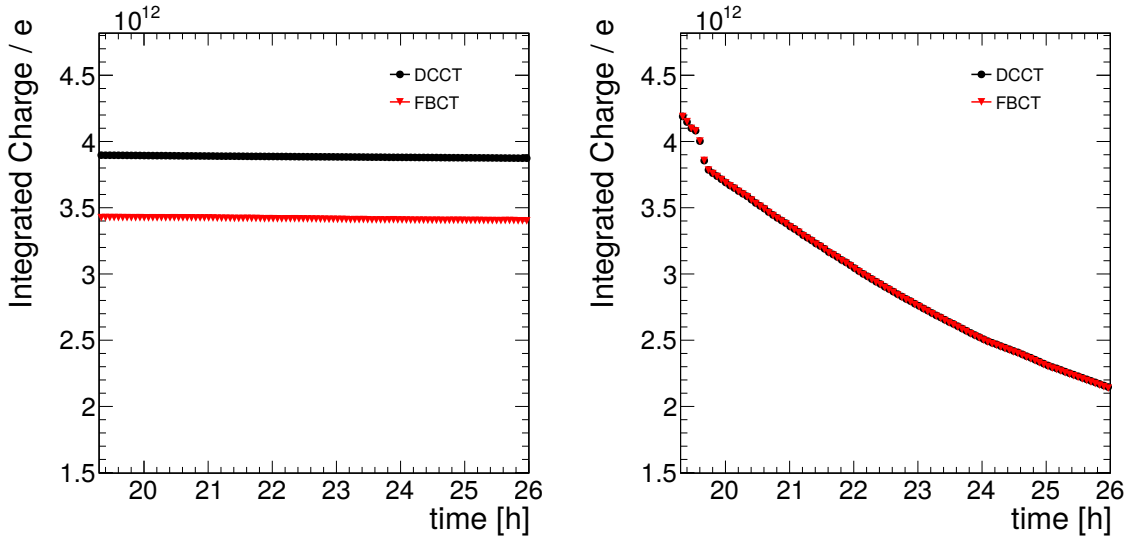


Figure 55: Charge in units of elementary charge integrated over all bunches for the proton beam (left) and lead beam (right). The plots show data from LHC fill 3503 (January p-Pb scan). The x-axis in all plots shows hours after 2013-01-28 00:00:00.

(see Section 3.1.1). The relative bunch-to-bunch charges are measured with the fast FBCT devices. Usually they are calibrated to the DCCT response. It follows that the integrated charge over all bunch slots must be identical to the integrated charge measured with the DCCT devices. For the January scan, the calibration of beam 1 malfunctioned. This faulty calibration is repeated and fixed offline by extracting data from the DIP database[1], where both, FBCT and DCCT currents, are stored together with their corresponding UTC time. The average is calculated over all bunches of FBCT and is summed up over all filled bunch slots including the charges in neighbouring unfilled ones. Figure 55 shows the integrated charges of beam 1 and 2 as measured by the two devices. It is clearly visible that the charge measured before the re-calibration for the protons (beam 1) by FBCT gives a too low intensity.

The ratio is shown in Fig. 56. The weighted average for the correction in % for the period of the vdM scans is determined to be:

1 The Data Interchange Protocol (DIP) database stores global information exchanged between the LHC collider and all its experiments. Among others the charges of the beams as measured by the FBCT and DCCT devices. The data fields storing the charges of in beam 1 are

- DCCT A: LHC.BCTDC.A6R4.B1:BEAM_INTENSITY
- DCCT B: LHC.BCTDC.B6R4.B1:BEAM_INTENSITY
- FBCT: LHC.BCTFR.A6R4.B1:BUNCH_INTENSITY

and their corresponding counterparts for beam 2 (B1→B2). For DCCT there is one device on each side of the interaction point to pick up the charge (referred to as A and B above).

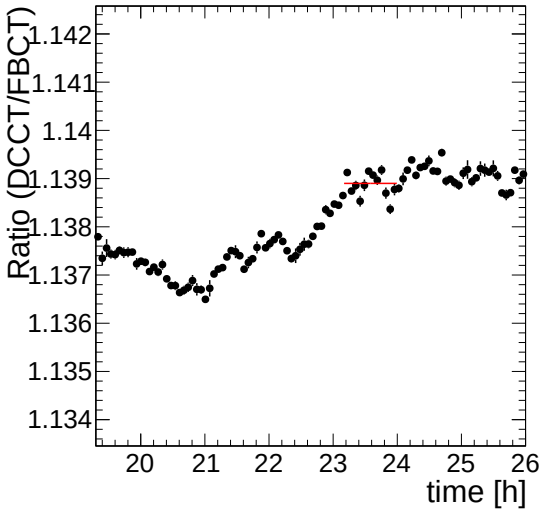


Figure 56: The ratio of the charge measured by the beam monitoring devices DCCT to FBCT. The vertical red line represents the average ratio during the vdM data taking period. The x-axis in all plots shows hours after 2013-01-28 00:00:00.

Beam configuration	**Beam 1**	**Beam 2**
p-Pb (January)	13.89	0.28
Pb-p (February)	0.27	0.14

A 13.9 % correction to the current of beam 1 during the January vdM scans is applied. The correction to the currents of beam 2 is small (<0.3 %) and lies within the systematic uncertainty of the current measurement itself. Hence, no correction is applied.

The p-Pb charge measurement by the FBCT not only had a wrong normalisation, but also the pedestal values were not set correctly during the January scans. In the currents of the first beam a unique pattern in the bunch slots is visible: a pair of bunch slots with high currents alternates with a pair of bunch slots with lower currents. This effect can be corrected since the integrated current is normalised to the DCCT measurement. First the pairs are assigned to two groups depending on their position in the filling scheme. Then, the average current of each group is normalised in a way that does not influence the integrated current over the bunches of both groups. For this the low- and high-current groups are divided by correction factors $\alpha_{\text{low}} = 1.0154$ and $\alpha_{\text{high}} = 0.9852$, respectively. The correction leads to a consistent description of the vdM calibration factor, k, and keeps the normalisation of the integrated charge to the DCCT measurement (more details can be found in [CMS14d]). In the following, values of N_1 and N_2 are always presented with taking all discussed corrections into account.

4.1.2.4 *Measuring and Fitting the Profile Functions*

The profile functions are measured with the HF calorimeter. The advantage over previous analyses at CMS, which were using the pixel detector is a higher visible cross section. This reduces statistical fluctuations in the tails of the profile and is necessary for the shorter scan periods of 2013.

The average fraction of empty towers of HF, p_0, is related to the average number of interactions or pileup, μ, by

$$p_0 = e^{-b\mu}\,, \tag{69}$$

where b is the probability for a tower to be non-empty, i. e. to see a signal in a collision. The average interaction rate is then

$$R(\Delta_x, \Delta_y) = \text{const} \times (-\log(p_0))\,. \tag{70}$$

To maximise linearity only two 2π-rings in $3.5 < |\eta| < 4.2$ of the HF detector are used. For low luminosities the fraction of empty towers is always $p_0 \gg 0$ allowing to neglect any non-liner terms. The rate is measured for displacements in x and y from $-0.15\,\text{mm}$ to $+0.15\,\text{mm}$.

For p-Pb collisions the beam profile turns out to be more difficult to describe than for p-p collision. It is useful to introduce "supergaussian" functions that can be successfully fitted for more rectangular-like beam profiles, i. e. a profile that decreases at a slower rate around $(\Delta_x, \Delta_y) = (0, 0)$ than a Gaussian profile [Dec94]. The supergaussian has an exponential term known from the Gaussian distribution and one of higher order

$$\cdots \exp\left[-\left(\frac{x-\mu}{2\sigma_1}\right)^2\right] + \cdots \exp\left[-\left(\frac{x-\mu}{2\sigma_2}\right)^{2+\epsilon}\right]\,. \tag{71}$$

The adjustment of the power using a positive ϵ has been shown in the same reference to be able to describe the more rectangular-like shape. For $\epsilon = 2$, reasonable fits to the p-Pb beams profiles are produced. The exact equation is then

$$f_{\text{SG}}(x) = A\left[h\exp\left(-\frac{(x-\mu)^2}{2\sigma_1^2}\right) + \frac{2^{3/4}\sqrt{2\pi}}{\Gamma(0.25)}(1-h)\exp\left(-\frac{(x-\mu)^4}{2\sigma_2^4}\right)\right] \tag{72}$$

with

$$\sigma_1 = \frac{\Sigma r}{h r + 1 - h} \qquad \text{and} \qquad \sigma_2 = \frac{\Sigma}{h r + 1 - h}, \tag{73}$$

The normalisation is chosen such that its integral corresponds to

$$2\pi A\sqrt{\sigma_1^2 + \sigma_2^2}\,. \tag{74}$$

The supergaussian is then fitted to the data for each colliding bunch pair. The results for typical bunch pairs together with the residuals of the fit are shown in Fig. 57. The two selected bunch crossings can be fitted with $\chi^2/\text{ndf} = 27$ and 8 for x and y, respectively. The

average fit quality is roughly similar in terms of scan periods and coordinates. On average, χ^2/ndf is about 15 for all scans except for the displacements in x of the p-Pb beam direction, which is about 25. These are rather large values but are by far better than the fit quality of a Gaussian+const function as shown in Fig. 58. The relative residuals for the supergaussian function (bottom) are much closer to unity over a wider area in displacement for both x and y. The Gaussian function (top) first underestimates at $\pm 0.05\,\text{mm}$ and overestimates at $\pm 0.1\,\text{mm}$ by more than 20 %. The large χ^2 values are also induced in the region of small displacements where many collisions take place and the statistical uncertainty on the rate is very small.

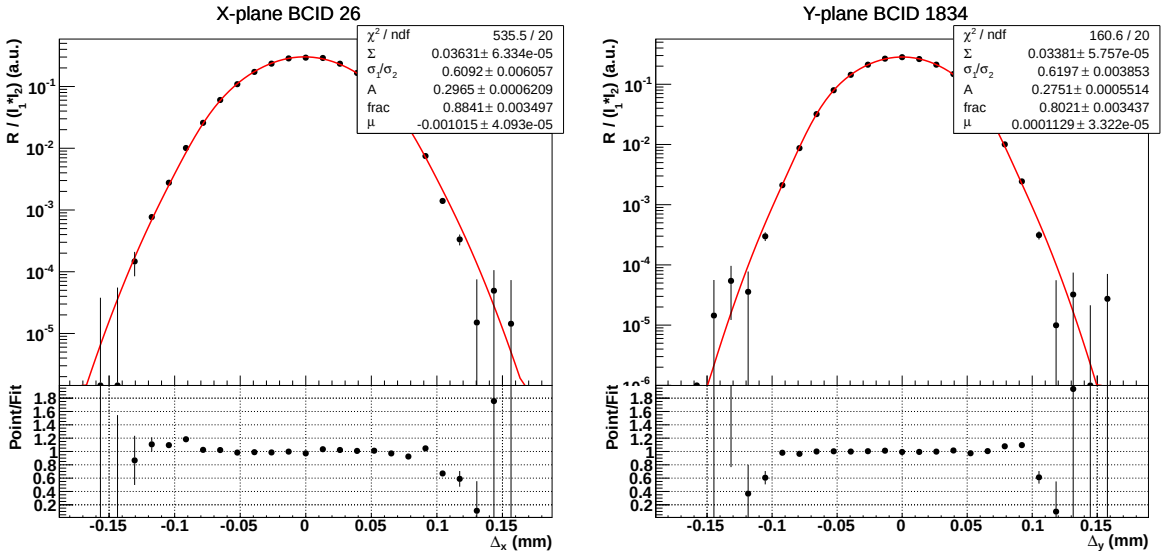


Figure 57: Measurement and fits of the p-Pb (first scan) profile function in x (left panel) and y (right panel). Example profiles are shown for two bunch slots with ID 26 and 1834. The red line shows the fit with the supergaussian function, Eq. (72).

The peak rate is obtained from the fits

$$\widetilde{R}(0,0) = \frac{1}{2}\left[\widetilde{R}_x(0,0) + \widetilde{R}_y(0,0)\right] . \tag{75}$$

Combined with the integral of the supergaussian, the calibration factor can be calculated. Since the value of the factor is meaningless and only the integrated luminosity of the whole data set is relevant, we will in the end quote k as "how much the luminosity changed compared to the previous estimate, k_{old}".

4.1.2.5 *Corrections and Uncertainties*

Multiple corrections related to the measurement of the beam charges have to be applied to achieve an accurate measurement with small systematic uncertainty. From the FBCT calibration, we obtain a 0.3 % uncertainty, which is determined from the difference of the

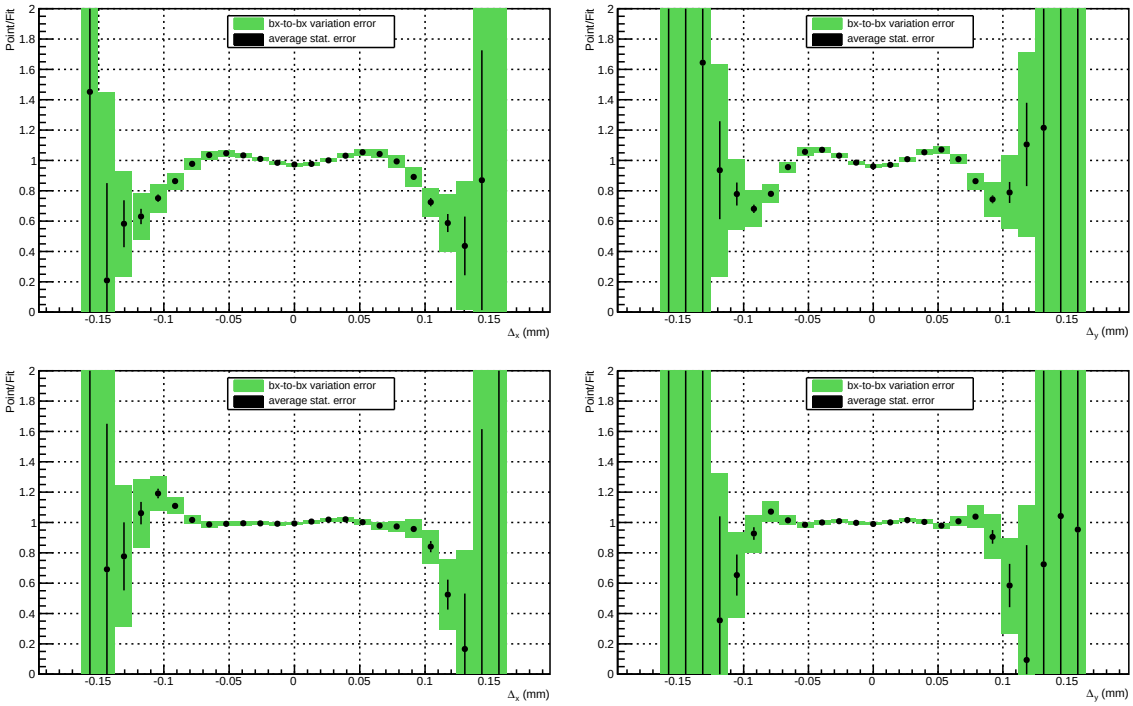


Figure 58: Comparison between Gaussian (top) and supergaussian (bottom) fit functions. The plots correspond to the lower part of each plot in Fig. 57 and show the point/fit ratios for the second Pb-p scan fitted with a Gaussian function (top). The point/fit ratios are averaged over all bunch crossings.

integrated charged visible in the previous section. This takes the recalibration of beam 1 for the January scan into account. The ghost fractions were measured by the LHCb experiment with the beam-gas method [Bar14]. The values are listed in Table 8. The ghost charges are subtracted from the current values.

The satellite and ghost fraction were measured by the LHC LDM group [Boc+13] in two scans, once at the beginning and once at the end of the run. The fraction of charges in the wrong buckets is shown in Fig. 59. The yellow and the red distributions, which cover the p-Pb and the Pb-p measurements, have a mean of 0.001 and 0.002, respectively. This is even smaller than the ghost fraction and both effects combined introduce only a small systematic uncertainty, conservatively estimated to be 0.1 %.

Coulomb fields lead to two corrections. The first correction is due to beam-beam effects. They describe the deflection one beam experiences in the electromagnetic field of the other beam. The deflection is calculated using the closed form of the Bassetti-Eskine

equations [BE80], that depend on the beam widths σ_x and σ_y. The two components of the electric fields, E_x and E_y are given by

$$E_{x/y} = \frac{Q}{2\epsilon_0\sqrt{2\pi\left(\sigma_x^2-\sigma_y^2\right)}} \times \mathrm{Re/Im}\left[w\left(\frac{x+iy}{\sqrt{2\left(\sigma_x^2-\sigma_y^2\right)}}\right) - \exp\left(-\frac{x^2}{2\sigma_x^2}+\frac{y^2}{2\sigma_y^2}\right) w\left(\frac{x\frac{\sigma_y}{\sigma_x}+iy\frac{\sigma_x}{\sigma_y}}{\sqrt{2\left(\sigma_x^2-\sigma_y^2\right)}}\right)\right], \tag{76}$$

where $w(z)$ is the complex error function, ϵ_0 is the vacuum permittivity and all other variables are known from previous equations. For the x-direction, the real part, and for y-direction, the imaginary part of the term in square brackets is needed. The deflection can then be calculated as

$$\delta_{x,y} = K\,N\,E_{x,y} \tag{77}$$

with $K = 2\,r_p\,\gamma$ and the radius of the proton, r_p, and the number of particles in the opposite bunch, N. The result of the calculation is shown in Fig. 60. The correction is on the order of 0.1 µm.

The second effect related to Coulomb fields is the dynamic evolution of the betatron function. This relates to an intrinsic repulsion of the beam particles causing a defocusing with time. This effect is investigated in [CMS14d] for p-p beams and a correction of 0.2 % is found. The effect scales just as the beam-beam deflection from p-p to p-Pb. Beam-beam deflections for p-Pb beams are 2 orders of magnitude below the deflection for p-p beams of 2013 due to the different centre-of-mass energies. It follows that the dynamic-β effect is negligible for p-Pb beams.

Finally, the estimations of the beam separations Δ_x and Δ_y are based on knowledge of magnetic fields that are responsible for beam displacements. This procedure is, however, not accurate enough and can be verified with the CMS tracker and its vertex reconstruction. Special data were taken where both beams are shifted simultaneously in one direction, which also displaces the vertex. The analysis procedure is described in detail in Appendix A.2. The correction factors that are derived from comparing the length scale of the LHC magnets and the shift in the vertex position are

$$c_{\mathrm{LS,x}} = 0.984 \pm 0.009 \tag{78}$$

$$c_{\mathrm{LS,y}} = 0.992 \pm 0.007\,. \tag{79}$$

Here, the systematic and statistical uncertainties of the length scale measurements are combined by adding them in quadrature.

The effects of all previously discussed corrections on the luminosity calibration are summarised in Table 9. The table also contains uncertainties from statistics, bunch-to-bunch variations, and the fit model. The statistical uncertainty (from fitted parameters) is 0.7 %, and additional uncertainty from bunch-to-bunch variation arise from taking the average

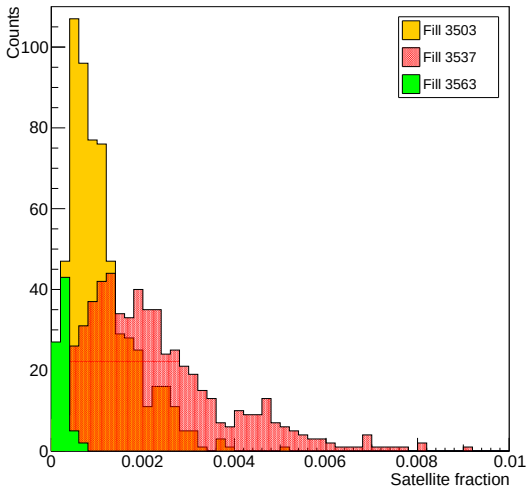


Figure 59: Distributions of ratio of satellite charges to the nominally filled RF-bucket charge for both beams. LDM measurements for three fills are shown: 3503 (p-Pb), 3537 (Pb-p), and 3563 (p-p).

over all per-bunch-pair normalisations. The uncertainty on the fit model is derived from the difference between obtaining the area of the profile with numerical integration over the bins and the supergaussian fits.

Table 8: Ghost fractions as measured by the LHCb experiment. In the last column, statistical and systematic uncertainties are added in quadrature.

LHC fill	Beam	Ghost fraction	Uncertainty
3503	1	0.002	7.1 %
	2	0.005	3.5 %
3537	1	0.005	6.9 %
	2	0.009	3.2 %

4.1.2.6 *Results on the Luminosity Calibration*

The luminosity calibration change, k/k_{old} can be easily calculated by using the rate R(0,0) and the integral of the supergaussian fits and substituting the values in Eq. (65). The normalisation change is shown in Fig. 62 for each bunch crossing for p-Pb and Pb-p. Averaged over all bunch crossings it is

$$\text{p-Pb}: \; k/k_{\text{old}} = 1.140 \pm 0.041 \tag{80}$$

$$\text{Pb-p}: \; k/k_{\text{old}} = 1.138 \pm 0.039\,. \tag{81}$$

The uncertainties are taken to be fully correlated, thus, we take the mean of the central value and the systematic uncertainty as a combined result

$$\text{combined}: \; k/k_{\text{old}} = 1.139 \pm 0.040\,. \tag{82}$$

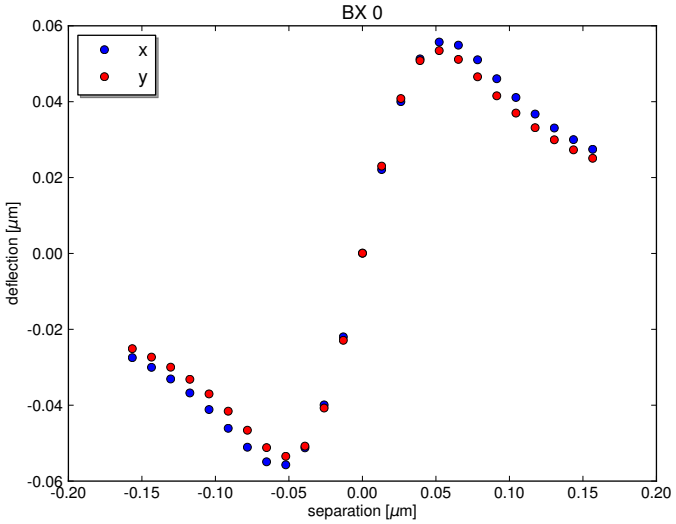


Figure 60: The impact of the beam-beam deflection effect. The plot shows the displacement due to deflection as a function of the beam separation for the first bunch crossing.

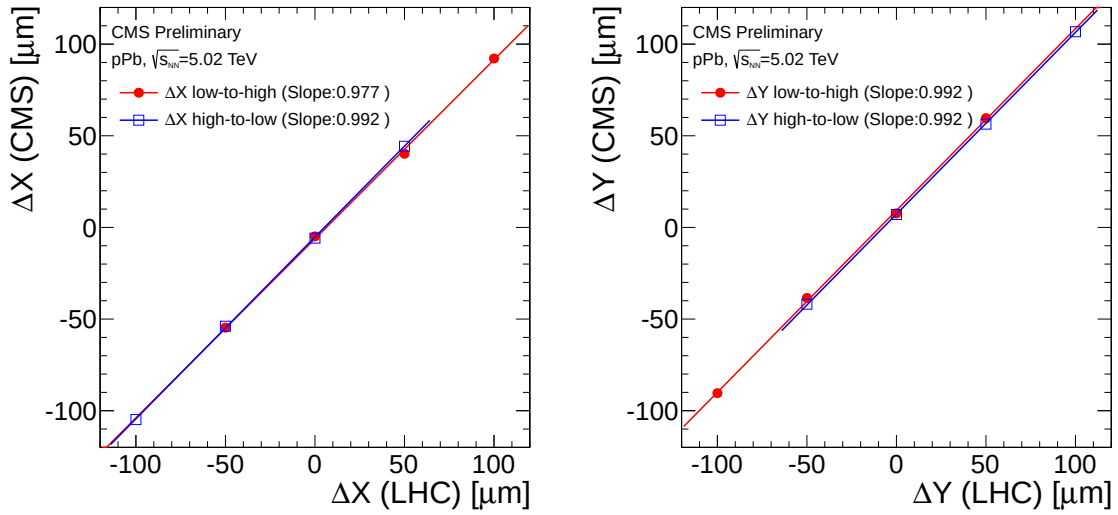


Figure 61: Fitted vertex position compared to LHC reference derived from the magnet settings. The red and blue lines are the first-order polynomial fits for the separation values divided into two subsets as defined in the text.

Table 9: Uncertainties on normalisation changes for p-Pb and Pb-p vdM scans. The table includes uncertainties arising from corrections. In row "Total", errors are added in quadrature.

Uncertainty	**p-Pb**	**Pb-p**
Statistical	0.7 %	0.7 %
Bunch-to-bunch	1.3 %	0.8 %
Fit model	3 %	3 %
Length scale	1.2 %	1.2 %
Ghosts and satellites	<0.1 %	<0.1 %
Beam current calibration	0.3 %	0.3 %
Beam-beam deflection	0.3 %	0.3 %
Afterglow	negligible	negligible
Dynamic β	negligible	negligible
Total	3.6 %	3.4 %

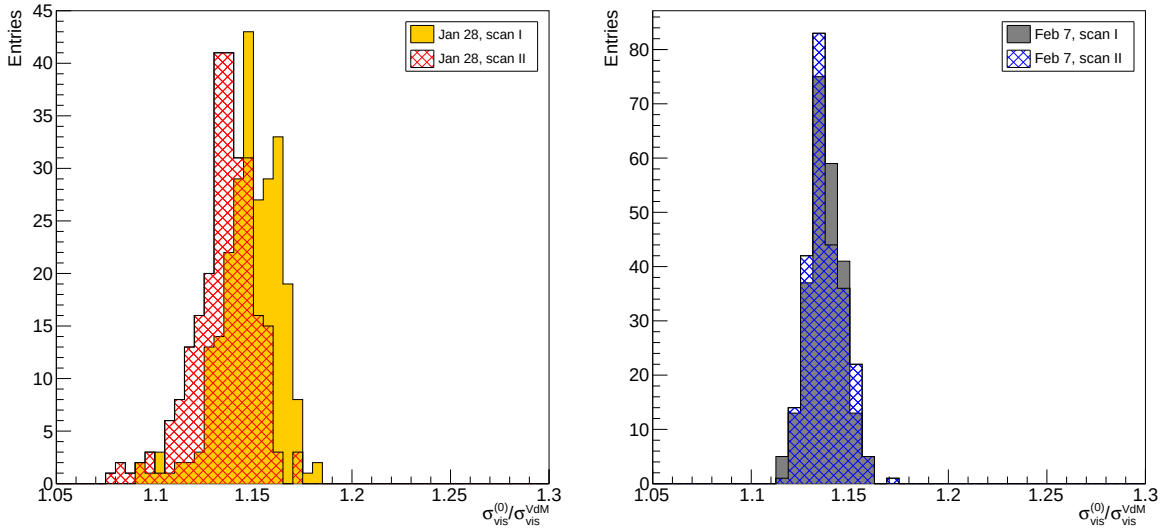


Figure 62: Distributions of normalisation changes for the p-Pb (left panel) and Pb-p (right panel) vdM scans.

The recorded integrated luminosity by CMS in 2013 for p-Pb [Pb-p] collisions is therefore (20.7 ± 0.7) nb^{-1} $\left[(14.0 \pm 0.5)\ \text{nb}^{-1}\right]$.

4.1.3 DETERMINING THE NUMBER OF COLLISIONS WITH THE HF DETECTOR

In collisions with particle production, secondary particles can be detected with a multitude of detectors of the CMS experiment. In the dominating case of non-diffractive collisions between protons and lead ions, many tracks are produced and only small rapidity gaps between the particles exist. The central detector can identify such events as inelastic with a high efficiency.

In case of a pomeron exchange, the diffractive processes $I\!P + p$ and $\mathrm{Pb} + I\!P$ can be induced. Dissociation of the projectile and/or the target (double/single diffraction) is possible. Furthermore, light mesons or heavier objects like jets can be produced in double pomeron exchange, which are interactions with very low cross sections [Alb11]. In the following, this is referred to as central diffraction. For measuring the production cross section, non-, single-, double-, and central-diffractive collisions have to be taken into account. A depiction of a single-diffractive collision at the CMS experiment is shown in Fig. 63. Here, the dissociating side produces particles that are only measurable with forward detectors. For smaller diffractive masses, even the indicated HF calorimeters can no longer detect these particles; the secondaries leave through the hole left by the beam pipe in the almost hermetic detector.

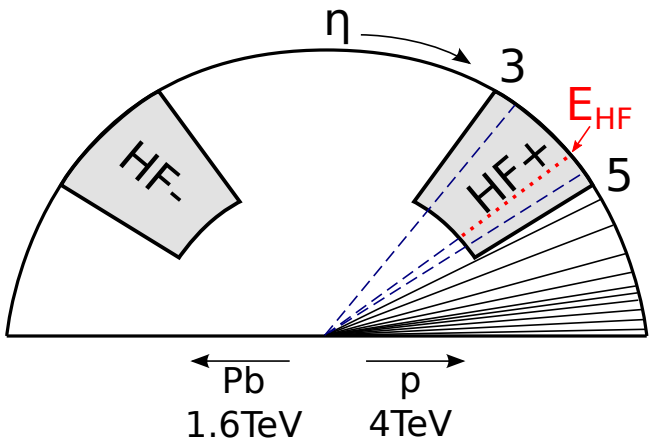


Figure 63: Illustration of a single-diffractive event where the proton dissociates and three secondaries (dashed lines) are produced within the acceptance of the HF calorimeter. The particle with the highest energy (dotted line) determines the quantity E_{HF}.

Measuring diffractive collisions is important even though their part of the cross section only accounts for $\simeq$10 % of the production cross section. An argument stems from the elusiveness of diffractive events. Non-diffractive collisions are better understood, both theoretically as well as experimentally. However, any measurement at a hadron collider is always contaminated by diffractive events. A poorly understood diffractive component can induce significant uncertainties. We evaluated the efficiency to detect non-diffractive and diffractive collisions with the central tracker, the hadronic forward calorimeter and the

CASTOR detector and the results are shown in Fig. 64. The left plot shows that forward detectors have a high efficiency to select inelastic events. The HF and the CASTOR detectors can reach efficiencies up to 95 % to 96 % based on the EPOS and QGSJET generators. Interestingly, also events that have only one reconstructed central track in the pixel tracker make up about 94 % of the diffractive and non-diffractive component. For the HIJING generator, diffraction is not implemented properly and the results determined with this generator can be regarded as only the non-diffractive component of the production cross section. For this reason, the efficiencies artificially seem to be higher. More importantly, the difference between the other two generators and HIJING reveals the acceptance for diffractively produced events. The tracker can pick up the least amount of diffractive events, the HF detector can pick up more, and CASTOR has the best acceptance for diffractive events.

If one would add the tracker to a selection based on HF, no improvements for diffractive collisions are expected (middle and right plot) and the CASTOR detector would recover about 1 % of the acceptance for diffractive events. In this thesis, CASTOR is used to evaluate the performance of the generators in terms of their modelling of diffraction and to constrain the acceptance extrapolation to the full phase-space. The CASTOR detector is not used directly for the event selection, as it is installed only on one z-side of the CMS detector.

In the following, we will employ the two HF detectors in both hemispheres in pseudorapidity to identify and count inelastic events. We will do so with two different event selections: not requiring (**single-arm**) and requiring coincidence (**double-arm**) of the signal measured in each detector. A minimum energy E_{HF} is required to count an event as an inelastic collision candidate. E_{HF} is determined using the highest energy in a tower of the HF calorimeter positioned at positive and negative pseudorapidity, $E_{\mathrm{HF}+}$ and $E_{\mathrm{HF}-}$. For the two selections, E_{HF} is then defined as

$$E_{\mathrm{HF}} = \begin{cases} \max(E_{\mathrm{HF}+}, E_{\mathrm{HF}-}) & \text{for single-arm selection} \\ \min(E_{\mathrm{HF}+}, E_{\mathrm{HF}-}) & \text{for double-arm selection .} \end{cases} \tag{83}$$

The double-arm selection has a larger capability to reduce background events, whereas the single-arm selection has a higher acceptance to select inelastic and especially single-diffractive events.

Simulation and Detector Performance

Proton-lead collisions need to be simulated to determine the efficiency to select inelastic events with particle production. The events are generated by hadronic (Section 3.3) and photo-nuclear interaction models (Section 3.3.0.5). The models DPMJET3.06, EPOS-LHC, HIJING1.383, and QGSJETII-04 are used to generate proton-nucleus collisions. The STARLIGHT model is combined with the PYTHIA and DPMJET frameworks to simulate photo-nuclear ($\gamma^* + \text{parton} \rightarrow X$) collisions, where the photon is part of the electric field of the lead nucleus. The detector simulation of the HF detector is performed with the GEANT4 framework and the GFlash extension (Section 3.2.2). The generation and simulation steps, the digitisation of the signal, and reconstruction of the events (Section 3.2.2)

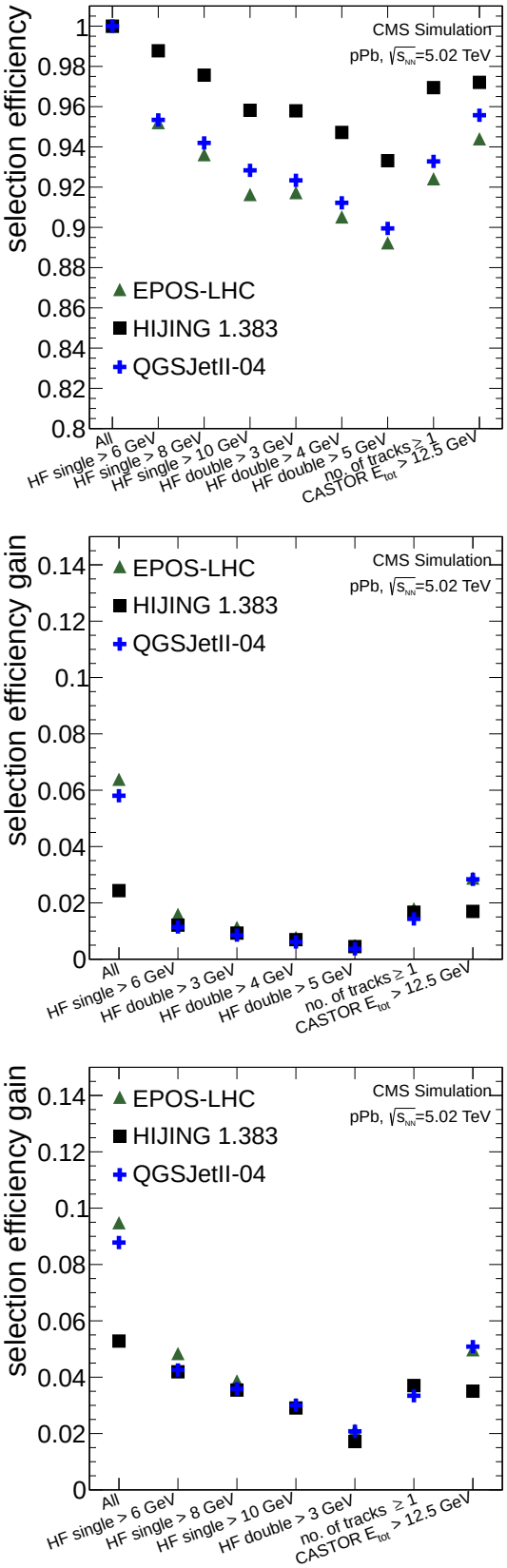


Figure 64: Top panel: Efficiency of various detectors and trigger thresholds. The values are determined from fully reconstructed p-Pb collisions at $\sqrt{s_{NN}} = 5.02\,\text{TeV}$ of the CMS detector simulation. Different thresholds, E_{th} for the single-arm and the double-arm cases are shown. The efficiency is also given for the requirement of 1 reconstructed track in the central tracker. The trigger based on CASTOR requires the total energy in the detector to be above 12.5 GeV, which corresponds to 4 pedestal widths in the first 5 modules.
Middle and bottom panel: The increase in efficiency in case other selections are combined with a logical AND is shown for single-arm $E_{\text{HF}} > 8\,\text{GeV}$ (middle) and double-arm $E_{\text{HF}} > 4\,\text{GeV}$ (bottom) selections. For all plots the label *All* corresponds to a perfect hermetic detector. For the middle and bottom plots, this is equivalent to the maximum possible efficiency gain.

run within the CMSSW framework. For each model about 100 000 events were simulated with the exception of EPOS and HIJING, for which centralised productions with 5 million events are available. To evaluate the performance of the detector in terms of electronic noise and the simulated signals, we compare the reconstructed energy to the measured energy in 4.7×10^6 events recorded by an unbiased trigger. The unbiased events were recorded at a constant rate of 20 Hz for which the only requirement is the presence of both beams. The interaction probability, λ, is about 5 % and the events can therefore have no, exactly one, or more collisions taking place[2]. The additional collisions of the latter type are called *pileup* collisions. Of interest to us are only events with collisions, and they are selected by requiring a threshold energy, E_{th}, fulfilling $E_{\text{HF}} > E_{\text{th}}$.

When comparing the simulation to data from 2013, the signal has to be adjusted for radiation damage that the fibres and PMTs accumulated during prior data taking periods. In the approach employed by this analysis, the signal of the simulation is rescaled to describe that of data with radiation damage. This is an acceptable measure since the HF detector is not used to actually measure the energy deposit, E_{HF}, but instead is used to select events based on, E_{HF}. The radiation damage is dependent on the pseudorapidity coverage of each ring of HF because the deposited energy increases with $|\eta|$. The scaling parameters are therefore obtained ring-wise by normalising the summed energy over towers averaged over the values of the interaction models to that of data. Events with collisions only are enforced by requiring one track with $p_{\text{T}} > 0.1\,\text{GeV}$ to be reconstructed within $|\eta| < 2.5$. This selection does not bias the HF response and has low contribution from γp collisions. The response of data lies below that of simulation as is shown in the top panel of Fig. 65. The rescaling factors are calculated from the total energy as

$$C_i = \langle E^{\text{MC}}_{\text{tot},i} \rangle / E^{\text{data}}_{\text{tot},i} \tag{84}$$

in each $i\eta$-ring. The rings are defined by the detector geometry and their pseudorapidity coverage is listed in Table 3.

An independent measurement of the radiation damage exists, where the energy response of p-p collision data at $\sqrt{s} = 2.76\,\text{GeV}$ from 2013 is compared to data from 2011. The radiation damage in the earlier period is negligible. These values are only obtained for long fibres of HF and cannot be used to correct the energy of a tower that is a combination the signal in long and short fibres. The values agree, nevertheless, which implies that radiation damage has similar impact in both long and short fibres. The values are compared in the lower panel of Fig. 65 and listed in Table 16 in Appendix A.3. For this analysis the rescaling factors from Eq. (84) are applied.

Figure 66 compares the quantity E_{HF} as determined in simulation and data. Here, the background that will be discussed in detail later is shown as a colour-filled area and quickly dies out, and hadronic collisions (coloured markers) dominate the number of events with E_{HF} above 8 GeV (4 GeV) for the single-arm (double-arm) selection. At low energies left of the dashed line, events stem from detector noise and photo-nuclear collisions and only a few hadronic events fall into this region.

2 The distribution of the number of collisions can be calculated with Eq. (56).

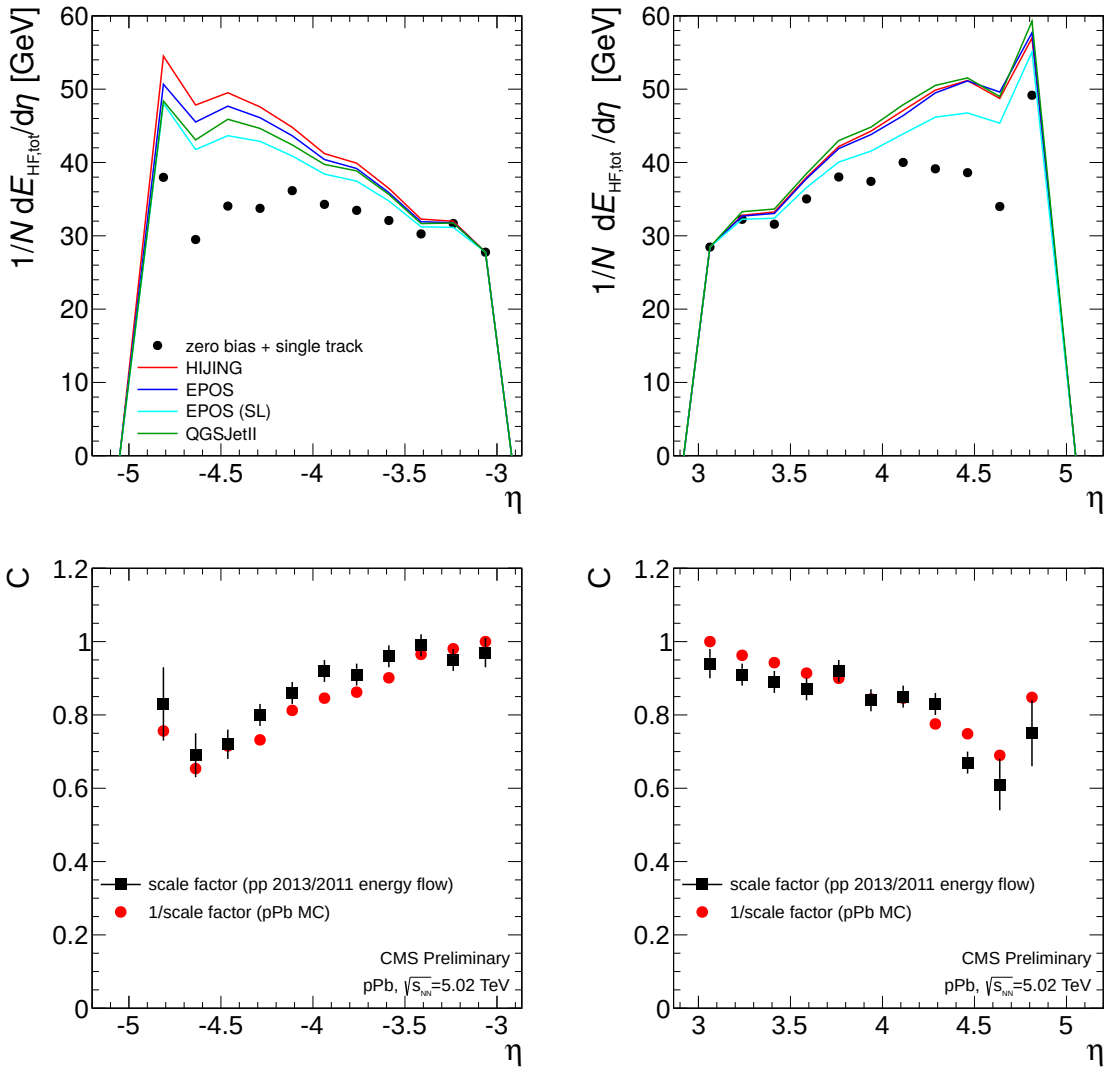


Figure 65: Top panels: Average energy in HF rings for the unbiased trigger requiring a single track from data compared to three Monte Carlo simulations and EPOS simulated with shower library (SL) instead of GFlash. Left panels: For HF on $\eta-$ side. Right panels: For HF on $\eta+$ side. The energy is normalised to the innermost $i\eta$-ring where the agreement is good. This corresponds to scaling up the signal in simulations by $\simeq$14 %.
Bottom panels: Scaling factors for HF derived from the p-p energy flow calibration compared to MC rescaling as it is used in this analysis.

Since the correction is applied ring-wise, it is of interest how the distribution of E_{HF} changes by this correction in each pseudorapidity region. The comparison is found in Figs. 100 to 103 of the Appendix and a good agreement is found for all rings. Only a few $i\eta$-rings (mainly 39 and -30) show slight disagreement but are still bracketed by the distributions of the different generators. Also the simulated distribution of all tower energies (not only the highest energy as in Fig. 66) shows good agreement for energies above 2 GeV when the rescaling factors are applied (Fig. 102).

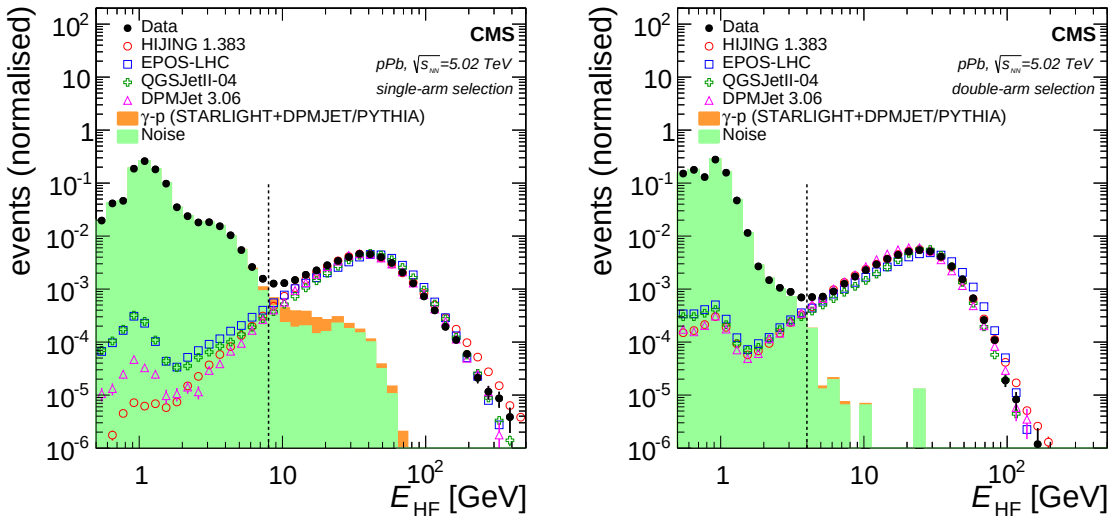


Figure 66: Distribution of E_{HF}, where data are recorded using the unbiased trigger. The contribution from noise is obtained from a random trigger normalised to the same number of triggers as that in the collision data. The average number of γp processes simulated with STARLIGHT+DPMJET and STARLIGHT+PYTHIA is indicated as background. Four hadronic interaction models (EPOS, DPMJET, HIJING, and QGSJET) are overlaid and normalised to the number of data events above 10 GeV, where the contribution from the background is small. The vertical line represents the threshold energy of 8 GeV used in this analysis.

An accurate efficiency can only be determined when the rescaling also works to improve the description of the probability of an $i\eta$-ring to contain the tower with the highest energy, which consequently determines E_{HF}. Taking the rescaling factors into account the simulation agrees well with the collisions in data and even the dip at $i\eta = \pm 39$ due to a large material budget in front of the detector (beam pipe projection) is described accurately (cf. Fig. 67).

Background Estimation and Selection Efficiency

Electronic noise and beam gas collisions lead to misidentified events in the sample of inelastic collision candidates. These events must be subtracted from the number of events

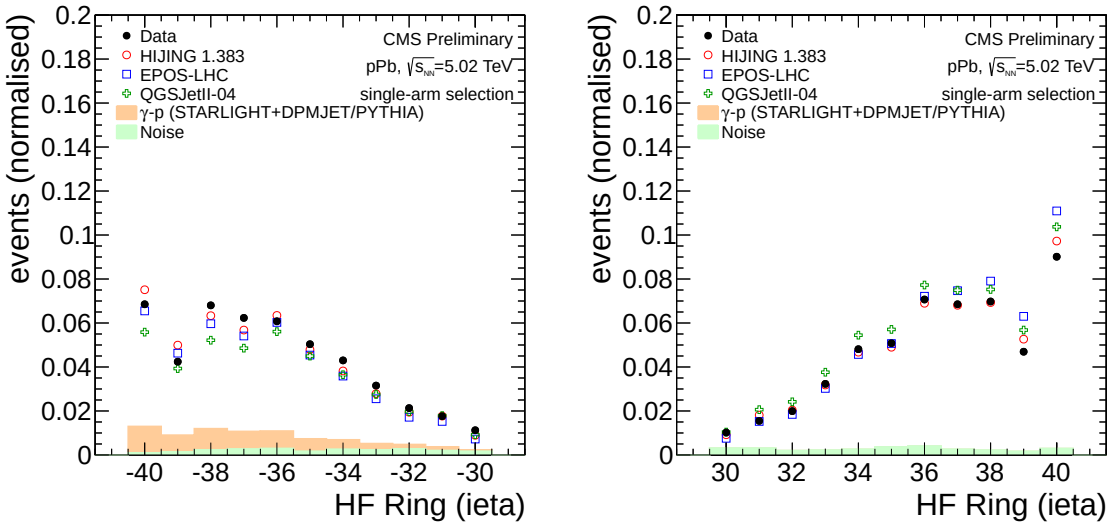


Figure 67: The plot shows the probability that an $i\eta$-ring contains the tower with the highest energy. Here, the single-arm selection from data is compared to simulated events with three different generators. Their distribution is added to the background contributions of noise and γp collisions.

recorded with the unbiased trigger, N. With the fraction of randomly read-out no-beam events that pass the selection criteria, f_{noise}, their number is given by

$$N_{\text{noise}} = N\, f_{\text{noise}}\,. \tag{85}$$

The value of f_{noise} is estimated from events recorded by a random trigger that selects events after random time intervals and requiring the absence of both beams as determined by the "Beam Pickup Timing for the eXperiments" devices. Events are recorded with 20 Hz and provide enough statistics to study the detector noise. Additionally, there are events where only one beam is present and that provide the opportunity to study beam related background, e. g., unanticipated collisions with the remaining gas in the beam pipe. The corresponding bunch crossings are dependent on the filling scheme and are selected by their appropriate bunch crossing numbers (half filled red markers in Fig. 68). In Fig. 69 the signal in the HF detectors is shown for both, the no-beam and the one-beam contributions. Within statistical uncertainties no differences are visible, suggesting a small contribution from beam-related background. The noise is therefore estimated with the no-beam events, where larger statistics is available. The electronic noise is not included in the detector description of the simulations.

The estimate of N_{noise} includes $N_{\text{vis+noise}} = N_{\text{vis}} f_{\text{noise}}$ events that contain both, a significant signal from noise but also a visible inelastic collision, where N_{vis} is the number of visible events with particle production. This is corrected for by explicitly subtracting $N_{\text{vis+noise}}$ from N_{noise}.

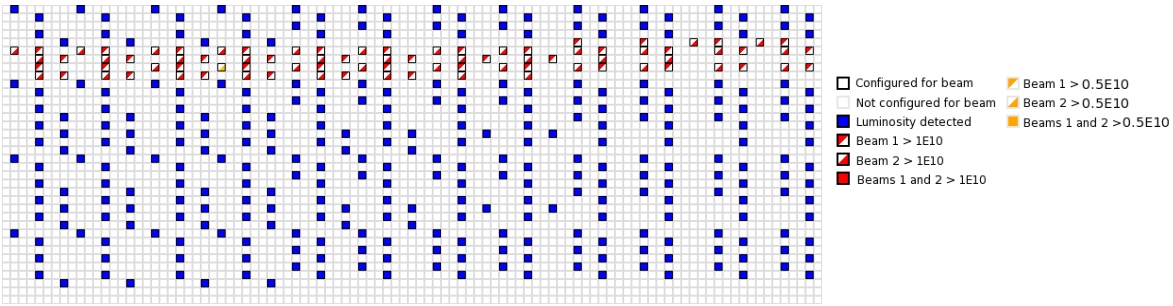


Figure 68: The beam filling scheme for 296 bunch pairs (filled blue). Each bunch slot is represented by a marker for each bunch crossing (bunch crossing 0 to bunch crossing 3563 where the first line represents 0–98). This scheme is used, e. g., for LHC fill number 3496.

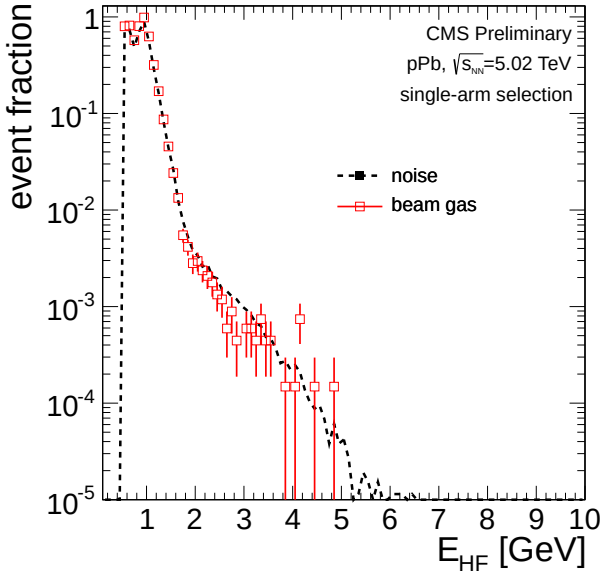


Figure 69: Distribution of signal in HF for two types of events where the bunches do not collide. The single-beam distribution is normalised to the first 5 bins of the no-beam distribution.

Figure 72 (top panel) shows the number of events divided by luminosity for noise events (black line) for different thresholds of E_{HF}. The quantity drops faster for the double-arm selection because it is less likely to have high-noise signals in coincidence in both HF calorimeters.

Another background component comes from photo-nuclear collisions. At ultra-peripheral impact parameters, where hadronic interactions become unlikely, interactions of the protons with the electromagnetic field of the lead ion dominate (Section 3.3.0.5). The processes involving two photons, for which purely leptonic final states are produced, have a cross section too low to be important. The dominant process for this case produces an electron-positron pair. Its cross section is estimated by a simulation with STARLIGHT and is less

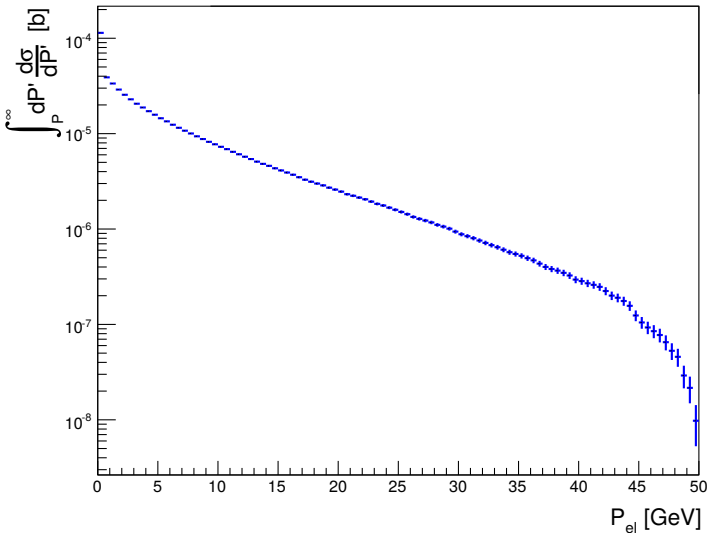


Figure 70: Integrated cross section for process $\gamma + \gamma \to e^+ + e^-$ in p-Pb collisions passing a selection that corresponds to the single-arm selection based on the HF detector. The highest energetic $e^\pm$ in $3 < |\eta| < 5$ must have at least the momentum P_{el}. The processes are simulated with STARLIGHT with an invariant mass of the two photons between $1 < W/\text{GeV} < 50$.

than 10 µb for electrons with momenta larger than 7 GeV that hit one of the HF detectors as is shown in Fig. 70. On the other hand, the processes that cannot be neglected are shown in Fig. 71; these are the production of vector mesons (left) and the dissociation of the proton (right). The background contribution due to energy deposit in the HF detectors is estimated for both types using simulations. The outcome is shown in Fig. 72 (top panel) as the red solid and blue dashed line for the frameworks that treat the hadronic part of the collisions, DPMJET and PYTHIA, respectively. The contribution for the double-arm selection is about 2 orders of magnitude lower and is only slowly decreasing for the single-arm election but staying below 0.1 mb for the relevant energy thresholds. The observed number of γp collisions is $N_{\text{em}} = f_{\text{em}}\,\sigma_{\text{em}}\,\mathcal{L}$, where f_{em} is the fraction of simulated photo-nuclear events passing the selection and σ_{em} is the predicted cross section by the STARLIGHT program. Due to a different minimum photon energy in the simulations, the cross sections for the two samples differ. The cross section is $\sigma_{\text{em}} = 195\,\text{mb}$ $(122\,\text{mb})$ for STARLIGHT+DPMJET (STARLIGHT+PYTHIA). The larger cross section of the former sample is the consequence of a smaller value of the minimal sampled photon energy. This is visible as an excess at very small values of E_{HF}. Subtracting photo-nuclear events is consistent with the purpose to measure a cross section most relevant for the modelling of extensive air showers. The reason is related to the different target particles. The charge of the lead ion is $Z = 82$ but in air, mostly nitrogen with charge $Z = 7$ is responsible for the electromagnetic field. According to Eq. (46) the photon density scales with $\propto Z^2$ and consequently it will be suppressed by a factor of $7^2/82^2$ in p-Air with respect to p-Pb. Taking the cross section estimated by STARLIGHT+DPMJET results in a contribution of 1.4 mb to the proton-air cross section, which is far below 1 %.

The purity of the event selection is defined by the amount of contamination with the background as $p = N_{\text{prod}} / \left(N_{\text{prod}} + N_{\text{em}} + N_{\text{noise}}\right)$.

Figure 71: Examples of ultra-peripheral proton-nucleus interactions. Left panel: A quasi-real photon from the field of the lead ion fluctuates into a neutral ρ meson (circle) which interacts with the incoming proton via a colour-singlet state (pomeron exchange) and decays into π^+ and π^- mesons. Right panel: A quasi-real photon γ^* interacts with a parton leading to the break-up of the proton.

The efficiency within the acceptance of HF is defined as the fraction of events with hadronic particle production that are detectable, $\epsilon_{\text{acc}} = N_{\text{vis,had}}/N_{\text{prod}}$. The dependence on the energy threshold, E_{th}, used for the event selection is shown in the bottom panel of Fig. 72. To specify the selection thresholds the efficiency and purity are compared in Fig. 73. The chosen optimal value of E_{th} is denoted with a square marker. Since the two event selections have intrinsic differences in their acceptance to physics processes, we obtain a high efficiency in the single-arm case and a high purity for the double-arm case. The thresholds are $E_{\text{th}} = 8\,\text{GeV}$ and $E_{\text{th}} = 4\,\text{GeV}$, respectively, for the single-arm and double-arm selection.

For these working points the background contributions and efficiency are clearly defined. The value ϵ_{acc} for the chosen threshold is determined by averaging over the results of the models EPOS and QGSJET. The results of HIJING, which does not include nuclear effects for diffraction, and DPMJET, which does not contain coherent diffraction, are not considered for this average. The value is $\epsilon_{\text{acc}} = 0.939 \pm 0.005$ (0.910 ± 0.014) for the single-arm (double-arm) selection.

Similarly, N_{em} is determined by averaging over the same interval for the used generators STARLIGHT+DPMJET and STARLIGHT+PYTHIA.

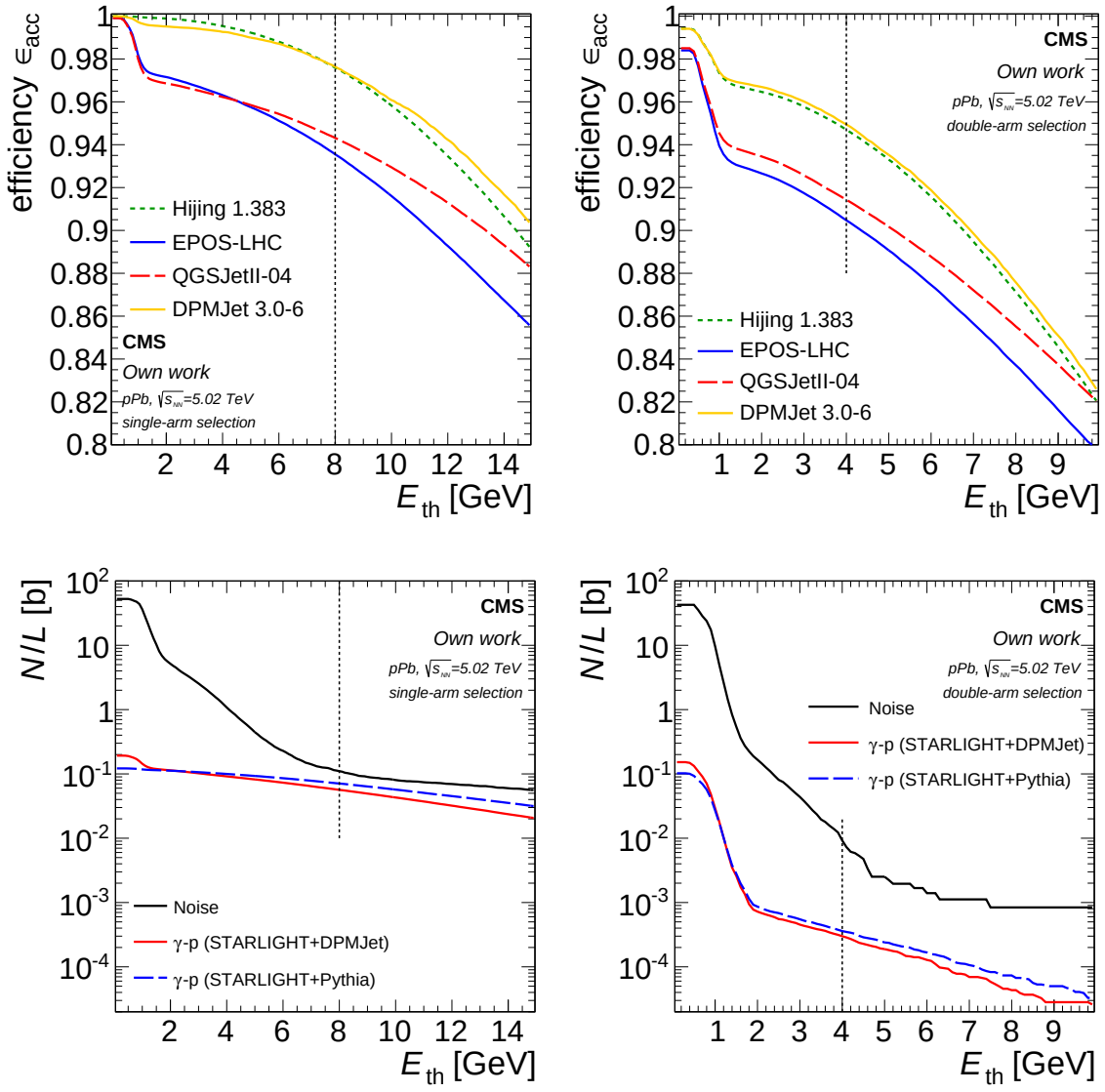
The uncertainties on ϵ_{acc} and N_{em} are obtained from the difference, δ, between the respective estimates of the generators. It is taken as the average of $|\delta|$ over the wider energy interval for E_{th} between 2 and 10 GeV.

The fraction of noise events is estimated by averaging over all run periods, where the observed standard deviation determines the uncertainty. Its value is $(2.0 \pm 0.5)\times10^{-3}$ ($(1.8 \pm 0.8)\times10^{-4}$) for the single-arm (double-arm) selection and relates to $N_{\text{noise}}/\mathcal{L}$ / mb $= 102 \pm 25$ (9 ± 3) of noise contribution to the cross section. However, these values still contain a small amount of double-counted events that will later be subtracted. The determination of the uncertainty on N_{noise} is illustrated in Fig. 74.

The values of ϵ_{acc}, N_{em}, N_{noise}, f_{noise}, and c_{had} are summarised in Table 10.

Corrections and Uncertainties

The cross section is measured for both selection methods in terms of three fully-corrected quantities:



(a) single-arm (b) double-arm

Figure 72: Efficiency for different selection thresholds for E_{HF}. Top panels: Three hadronic interaction models. Bottom panels: Noise (randomly triggered, empty bunch crossings) events and two photo-nuclear simulation samples. The vertical lines represent the threshold energies used in this analysis.

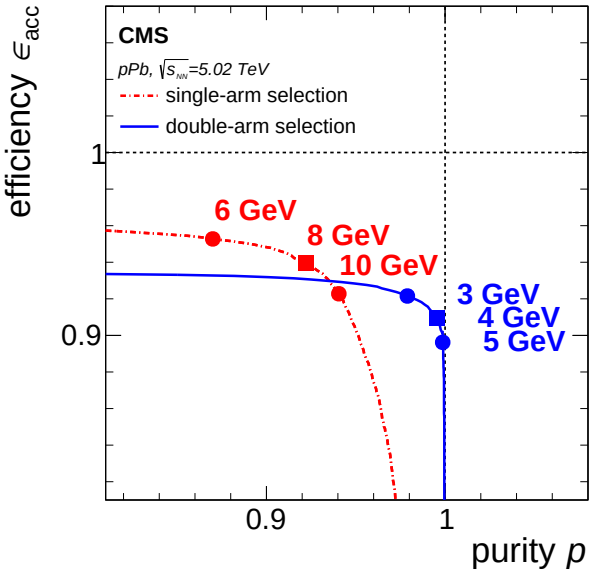


Figure 73: Efficiency of the event selection against the purity as derived from the EPOS and QGSJET generators. Different values of E_{th} are shown with markers. The selected thresholds for E_{HF} are marked with squares.

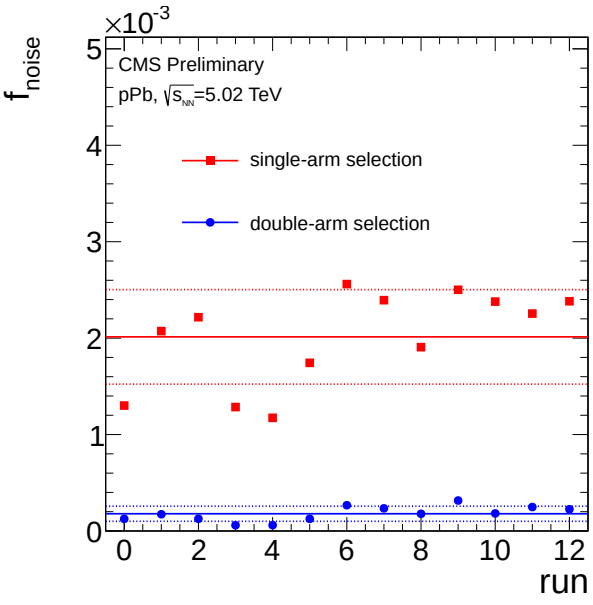


Figure 74: Noise fraction, f_{noise}, over run number. The weighted average and the 1σ error band are shown as solid and dotted lines, respectively.

Table 10: Efficiencies and background events present in the measurement.

Selection	$\epsilon_{\mathbf{acc}}$	$N_{\mathbf{em}}/\mathcal{L}$ **[mb]**	$N_{\mathbf{em}}/\mathcal{L}$ **[mb]**	$f_{\mathbf{noise}}$	$c_{\mathbf{had}}$
Single-arm	0.939 ± 0.005	63 ± 11	102 ± 25	$(2.0 \pm 0.5) \times 10^{-3}$	1.000 ± 0.004
Double-arm	0.910 ± 0.014	0.33 ± 0.05	9 ± 3	$(1.8 \pm 0.8) \times 10^{-4}$	1.000 ± 0.002

- the *visible* cross section σ_{vis}, which is the sum of all processes producing events that pass the selection criteria used in this analysis
- the *visible hadronic* cross section σ_{had}, which is the same as above without the contribution of photo-nuclear processes. This cross section is also corrected to a hadron-level definition allowing it to be compared to generators.
- the *hadronic production* cross section σ_{prod}, which is the above quantity extrapolated to the full phase-space and acceptance. This cross section can be compared to theory but has slightly larger uncertainties due to the extrapolation.

To obtain σ_{vis}, we first correct N_{sel} for noise and photo-nuclear collisions. Furthermore, pileup collisions need to be taken into account since these are multiple simultaneous collisions that count only as one visible collision towards N_{sel}. The distribution of *pileup* collisions is Poissonian. If one collision is selected with efficiency ϵ_{acc}, then i simultaneous collisions are selected in very good approximation with probability $P_i \approx 1 - (1 - \epsilon_{\text{acc}})^i$. We can correct for the uncounted pileup collisions by using the pileup-correction factor

$$f_{\text{PU}} = \frac{\sum_{i=1}^{\infty} i\epsilon_{\text{acc}} \,\text{Pois}(i;\lambda)}{\sum_{i=1}^{\infty} P_i \,\text{Pois}(i;\lambda)} = \frac{\epsilon_{\text{acc}}\lambda}{\sum_{i=1}^{\infty} P_i \,\text{Pois}(i;\lambda)} \,. \tag{86}$$

The interaction probability λ is calculated recursively by using Eq. (55). In the recursive steps, the result of the cross section enters the equation again, until no significant change of $\mathcal{O}(0.1\,\text{mb})$ is found which typically already happens after the second iteration. The pileup correction procedure increases the measured cross section by 2 % and induces an uncertainty on it of $\simeq$0.03 %.

Deriving N_{vis} from the measurement of N_{sel}, which is the part of N passing the selection, is done by applying the noise, double counting, and pileup corrections. Dividing by luminosity yields the visible cross section

$$\begin{aligned} \sigma_{\text{vis}} = \frac{N_{\text{vis}}}{\mathcal{L}} &= \left(N_{\text{sel}} - N_{\text{noise}} + N_{\text{vis+noise}}\right) \frac{f_{\text{PU}}}{\mathcal{L}} \\ &= \frac{1}{\mathcal{L}} \frac{N_{\text{sel}} - N_{\text{noise}}}{1/f_{\text{PU}} - f_{\text{noise}}} \,. \end{aligned} \tag{87}$$

The *hadron-level-corrected* cross section is derived by removing the photo-nuclear contamination and applying the correction factor c_{had}. Its numerical value is by this definition identical to the hadronic part of the visible cross section

$$\sigma_{\text{had}} = \frac{1}{\mathcal{L}} \frac{N_{\text{sel}} - N_{\text{noise}} - N_{\text{em}}}{1/f_{\text{PU}} - f_{\text{noise}}} c_{\text{had}} \,. \tag{88}$$

The *hadronic production* cross section is obtained by correcting the cross section for the limited detector acceptance

$$\sigma_{\text{prod}} = \frac{\sigma_{\text{had}}}{\epsilon_{\text{acc}}} \,. \tag{89}$$

The calculations are performed for each individual luminosity section and the derived σ_{prod} values are shown in Fig. 75 for the example of run 210885 in the top panels and averaged over the luminosity sections for all runs in the bottom panels. Determining the size of the corrections in the short period of 23.31 s has the advantage that any dependencies on the instantaneous luminosity are implicitly removed. To account for the decreasing number of collisions towards the end of the run, the values entering the average over all luminosity sections, and subsequently over all runs, are weighted by the inverse of their variance, which is determined from the Poisson uncertainty. The standard deviation of the "pull" distribution is not unity, which would be expected from a Gaussian distributed sample, but is still small. It is 2.1 ± 0.3 (1.7 ± 0.2) for the single-arm (double-arm) selection. This effect stems from the first runs with p-Pb beam configuration. The observed run-to-run variation seen are taken into account as uncertainties on the final value.

Previously, it was shown that uncertainties on f_{noise}, f_{em}, and f_{pu} exist and all of them enter the systematic uncertainties on σ_{vis}. For σ_{had} and σ_{prod}, also uncertainties due to hadronic interaction models need to be taken into account. For the former a hadron-level definition, requiring minimal extrapolation is used. To achieve this, the reconstructed quantity, E_{HF}, shall be replaced by the hadron-level momentum of a generated particle within the HF acceptance. This quantity is p_{HF} and is defined analogously to Eq. (83) but replacing E_{HF} by the largest absolute value among the momenta, $|\vec{p}|$, of all generated final-state particles within the pseudorapidity intervals of the HF calorimeters, excluding muons and neutrinos. The hadron-level quantity is then derived from the photo-nuclear corrected visible cross section by applying the following correction

$$\sigma_{\text{had,RECO}} \frac{N^{\text{MC}}_{\text{GEN}}}{N^{\text{MC}}_{\text{RECO}}} = \sigma_{\text{had,RECO}}\, c_{\text{had}} = \sigma_{\text{had}} \,, \tag{90}$$

where $N^{\text{MC}}_{\text{RECO}}$ is the number of simulated events that fulfil the selection on the reconstructed quantity E_{HF} and $N^{\text{MC}}_{\text{GEN}}$ is the number of events that fulfil the selection on the hadron-level quantity p_{HF}. We define the purity and efficiency of the hadron-level correction by

$$\text{efficiency} = \frac{N(\text{RECO} \cap \text{GEN})}{N(\text{RECO})} \tag{91}$$

$$\text{purity} = \frac{N(\text{RECO} \cap \text{GEN})}{N(\text{GEN})} \,. \tag{92}$$

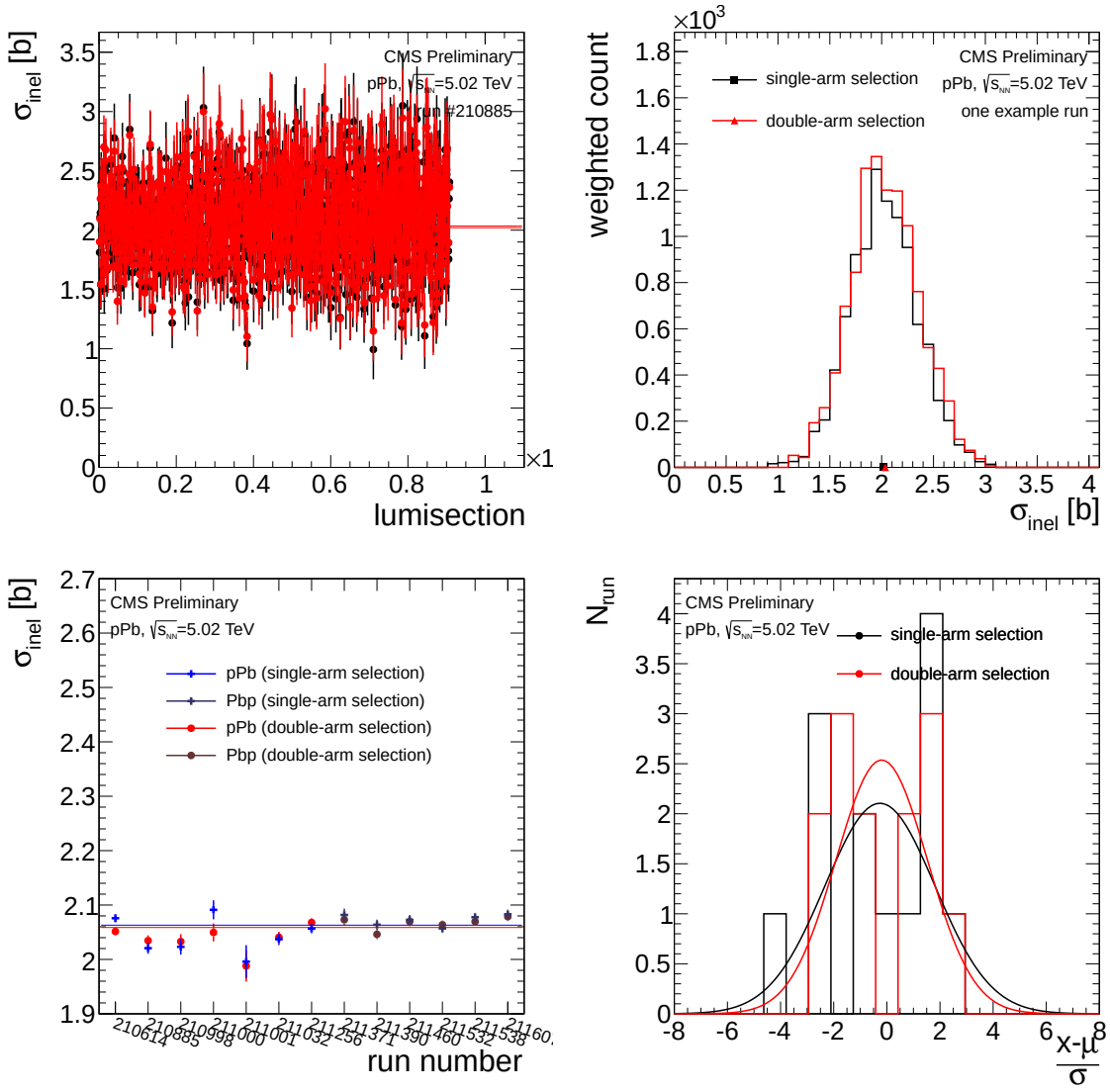


Figure 75: Top panel: The inelastic cross section over time for run 210885 as a function of luminosity section (left) and as the projection (right). Bottom left panel: Measured inelastic cross sections for the two selection criteria as a function of the run number. The error bars denote the statistical uncertainty. The horizontal lines show the weighted mean of the values for each selection criterion. Bottom left panel: The "pull" distribution is shown. The difference of σ_{prod} to the mean divided by the statistical uncertainty is counted for all runs.

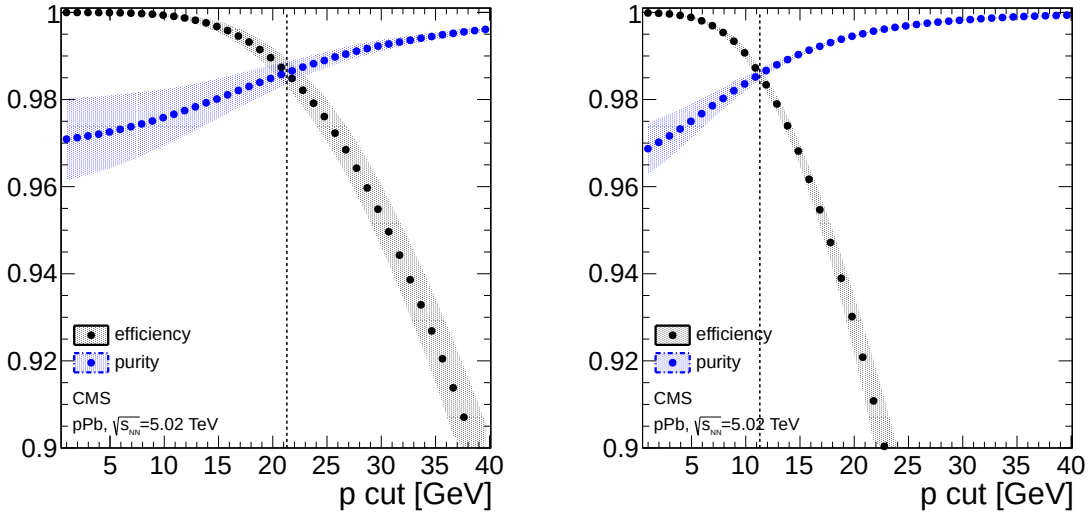


Figure 76: Comparison of the hadron-level selection to the detector-level selection in simulation. The quantities efficiency and purity are defined in the text and are plotted as a function of the threshold for the absolute value of the particle momenta. The shaded uncertainty region is determined by the spread of four hadronic interaction models EPOS, DPMJET, HIJING, and QGSJET. The comparison is given for both event selections: single-arm (left panel) and double-arm selection (right panel).

Please note that the purity and efficiency are not related to the previously introduced p and ϵ_{acc}. Both quantities are plotted as a function of the threshold energy in Fig. 76. We choose the thresholds in a way that c_{had} becomes equal to unity. This yields $p_{\mathrm{HF}} > 21.3\,\mathrm{GeV}/c$ ($11.3\,\mathrm{GeV}/c$) for the single-arm (double-arm) case. Now, c_{had} has no numerical effect on the value of the σ_{had} but since the thresholds are slightly dependent on the event generator, a systematic uncertainty derived from all four hadronic interaction models is assigned to c_{had}. The values of c_{had} with their uncertainties are found in Table 10 for both event selections.

The number of diffractive events that are part of the extrapolation for σ_{had} is, thus, dramatically reduced. For σ_{prod} also all diffractive events contribute, which was argued earlier to be one of the less precisely understood quantities in p-Pb collisions.

For higher (lower) selection thresholds on E_{HF}, ϵ_{acc} changes because less (more) mostly diffractive events are selected. The efficiency is described well by the MC simulation as is illustrated in Fig. 77 but a slight dependence still remains for the extrapolated cross section (black filled markers) for both selections. The visible hadronic cross section agrees with the simulation but deviates towards lower energies. This can be due to the less accurate noise subtraction in this range but also due to more diffractive events that cannot be detected with the HF calorimeters. For illustrative purposes, a modified model with twice the single-diffractive cross section is also shown. For high thresholds the effect is small but the extrapolation at 0 GeV changes. From this we derive three systematic uncertainties.

Modifying the event selection thresholds to $E_{\mathrm{HF}} > 10\,\mathrm{GeV}(5\,\mathrm{GeV})$ influences the cross section by $0.6\,\%(0.2\,\%)$ for the single-arm (double-arm) selection. The diffractive cross section is scaled within allowed upper/lower limits determined by this analysis of 1.13/0.88 and 1.20/0.84 (cf. Section 4.2). They relate, however, to an uncertainty on σ_{prod}, conservatively assumed to be symmetric, of $0.8\,\%(1.1\,\%)$ and is denoted in Table 10 by "ϵ_{acc} (σ_{diff})". The intrinsic uncertainty of ϵ_{acc} determined by the different hadronic interaction models results in an uncertainty of $1.6\,\%(0.5\,\%)$ and is denoted by "ϵ_{acc} (models)".

Since the detector simulation does not include radiation damage, the induced uncertainty is estimated by repeating the measurement without the radiation damage correction in the simulations, which introduces an effect of $1.7\,\%(0.8\,\%)$ on the cross section.

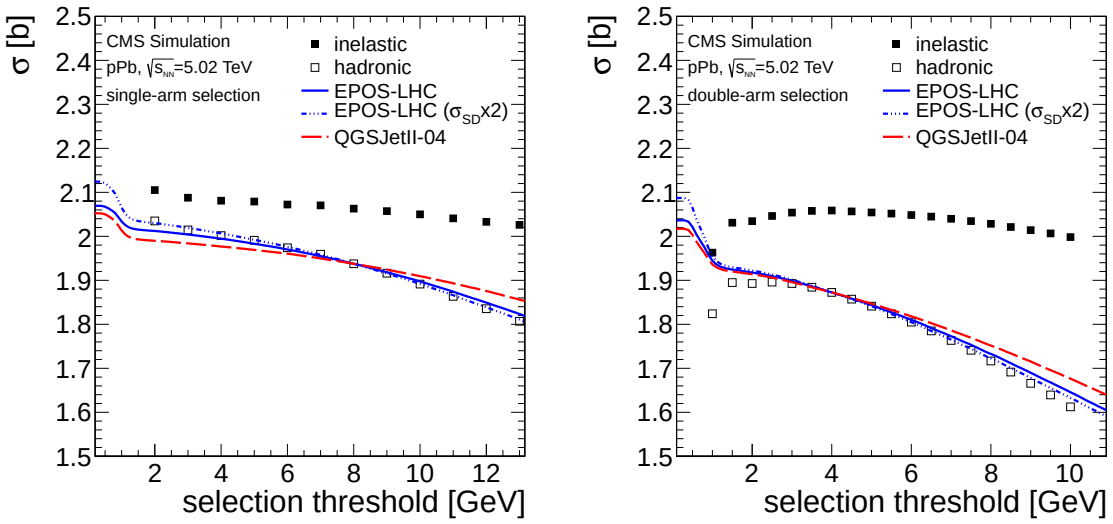


Figure 77: Dependence of the cross section on the selection threshold of E_{HF}. The distributions are normalised at their respective threshold of 8 GeV and 4 GeV. Left panel: for single-arm selection. Right panel: For double-arm selection.

The uncertainties for the single-arm (double-arm) selection are added in quadrature and without taking the luminosity contribution into account are given by

- σ_{vis}: 1.4 % (0.3 %) for which neither c_{had} nor ϵ_{acc} is relevant,
- σ_{had}: 2.3 % (0.9 %) for which the uncertainty on c_{had} contributes, and
- σ_{prod}: 2.5 % (2.1 %) for which the uncertainty on ϵ_{acc} is taken into account.

For all cross sections the uncertainty of 3.5 % due to the luminosity determination has to be added, which dominates the total uncertainty.

Table 11: A list of the systematic uncertainties for two event selection methods.

Source of uncertainty	**Single-arm**	**Double-arm**
Luminosity calibration	3.5 %	3.5 %
Noise subtraction	1.3 %	0.2 %
Photo-nuclear subtraction	0.6 %	<0.1 %
Pileup correction	<0.1 %	<0.1 %
c_{had} (hadron-level correction)	0.4 %	0.2 %
ϵ_{acc} (models)	0.5 %	1.6 %
ϵ_{acc} (σ_{diff})	0.8 %	1.1 %
Event selection	0.6 %	0.2 %
Detector simulation	1.7 %	0.8 %

Results on the Production Cross Section

The two event selections are compared in Table 12. Only for σ_{prod} the values of both selections are expected to agree because of the different acceptances of the two visible cross sections. With a small difference of 4 mb this the agreement is excellent. The values for σ_{had} differ by 65 mb and this difference is also observed for the corresponding predicted cross sections by the interaction models.

Table 12: Summary of cross sections obtained by the two different event selections.

	Selection	σ_{vis} **[mb]**	$\sigma_{\text{vis,had}}$ **[mb]**	σ_{prod} **[mb]**
Data	Single-arm	2003±76	1937±82	2063±89
	Double-arm	1873±66	1872±68	2059±84
EPOS-LHC	Single-arm	-	1947	2082
	Double-arm	-	1883	
QGSJETII-04	Single-arm	-	2059	2181
	Double-arm	-	1998	
DPMJET3.06	Single-arm	-	2116	2166
	Double-arm	-	2055	

Finally, the σ_{prod} values measured with both selections are combined by taking the central value. The statistical uncertainties and the uncertainty on the luminosity are correlated. For the remaining systematic uncertainties the correlation is less direct and they are taken to be uncorrelated. This yields a final result for the hadronic production cross section of

$$\sigma_{\text{prod}} = (2061 \pm 3\,(\text{stat}) \pm 34\,(\text{syst}) \pm 72\,(\text{lumi}))\,\text{mb}\,. \tag{93}$$

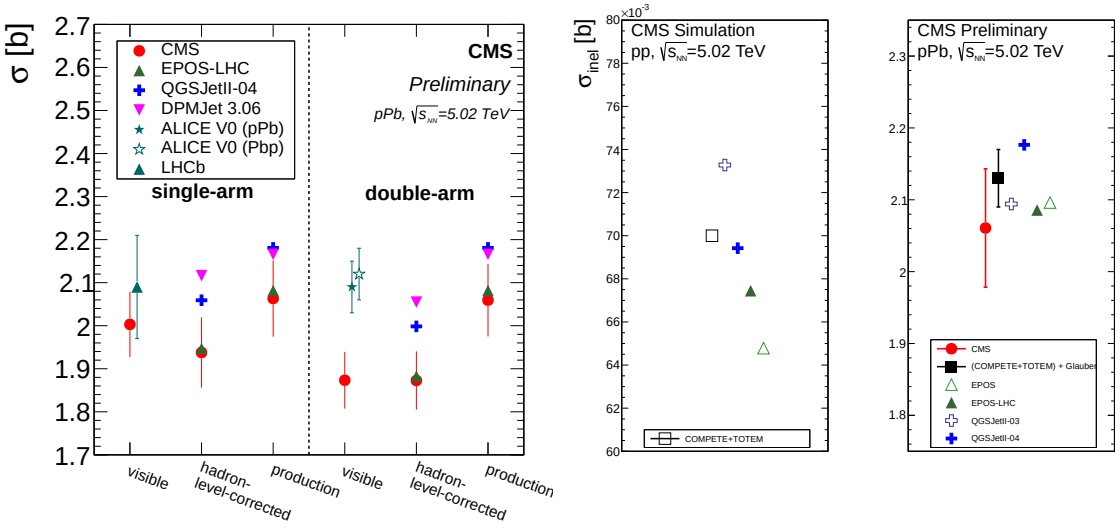


Figure 78: Left panel: The measured proton-lead cross section for visible, hadronic and inelastic contribution. The systematic uncertainty on σ_{prod} is shown as defined in Table 11. For σ_{had} the extrapolation and modelling uncertainties do not enter. In addition, the uncertainty on electromagnetic events does not enter for σ_{vis}. Model predictions from the hadronic interaction models EPOS and QGSJET are given for the hadronic and inelastic cross section. Middel and right panels: Predictions for the inelastic cross section for proton-proton (middle) and proton-lead (right) collisions.

Discussion

One goal is to evaluate the size of corrections to the Glauber calculations of the proton-lead production cross section. To achieve this, the measurement is compared to a calculation that does not include the screening or short-range-correlation effects. The Glauber calculation is based on a proton-proton inelastic cross section, which is derived by the COMPETE parametrisation [Cud+02]. The parametrisation takes the measurement by the TOTEM Collaboration at $\sqrt{s} = 7\,\text{TeV}$ [Ant+13b] into account, and to good approximation the uncertainty of the measurement at $\sqrt{s} = 7\,\text{TeV}$ can be used for the value at 5.02 TeV. The obtained inelastic cross section is $\sigma_{\text{inel,pp}}(5.02\,\text{TeV}) = (70.0 \pm 1.5)\,\text{mb}$. The resulting proton-lead cross section is evaluated with a MC-based Glauber calculation to be $\sigma_{\text{Glauber,pPb}}(5.02\,\text{TeV}) = (2130 \pm 40)\,\text{mb}$. The uncertainty does not include screening and short-range-correlation effects but stems from intrinsic uncertainties of the input parameters. It is obtained from varying, e. g., the proton-proton cross section, the nucleon density function or the size of the lead nucleus. All systematic checks and all input parameters for this calculation can be found in Appendix A.3. Comparison of the value of the Glauber method to our measurement reveals that it lies within the 1σ boundaries determined by its

uncertainties. A 10 % larger (smaller) cross section due to short-range correlations (screening) as is predicted for proton-air cross section can be excluded at $3.5\,\sigma$ ($1.8\,\sigma$) confidence level, respectively. However, if both effects are of similar magnitude they compensate each other, thus, their individual contribution may be hidden from this measurement. Conclusively, it can be said that the net effect of both contributions is small.

The visible proton-lead cross section has also been measured at the same energy by the ALICE Collaboration to be $\sigma_{\text{vis,v0,ALICE}} \approx (2090 \pm 70)\,\text{mb}$ [Abe+14a] and by the LHCb Collaboration to be $\sigma_{\text{vis,LHCb}} = (2090 \pm 120)\,\text{mb}$ [LHC12]. However, the definitions of *visible* are different to that employed here, since photo-nuclear and hadronic collisions are not distinguished and no corrections for the detector acceptance are applied. A direct comparison is therefore not possible. For completeness, their measurements are shown in Fig. 78 together with the visible cross section measured by other LHC experiments. It is remarkable that both values are above that of the hadronic production cross section measured by CMS. The tension is not statistically relevant but might be related, at least in the case of LHCb, with the fact that the measurement had been performed before the luminosity was precisely calibrated.

Finally, the result is compared to other theory predictions. Values predicted by EPOS, DPMJET, and QGSJET are shown in Fig. 79. HIJING does not calculate this value, instead the colour-dipole model DIPSY [AGL05; FGL11] is included. Predictions by EPOS-LHC and DIPSY agree to the measurement given the experimental uncertainties. DPMJET and QGSJET predict a value slightly above the measurement of σ_{inel} with an even larger deviation when compared to the hadron-level-corrected cross section.

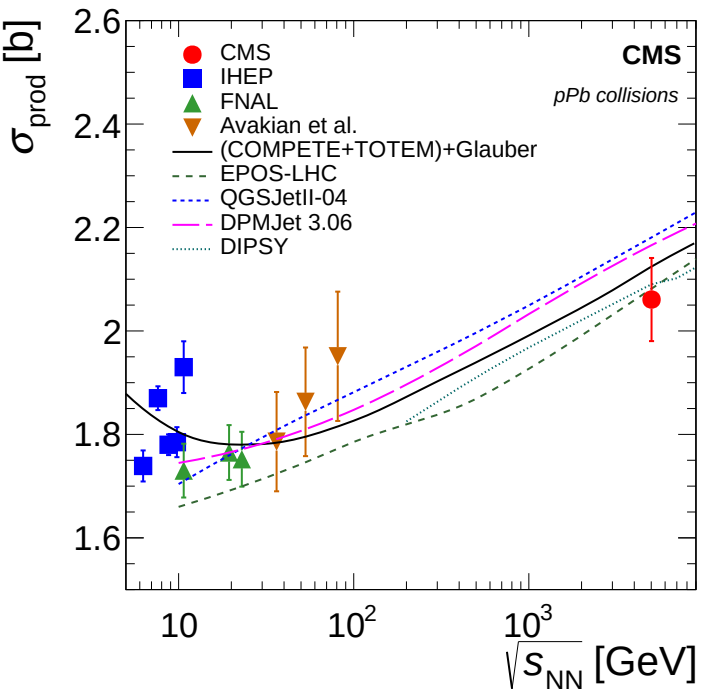


Figure 79: Comparison of the production cross section measurement and lower energy data [Den+73; Car+79; Ava+86] to model predictions as a function of the centre-of-mass energy.

4.2 DIFFRACTIVE COLLISIONS IN PROTON-LEAD DATA

The measurement of the production cross section in proton-lead collisions suffers from an uncertainty in the extrapolation of 0.9 % to 1.6 % depending on the selection, which is one of the largest uncertainties related to the event counting method. At the same time, it is unsatisfactory to base this judgement solely on the difference in ϵ_{acc} between the only two generators that can model diffractive collisions for this beam configuration. Additional measurements of diffractive processes for this type of collisions can verify the argument that in such way the uncertainty is well estimated. Such a measurement has never been performed for proton-lead collisions at this beam energy. In fact, little is known about how nuclear effects influence the diffractive cross section. At the same time, changes in diffraction have a large impact on air shower observables [Lun+04; Pie13; BUE15].

An accurate measurement is difficult to achieve and beyond the scope of this thesis. This can be attributed to the sparse instrumentation in the forward region of experiments and to overlapping signal from photo-nuclear interactions, whose collision process involves electromagnetic exchange particles but is similar to diffraction otherwise. The former problem is related to a fundamental properties of soft processes. In such collisions, a gap in the sorted ensemble of rapidities of the final-state particles is an indication for a diffractive col-

lision. The size of the gap is denoted by Δy. Whereas non-diffractive processes Δy have a roughly exponentially decreasing distribution, diffractive collisions approximately follow

$$\frac{\mathrm{d}N}{\mathrm{d}\Delta y} \simeq \text{const}\,. \tag{94}$$

At large energies where the processes are exclusively described by an exchange of vacuum quantum numbers (or equivalently a pomeron) the rapidity gap is related to the mass of the diffractively dissociated products (diffractive mass, M) as [BP02]

$$\Delta y \simeq \ln\left(s/M^2\right)\,. \tag{95}$$

The right-hand side equals to the longitudinal momentum loss the undissociated particle experiences. The mass can be derived from final-state particles as the invariant mass of the sum of the four-vectors of the final-state particles of the diffractive system. The latter is defined by all particles on the side of the rapidity gap where the beam particle dissociates.

At the LHC, measurements of the diffractive cross section in p-p collisions have been performed based on the rapidity gap that is detectable as an absence of signal in calorimeters or trackers [Aad+12b; Abe+13; Kha+15]. The forward detectors are vital to measure large rapidity gaps. Therefore we employ the HF and CASTOR calorimeters to study diffractive collisions in p-Pb.

4.2.1 DIFFRACTIVE EVENTS MEASURED WITH THE HF DETECTOR

The values of σ_{had} measured with the HF detector contain information about diffractive events. Different acceptances for the single-arm and double-arm selections exist towards diffractive events. An example for the EPOS generator is given in Fig. 80, where the distribution of events in E_{HF} is shown, stacked for different diffractive and non-diffractive processes. The single-arm (double-arm) selection includes the events to the right of the dashed lines as shown in the left (right) panel. The diffractive events have almost no contribution to the distribution for $E_{\mathrm{HF}} > 100\,\mathrm{GeV}$ for the single-arm and $E_{\mathrm{HF}} > 40\,\mathrm{GeV}$ for double-arm selection. However, at low energies (<1 GeV) these events dominate. At such low energies, they are indistinguishable from the large number of misidentified noise events. For more generators the fraction of diffractive events and the acceptance in each category is listed in Table 13. For DPMJET and HIJING no flag for the type of event process is provided and the individual contributions can therefore not be resolved in a straightforward way. Due to aforementioned reasons the diffractive contribution to events of each selection is expected to be low. EPOS predicts more than twice as many double-diffractive (DD) collisions than QGSJET, which instead has a higher single-diffractive (SD) contribution. The requirement on the single-arm threshold roughly halves the single-diffractive contribution but the double-arm requirement reduces it by more than a factor of four. The influence of the event selection on non-diffractive events is smaller (<3 %). The loss in absolute values for the cross section is therefore similar to the loss due to diffractive events.

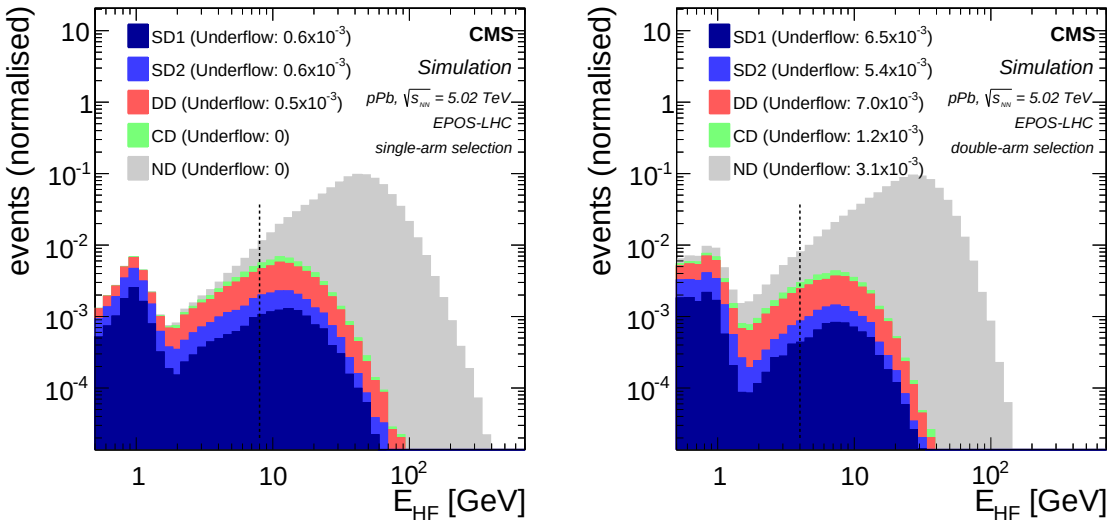


Figure 80: Distribution of E_{HF} for the EPOS generator. The single-arm selection is on the left side and the double-arm selection on the right side. Five stacked categories of inelastic collisions are shown: single-diffractive where the lead ion (SD1) or the proton (SD2) dissociates, double-diffractive (DD), central-diffractive (CD), and non-diffractive (ND) events are shown for each.

Since different fractions of SD and DD events survive in each selection, the ratio of σ_{had} obtained from the double-arm selection to σ_{had} obtained from the single-arm selection, $R_{\mathrm{d/s}}$, can constrain the diffractive cross section. For data, it is measured to be 0.966 ± 0.013, where the uncertainty is estimated by taking the quadratic sum of the photo-nuclear correction, the event selection, and the noise subtraction. The rest is assumed to cancel or is not necessary for the hadron-level corrected cross section. EPOS obtains the closest value to data for $R_{\mathrm{d/s}}$ but all models lie within the uncertainty. A scaling of all diffractive components by a particular factor is introduced and the dependence on $R_{\mathrm{d/s}}$ is studied (Fig. 81 left panel). The ratio increases for a downscaled diffractive component and increases otherwise. To be compatible within 2 standard deviations of the value of data, EPOS cannot be scaled up by more than 1.13 or down by less than 0.88, while for QGSJET those limits are 1.20 and 0.84. A restriction of the diffractive cross section could be obtained within the acceptance of the HF detectors.

An uncertainty from these limits is derived on the production cross section. One can translate the scale factor into a change in acceptance as is shown for the example of the double-arm selection in the right panel of Fig. 81. The allowed scaling results in an uncertainty of 0.8 % (1.1 %) for the single-arm (double-arm) cross section. It is denoted in the list of systematic uncertainties as ϵ_{acc} (σ_{diff}) (Table 11).

Table 13: Fractions of events in each type of collision before and after the two triggers $E_{th} = 8$ GeV (single-arm) and $E_{th} = 4$ GeV (double-arm). Four types of events are considered: single-diffraction (SD), double-diffraction (DD), central-diffraction (CD), and non-diffractive (ND) events. The last two columns show the ratios of the double-arm to the single-arm selection for MC and data (σ_{had} [double-arm]/σ_{had} [single-arm]).

Model	Selection	SD [%]	DD [%]	CD [%]	ND [%]	$\sum$ [%]	$R_{d/s}$ (MC)	$R_{d/s}$ (Data)
EPOS	No selection	4.5	4.5	1.1	90.0	100	-	0.966 ±0.013
	Single-arm	1.6	2.4	0.7	88.8	93.5	0.967	
	Double-arm	1.1	1.8	0.5	87.2	90.5		
DPMJET	No Selection	-	-	-	100	100	-	
	Single-arm	-	-	-	97.6	97.6	0.972	
	Double-arm	-	-	-	94.9	94.9		
HIJING	No selection	-	-	-	100	100	-	
	Single-arm	-	-	-	97.6	97.6	0.970	
	Double-arm	-	-	-	94.7	94.7		
QGSJET	No selection	5.1	1.8	0.0	93.1	100	-	
	Single-arm	1.2	1.1	0.0	92.0	94.3	0.969	
	Double-arm	0.3	0.7	0.0	90.5	91.4		

4.2.2 DIFFRACTIVE EVENTS MEASURED WITH THE CASTOR DETECTOR

To achieve a better acceptance towards low-mass diffractive events—or equivalently events with large rapidity gaps—the p-Pb collisions are analysed with the CASTOR detector. The pseudorapidity coverage of the CASTOR detector extends the measurement to the region, $5.2 < |\eta| < 6.6$, and its noise level is much smaller than the one of the HF detector, making it a well suited instrument for this study.

The first data of 2013 contain collisions for which the proton goes towards CASTOR and the lead ion away from CASTOR. Also data for the reverse beam configuration are available. For air showers, the relevant case is the dissociation of the proton by momentum transfer from the lead nucleus. This case will be studied here with the data of run 210885.

The overall detector performance during the 2013 data acquisition was very good [CMS13a]. One channel in the second module and another in the fourth module were malfunctioning and are excluded in data as well as in the simulated response. Intercalibration between the individual channels is obtained from halo muon events that are collected during periods shortly before and after stable beams are provided by the LHC. The absolute energy calibration of the CASTOR detector for 2013 is derived from analysing p-p collisions of 2011 and 2013. For this the energy response in the adjacent pseudorapidity region that is covered by the HF detector, is measured and then extrapolated with the

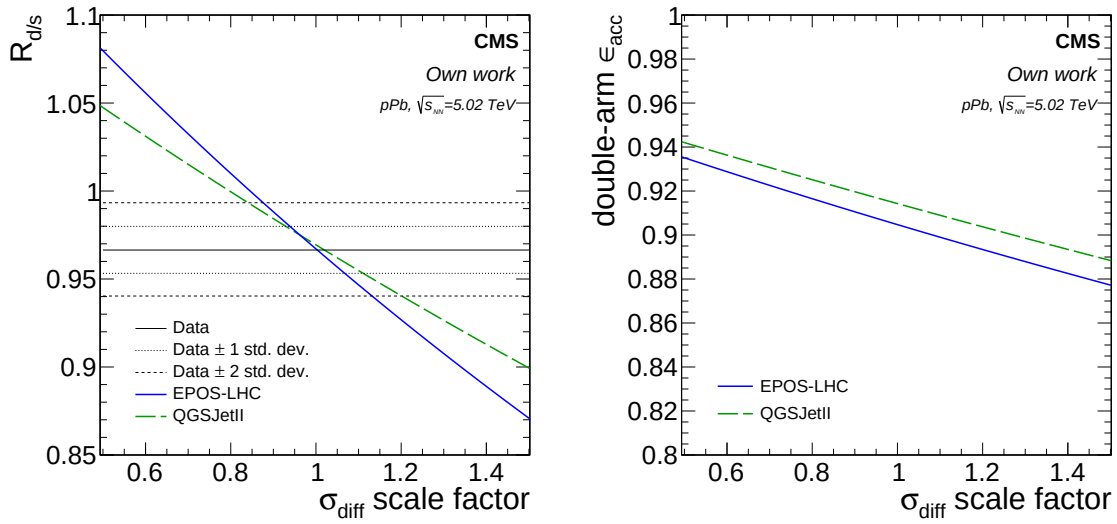


Figure 81: Left panel: Dependence of the ratio σ_{had} [double-arm]/σ_{had} [single-arm] for two different generators on a scaling of all types of diffractive cross sections of a model. The data with its uncertainties is shown as vertical lines. Right panel: Changes in the efficiency of the double-arm selection.

generators EPOS and HIJING into the pseudorapidity region of the CASTOR detector. A reconstructed energy in the CASTOR detector is corrected to match the extrapolated energy. The result is independent of p-Pb data but a rather large uncertainty of 22 % is attributed to this method in Pb-Pb collisions [CMS12]. It is assumed to be of similar order for p-Pb.

The total measured energy $E_{\text{cas,tot}}$ that corresponds to the sum of the reconstructed energy of the first 5 modules (2 electromagnetic and 3 hadronic) is used to study p-Pb collisions. Since stray magnetic fields, stemming from the central solenoid, reduce the efficiency of light collection by the PMTs of modules 6 and 7, they are excluded. For module numbers larger than 7 only the tails of the cascades that hadronic particles produce is seen. The simulation can accurately describe the first 5 modules but due to geometrical tilts of the detector halves that can appear when each half is mounted on its support table, large uncertainties for the rear modules are present. The tilts are not accounted for in the geometry description of the simulation but are expected to have negligible influence on the first 5 modules. The energy measured in the rear modules is not included in $E_{\text{cas,tot}}$.

Detailed studies of the noise in the detector have been performed, which are necessary to understand the contribution for events with small $E_{\text{cas,tot}}$. The left panel of Fig. 82 shows a comparison between unbiased events and the contribution of noise events estimated by the random trigger similarly to the studies that are performed for the HF detector. A disagreement is visible for the response of the random triggered events with empty bunch crossings that show a larger cross section than the unbiased events. This effect appears due to the pulse shape of the signal in a collision event. It slowly drops with time but can

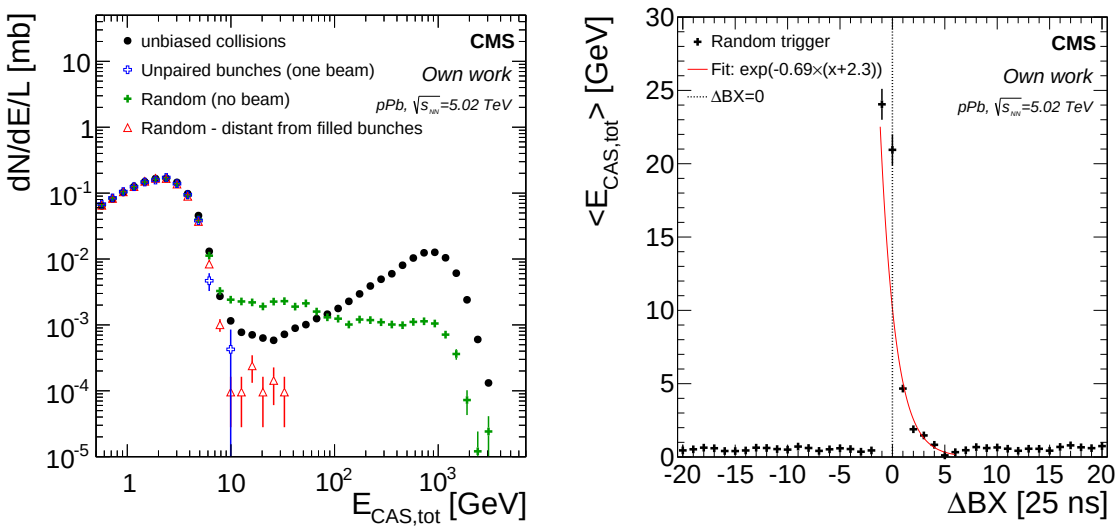


Figure 82: Left panel: The noise contribution in CASTOR estimated by different triggers and selections and compared to data taken with the unbiased trigger. Right panel: Mean of sum over signals in each channel of CASTOR. The events are recorded with the random trigger. The signal is shown as a function of the distance to the closest filled bunch slot, ΔBX.

introduce—for noise studies unwanted—signal in the empty bunch crossings after a filled one (cf. right panel). The noise is therefore estimated only for bunch slots at a certain distance in time, ΔBX, from the corresponding closest filled bunch slot. It is evaluated in the range $-5 > \Delta\text{BX}/25\,\text{ns} > 15$. Its distribution is shown as red triangles in the left panel and also agrees with the measurement from unpaired bunches (blue crosses), implying negligible contribution from beam gas events. The treatment of the detector noise is applied in the following for studying diffractive events using data.

The contribution of diffractive events seen in the CASTOR detector can be enhanced by requiring a rapidity gap. Since detector acceptances are defined in pseudorapidity, in the following all gaps are defined in pseudorapidity, η, and not in rapidity. The gap is denoted by $\Delta\eta \simeq \Delta y$. Three detectors that cover a more central region than CASTOR are analysed: the HF calorimeter on the positive pseudorapidity side (HF+), the pixel tracker, and the HF calorimeter on the negative pseudorapidity side (HF-). Together they instrument the region $-5 < \eta < 5$ with gaps at $2.5 < |\eta| < 3$. For every event, the states "full" or "empty" are assigned to each of them depending on whether they detect particles or not. The "full" state is defined by an energy requirement in HF, $E_{\text{HF}} > 8\,\text{GeV}$ (Section 4.1.3) and by an requirement of at least one reconstructed pixel track in the tracker.

Using this definition, four scenarios based on these detectors are studied. They are, in the order of decreasing $\Delta\eta$: all detectors full, all full except HF+, HF- full and the rest empty, and all empty. The latter corresponds to a very large gap of almost 10 units in η

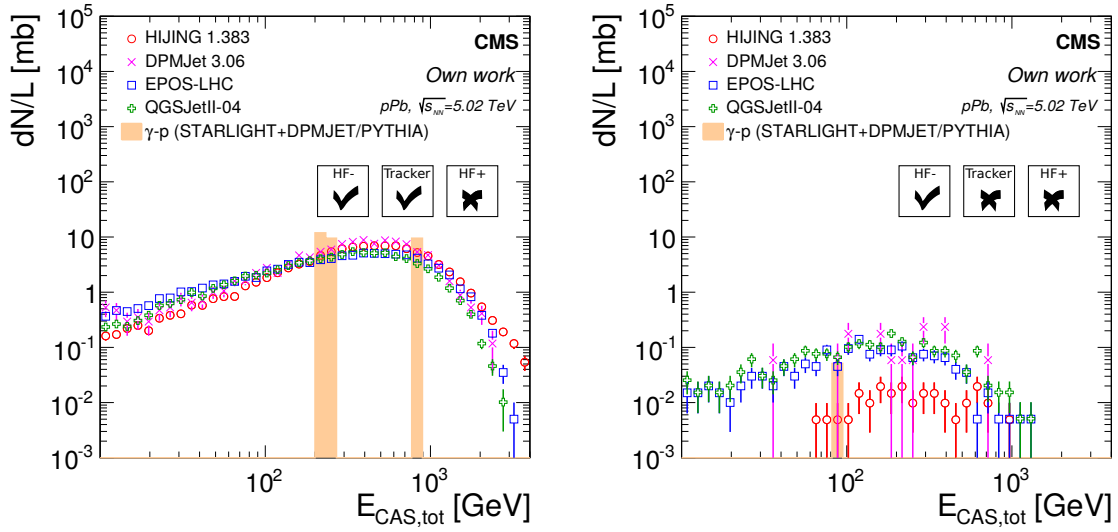


Figure 83: Distribution of the sum over the generator-level energy of all particles in the CASTOR acceptance is shown for multiple generators. Two different selections for a smaller (left) and a larger rapidity gap (right) are shown. Empty detectors are denoted by a cross and detectors with signal by a tick.

and photo-nuclear events dominate in this case since the particle production is at very forward rapidities in γp interactions. Collisions between $\gamma\gamma$ in ultra-peripheral events, which have a very small cross section, become important when the hadronic cross section is very low. Their contribution is expected to be smaller than for photo-nuclear events and will not influence the scenarios shown in Fig. 83. In this figure, the distribution of energy is shown at generator-level. The full and empty detectors correspond to a definition using the simulated final-state particles in the acceptance region of the detectors. For a detector to be flagged as full, the requirement is a particle with 21.3 GeV or 1 charged particle with $p_{\mathrm{T}} > 0.2$ GeV in the HF detector and tracker, respectively. For both, small and large gap, no obvious contribution from photo-nuclear events is observed even though the number of simulated events is too small to make any quantitative claim. For the larger gap the hadronic interaction models can be separated into a group that describes diffractive events (EPOS and QGSJET) and one that does not (DPMJET and HIJING). The latter group predicts a smaller cross section in the region of $E_{\mathrm{cas,tot}}$ from 100 GeV to 1000 GeV. The small number of simulated events for the DPMJET generator allows no quantitative differentiation.

The same quantities are checked with a detector simulation where a data-driven noise component is added and subsequently compared to data recorded by the unbiased trigger. The contribution of misidentified tracks due to electronic noise is negligible. For the case of no rapidity gap and small rapidity gap (the latter is shown in left panel of Fig. 84) the hadronic models agree with data except for $E_{\mathrm{cas,tot}} \simeq 10$ GeV, which could be due to an

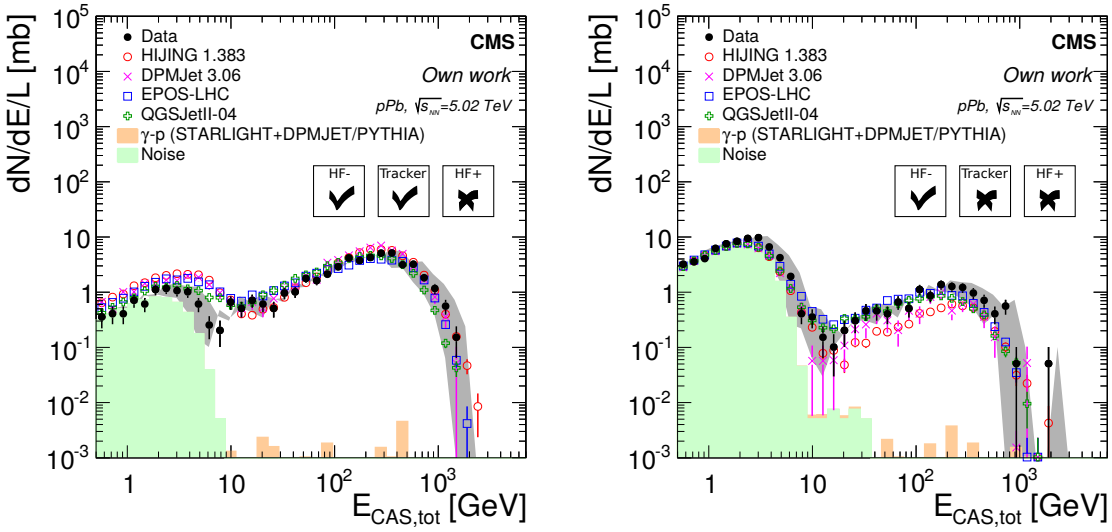


Figure 84: Distribution of the sum of recorded signals of all good channels in the first five modules of CASTOR. Data and the response of detector simulation with multiple generators are shown. Statistical uncertainties are given by error bars and the grey error band reflects the energy calibration uncertainty of the CASTOR data. Two different selections for a smaller (left) and a larger rapidity gap (right) are shown. Empty detectors are denoted by a cross and detectors with signal by a tick.

underestimated contribution of misidentified events from electronic noise. For large and very large $\Delta\eta$ (the former is shown in the right panel) only EPOS and QGSJET agree with data. The simulated response of HIJING and DPMJET is too low by about a factor of two. Without increasing statistics and quantitatively taking the energy uncertainty into account, no ultimate conclusion can be drawn about the missing diffractive component of the HIJING and DPMJET models. Nevertheless, the data suggests that the diffractive component is better described in the EPOS and QGSJET models.

This study outlines the first-ever look at diffraction in p-Pb collisions at LHC energies. It exploits the excellent acceptance of the CMS experiment in the forward region and bases its measurement on energy deposited in the CASTOR calorimeter. The effect of the diffractive component in the measured energy by the CASTOR detector is clearly visible and can be distinguished by requirements on an identified rapidity gap in the central part of the experimental coverage. Further studies are suggested, in which $\Delta\eta$ is measured by reconstructed objects with the particle flow method, which combines tacking as well as electromagnetic and hadronic calorimeter responses. In this way, the pseudorapidity coverage is improved and gaps can be identified more accurately. Additionally, studying the reversed beam direction, where the dissociation products of the nucleus can be detected in the CASTOR detector, offers clues about how the probability for bound nucleons to interact diffractively is influenced.

4.3 STUDYING THE SENSITIVITY OF PARAMETERS OF HADRONIC INTERACTION MODELS TO FORWARD PHASE-SPACE MEASUREMENTS AT THE LHC AND COSMIC RAY DATA

At LHC experiments multiple measurements that are concerned with the forward phase-space were published in the recent years. This calls to answer the question, which of these are the most constraining analyses for model parameters of generators for hadronic interactions that are used in air shower simulations. Here, we propose a study of a physically sensitive subset of parameter of one particular model to explore this effect. The application of this method encompasses the determination of constraints and sensitivity of the parameters to measurements and identifying an optimal parameter set that describes the measurements well. Ultimately it can be determined, which future measurements are the most relevant and how accurately observables need to be measured to bear an impact on the constraints of model parameters. Furthermore, the approach to simultaneously analyse data from UHECR experiments and from LHC can be used to constrain model uncertainties in a global picture.

4.3.1 ANALYSIS BASED ON EXPERIMENTAL DATA FROM ACCELERATORS AND FROM THE PIERRE AUGER OBSERVATORY

The workflow of this study is very straightforward whereas its application is technically complex due to rigid implementation of the models and tools that do not foresee modifying the model parameters on the fly. Also the amount of computation time is significant. The following steps have to be taken:

1. select a set of parameters for the event generator under investigation,
2. select measurements that are sensitive to the selected parameters,
3. define a phase-space grid for the parameter scan,
4. simulate events with Monte Carlo tools for each parameter set,
5. calculate the likelihood to describe all measurements for each parameter set, and
6. analyse many parameter sets to study the sensitivity.

The choice of a likelihood method makes the approach extremely flexible. Also complicated non-Gaussian data and assumptions can be properly implemented. For symmetric uncertainties a Gaussian function approximates the likelihood function sufficiently well. In such limited cases, a χ^2-method is fully equivalent to the new technique proposed here. However, the likelihood can also be calculated for asymmetric uncertainties given the known standard deviations σ_+ and σ_-. In this case an asymmetric piecewise Gaussian

distribution is used to obtain the probability for a model prediction y_i to agree with the measured value $y_{\text{data},i}$

$$P(y|y_{\text{data}},\sigma_+,\sigma_-) = \begin{cases} c_1\,\text{Gaus}(y_{\text{data}}-y,\,\sigma_+) & \text{for } y \geqslant y_{\text{data}} \\ c_2\,\text{Gaus}(y_{\text{data}}-y,\,\sigma_-) & \text{for } y < y_{\text{data}}\,, \end{cases} \tag{96}$$

where the constants c_1 and c_2 are chosen such that $\int_{-\infty}^{\infty} \mathrm{d}y\, P(y|y_{\text{data}},\sigma_+,\sigma_-) = 1$ and $c_1\,\text{Gaus}(0,\sigma_+) = c_2\,\text{Gaus}(0,\sigma_-)$. Obviously, for $\sigma = \sigma_+ = \sigma_-$ the method is again equivalent to the χ^2-distribution. The only technical difference is that the probability, P_i, for each data point y_i is $P_i = 1/2$ in case the model deviates by 1σ from the measurement. Besides (a)symmetric error bars, also Poissonian distributions as well as any other statistics can be used. Furthermore, priors and constraints of all kinds are very straightforward to consider.

Systematic uncertainties are often strongly correlated between for quantities given in multiple bins. To fully take this into account, the likelihood needs to be calculated separately for statistical and systematic uncertainties. For the latter, a nuisance parameter needs to be included to describe a systematic shift of data in each bin. Since the correlations are in general not known to us, multiple measurements are added and combined to a normalised logarithm of the likelihood

$$\log(L) = \sum_i w_i \log(P_i)\,, \tag{97}$$

where w_i is the inverse of the number of degrees of freedom $w_i = 1/\text{ndf}$. The normalisation enables the comparison between different analyses. This can be further improved if exact correlations of the uncertainties are known, which can only be achieved by further correspondence with the authors of the measurements.

4.3.1.1 *Interaction Model and Parameter Selection*

First of all it is required to choose parameters of interest that are as sensitive as possible to the selected measurements. Here, we focus on the influence of fragmentation, cross section calculations, and diffractive processes. All of these are in close relationship to the measurement of σ_{prod} in p-Pb collisions (Section 4.1). In the presented study we concentrate on the hadronic interaction model EPOS. It is one of the most actively developed models and exhibits good agreement with LHC data at central rapidities. It is also heavily used in UHECR physics. The model can easily be replaced by another one in future studies. Three parameters of EPOS are selected as input for the parameter grid. These parameters are scanned using five to six points in close vicinity to the default value as implemented in the most recent version EPOS-LHC. The selected parameters are listed, for which the default values are highlighted in bold font:

- Name: "wdiff(2)": w_{diff} = {0.865, 0.870, **0.875**, 0.880, 0.885}
  This parameter is related to the probability of simulating a hadronic collision diffractively or non-diffractively. The parameter is not the same as w_{diff} of Eq. (2.77) in

[Dre+01]. Instead it is a non-linear parameter introduced in EPOS only after this publication. It is a free parameter to scale both hard and soft diffraction and exists separately for pions (1), protons (2), kaons (3), J/ψ (4) but here only the second component (protons) is modified. For larger values more diffractive events are produced.

- Name "epscrx": ϵ_{crx} = {0.25, 0.3, 0.35, **0.4**, 0.45, 0.5}
  This parameter affects mainly the cross section extrapolation and multiplicity of particle production in an interaction. For elementary parton-parton interaction in EPOS, the scattering amplitude, $F(s,b)^2$, (cf. Eqs. (23) and (24)) are defined via a parametrisation $\alpha\,(x^+)^{\beta+\epsilon}\,(x^-)^{\beta+\epsilon}$, where α and β are two fit parameters and $x^\pm$ are the longitudinal momentum fractions of projectile and target. The exponent ϵ expresses a modification of mainly the hard component that dominates at x close to unity. This energy- and impact parameter-dependent modification has a saturation scale that is determined by ϵ_{crx}. The saturation scale is responsible for damping a too quick rise of the cross section with energy and effectively limits the cross section already at LHC energies. The cross section at LHC energies scales with the inverse of ϵ_{crx}.

- Name "alpdi": α_{diff} = {0.2, 0.4, 0.725, **1.05**, 1.35, 1.65}
  This parameter determines the shape of the remnant mass distribution for diffractive processes. The distribution follows $1/M^{2\alpha_{\text{diff}}}$. Therefore, for larger values of α_{diff} more low-mass states are present, which influences in particular the forward multiplicity.

The physically allowed region for each parameter is not precisely known. The parameter range can be justified only retrospectively, by the outcome of the study. If the measurements disfavour a certain parameter space, the range is chosen appropriately.

One important aspect is the computation time needed to test each point in the parameter space. The time scales with $\propto \prod_i n_i$, where n_i is the number of tested phase-space points for each parameter $i = 1 \ldots d$. When each parameter has the same number of grid points, n_{grid}, a regular grid requires $N = \left(n_{\text{grid}}\right)^d$ simulated points. From statistics it is known that phase-space grids for $d \gtrsim 3$ have a very poor performance relative to Monte Carlo sampling techniques. The computation time per phase-space point also critically depends on the number and nature of the chosen measurements. Here, we use three parameters and with these the fundamental power of the new approach is demonstrated. A sophisticated Markov-Chain chain sampling including more parameters is foreseen for the future.

4.3.1.2 *Selected Measurements and Data Sets*

The focus primarily lies on improving the description of physics relevant for air showers where many measurements at various centre-of-mass energies and with different colliding particles are potentially important. Uncertainties in air shower simulations come from the extrapolation in energy and in the phase-space of the forward region, nuclear effects,

and collisions of high-energy pions. It follows that measurements in the forward phase-space and in proton-lead collisions at the LHC at the currently highest achievable particle energies in a controlled environment offer the most constraining evidence. A selection of measurements is presented below and gives a holistic coverage of the physics of the underlying event, diffractive processes, and energy dependence of the cross section due to the pomeron component. As such, processes with large cross sections and a significant amount of particle production (inclusive) or processes that fuel a particular component in the air shower (e. g. number of muon produced by baryons) are selected.

The code basis, which contains the implementation of each analysis, the management of the production of simulated events on a large computer cluster, and the analysis comparison has been written from scratch. One possibility to avoid this would have been to realise the study with help of a combination of the RIVET and PROFESSOR tools [Buc+13; Buc+10]. However, there exist significant advantages not relying on them, which are the absolute flexibility in determining the simulation-data agreement and that no limits in terms of input to an analysis is given. The latter enables us to include also air shower-based analyses. So far this is implemented for the measured average X_{max} of air showers at the highest energies measured by the Pierre Auger Observatory. No assumption on the mass composition is made with the used method. In the future, other air shower-based analyses can be implemented. Constraints on specific model parameters may be significant since some air shower measurements exhibit a strong sensitivity to interaction models, e. g., the measurement by the Pierre Auger Observatory of the muon production depth, X^{μ}_{max}. Since many showers need to be simulated, the application of either high thinning levels or cascade equation-based approaches are favourable. For example, the inclusion of the X^{μ}_{max} measurement requires a realistic simulation of the lateral component of the shower at ground, which plays no role in the measurement of X_{max}.

Many of the available measurements are based on data from proton-proton collisions at the LHC and these data are valuable input to any model describing nuclear collisions (Section 2.3.4). The measurements that are currently used to determine the likelihood of the parameter set are the following:

I) An analysis based on air shower data: Measurement of $\langle X_{\mathrm{max}} \rangle$ for energies above 10^{17} eV including data of the low-energy extension of the Pierre Auger Observatory, High Elevation Auger Telescopes (HEAT) [Por14]. The measured X_{max} is closely correlated to the mass of the primary cosmic rays. Based on arguments of the astrophysical acceleration mechanism, the incoming particles are nuclei whose masses are between proton and iron. It is so far the only analysis that cannot directly make use of determining the likelihood via Eq. (96). A simple implementation is a flat likelihood when the data is in between the predictions made for proton, $X^{\mathrm{p}}_{\mathrm{max}}$, and iron primaries, $X^{\mathrm{Fe}}_{\mathrm{max}}$. However, when the data are outside this range, the likelihood drops with Gaussian tails that are determined by the uncertainties of the measure-

ment. Therefore, we define for the measured value $X_{\max} \pm \sigma$ with the corresponding prediction derived from iron and proton simulations $X_{\max}^{\mathrm{Fe/p}}$

$$P(X_{\max}^{\mathrm{Fe/p}} | X_{\max}, \sigma) = \begin{cases} c_1 \,\mathrm{Gaus}(X_{\max} + \sigma - X_{\max}^{\mathrm{p}},\, \sigma) & \text{for } X_{\max} \geqslant X_{\max}^{\mathrm{p}} - \sigma \\ c_2 & \text{for } X_{\max}^{\mathrm{Fe}} + \sigma < X_{\max} < X_{\max}^{\mathrm{p}} - \sigma \\ c_1 \,\mathrm{Gaus}(X_{\max} - \sigma - X_{\max}^{\mathrm{Fe}},\, \sigma) & \text{for } X_{\max} \leqslant X_{\max}^{\mathrm{Fe}} + \sigma\,. \end{cases}$$

The normalisation constants c_1 and c_2 are chosen such that the distribution is continuous and its integral is unity. Furthermore, a prior distribution must be multiplied that compensates the dependence of c_2 on the size of the difference $X_{\max}^{\mathrm{p}} - X_{\max}^{\mathrm{Fe}}$. This is achieved by dividing by c_2 as a last step.

The large number of showers needed for a comparison to data requires time-intensive simulations. To decrease the statistical fluctuations below the level of natural shower-to-shower fluctuations, we simulate 1000 showers for each point in energy. Since the data of the low-energy extension are employed, showers with $\lg(E/\mathrm{eV}) = [17, 17.5, 18]$ are simulated and linearly extrapolated in $\lg(E)$ to higher energies. From arguments of analytic treatment this is acceptable as is shown in Section 2.2.2 and is known from MC based studies.

II) Analyses based on LHC data:

- **i)** Measurement of the inelastic cross section at $\sqrt{s} = 7\,\mathrm{TeV}$ and $8\,\mathrm{TeV}$ by the TOTEM experiment. These two measurements belong to the most accurate measurements of the inelastic cross section in this energy region. The are obtained using strategies for deriving the value with and without the knowledge of the luminosity. Knowing the inelastic cross section for proton-proton collisions is one of the main requirements for calculating the proton-air cross section. References: [Ant+13c; Ant+13a]
- **ii)** Measurement of neutron spectra at $\sqrt{s} = 7\,\mathrm{TeV}$ by the LHCf experiment. The measurement probes the ultra-forward fragmentation region and is related to the number of baryons produced in a proton-proton collision. With differential cross sections of less than 500 pb this analysis needs a high number of simulated events. To have acceptable statistical fluctuations we simulate 4×10^5 events. Reference: [Kaw14]
- **iii)** Measurement of the multiplicity over pseudorapidity at $\sqrt{s} = 8\,\mathrm{TeV}$ in a combined analysis of CMS, whose tracker covers the central part, and TOTEM, whose T2 telescopes cover the very forward region. By including a large part of the phase-space this strongly constrains the number of charged particles produced in an interaction. The data contains results for three different event samples: an inclusive sample, triggered by any of the T2 telescopes in the two hemispheres, a non-diffractive sample, triggered by both T2 telescopes, and a single-diffractive-enhanced sample, triggered by exactly one of the two T2 telescopes.

The latter has the lowest cross section and 10^5 events are simulated to provide sufficient statistics. Reference: [Cha+14b]

iv) Measurement of σ_{prod} for proton-lead collisions at 5 TeV (Section 4.1). It is included to see how the nuclear cross section scales when the proton-proton interaction is changed. Apart from the cross section, 10^4 events need to be simulated to derive the efficiencies of the single-arm and double-arm selections. To reduce the bias of extrapolation, σ_{had} instead of σ_{prod} is compared. Reference: [CMS13b]

v) Measurement of diffractive cross sections in for p-p collisions at $\sqrt{s} = 7$ TeV. Here, single- and double-diffractive cross sections and their differential distributions as a function of the longitudinal momentum loss (ξ) in the range of -5.5 to -2.5 are studied. The analysis includes the CASTOR detector resulting in a very good acceptance to identify both, single- and double-diffractive collisions with low masses. This is the only analysis that measures diffraction directly and is very sensitive to at least two of the parameters under investigation. The analysis can be evaluated with the same events that are produced for item ii). Reference: [Kha+15]

To summarise the Monte Carlo production, 3000 proton-induced air showers, 3000 iron-induced air showers, 430 000 LHC-type collisions and 3 cross sections at different energies are simulated. For air shower simulations, CONEX [Ber+07] is used, which tabulates the energy deposit of low-energy hadrons, electrons, and photons using cascade equations. The parameters of the cascade equations change when the model parameters are adjusted. The tables `conex.p2epo` and `conex.wepo` that contain information about the collision probability and energy deposit have to be rebuilt for each set. This is done for 20 logarithmic energy bins from $10^{10.5}$ eV up to 10^{15} eV for the particle types of p, n, π, $K^{\pm}$, K^0_L with 20 000 events in each bin. This process is lengthy and lasts for about 600 CPU hours but saves an immense amount of time when calculating multiple air showers. Even then, the proton-induced and iron-induced showers need about 10 and 400 CPU hours, respectively. The threshold energy at which CONEX switches from full 3D Monte Carlo to cascade equations is set to 0.1 % of the primary energy for all particle types to assure that the tables are sufficient. Even before the tables can be built, also the interaction model requires regenerated tables that contain, e. g., the cross sections at different energies. Here, the tables `epos.inirj` and `epos.inics` depend on the parameter set and take about 1 CPU hour to produce.

The LHC-type collisions are generated with CRMC (Section 3.4.1). The updated tables for the model and parameters are configured via the `crmc.param` file. To simulate the events, which are needed at different energies and for p-p and p-Pb primaries, about 15 CPU hours have to be invested. Contrary to air showers that are saved in the well compressed ROOT file format, `zlib`-compressed `HepMC` event records are saved, amounting to a substantial size of 7 GB of storage space per parameter set. The huge storage requirements and a total simulation time of 270 000 CPU hours demand a powerful computer farm and, thus, `lxbatch` at CERN was used for this purpose.

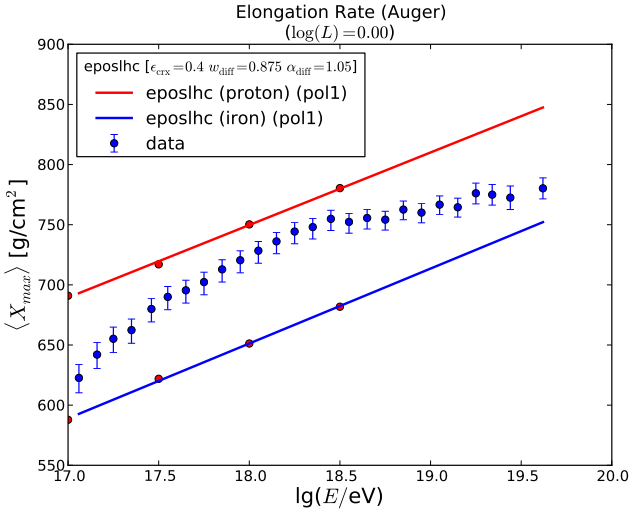


Figure 85: Comparison (part 1) of data and simulation obtained with the default setting of EPOS-LHC. The $X_{\max}$ measurement by the Pierre Auger Observatory (analysis I) based on cosmic ray data is shown. Predictions for proton and iron primaries that are linearly inter- and extrapolated are shown as well.

4.3.1.3 *One Example with the Default Parameter Set*

This section will introduce how the global likelihood is calculated and example plots for each analysis compared to EPOS-LHC with its default parameter set are given in Figs. 85 to 87. The $\log(L)$ values are given in the title of the individual subplots. The comparison to the data from $X_{\max}$ (analysis I) returns a likelihood of 1 as the data is within the proton and iron predictions for all energies. All other model predictions are fitted with a univariate spline (dashed line), $f(x)$, to filter out fluctuations due to the limited number of simulated events. The y-value of a data point at x will be compared to $f(x)$ with the possibly asymmetric Gaussian probability distribution. Summing up the logarithm of the likelihood in each data point results in the value specified in the title of each plot. Since the absolute value of the likelihood has no statistical relevance the reduced χ^2-value is given as well. From this, we see that that ξ-distribution (analysis II.v) are not described well, the neutron spectra (analysis II.ii) show a good agreement only in the highest rapidity bin. The multiplicity distributions (analysis II.iii) are overshot for the case where the non-diffractive contribution plays the major role but are slightly too low for the single-diffractive-enhanced case. The inclusive cross sections are reproduced well. The respective diffractive cross sections (analysis II.v) are however too high and too low for the single- and double-diffractive case, respectively.

None of the measurements is perfectly described and an improvement could possibly be achieved by finding a better parameter set. In the following all parameter sets are systematically compared to each other.

4.3.1.4 *Comparison of Likelihood Between all Parameter Sets*

Since the likelihood for each analysis depends on the parameter set $\vec{\theta}$, it is a function $L = L(\vec{y}_{\text{data}}, \vec{\theta})$, where the measurements $y_{\text{data},i}$ are not changed. The maximum likelihood

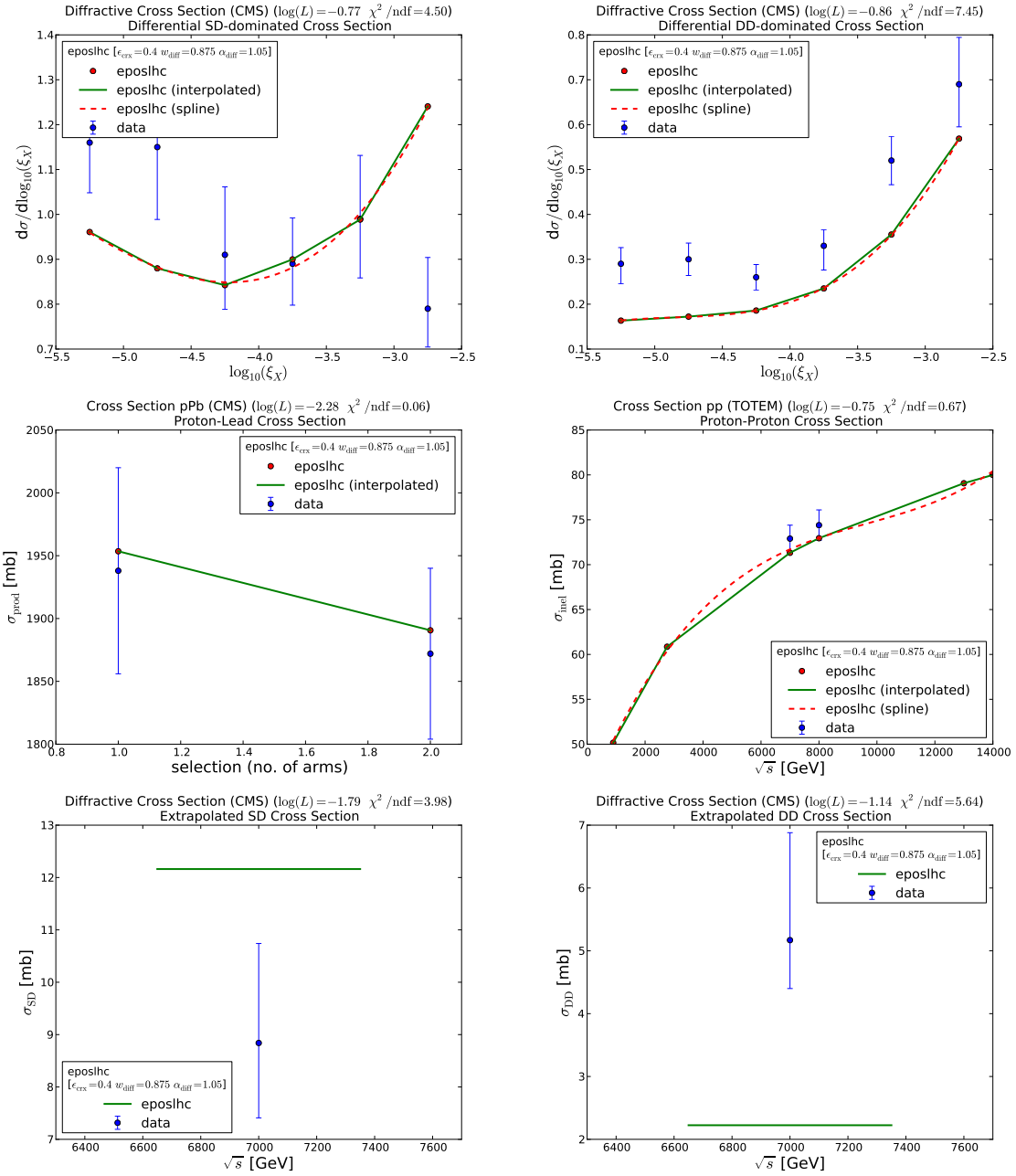


Figure 86: Comparison (part 2) of data and simulation obtained with the default setting of EPOS-LHC. The first row shows the diffractive cross sections over ξ of the diffractive system for single and double-diffractive processes (analysis II.v). The second row shows the proton-lead and the proton-proton cross section (analyses II.iv and II.i). The two data points of the former measurement belong to the single-arm and one for the double-arm selections. The last row shows the diffractive cross sections (as measured in analysis II.v). The fitted univariate spline used for interpolation of the simulated result is shown as a dashed red line.

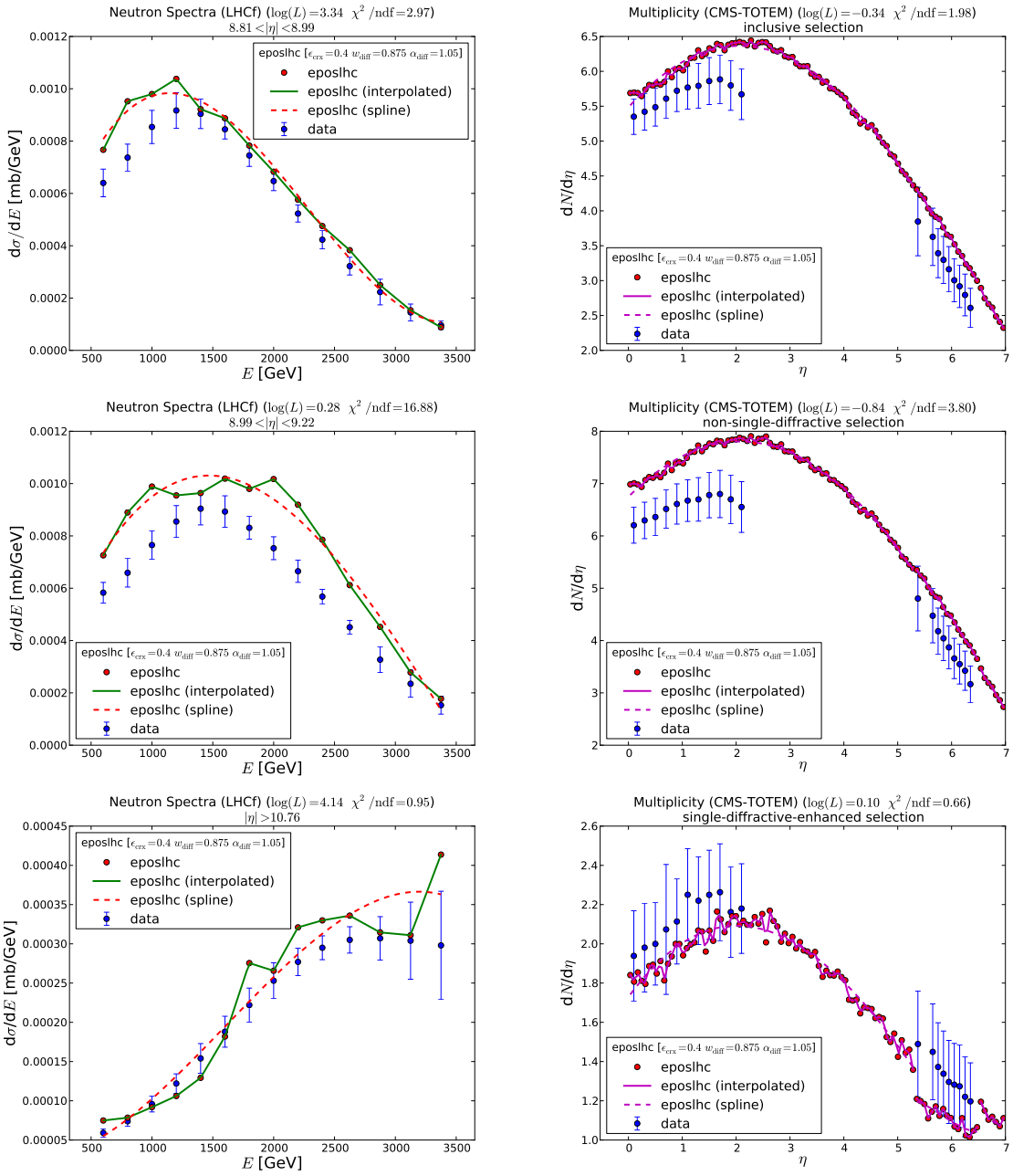


Figure 87: Comparison (part 3) of data and simulation obtained with the default setting of EPOS-LHC. The three plots on the left-hand side show neutron spectra for three different rapidity bins as given by analysis II.ii. The right-hand side shows the multiplicity per unit of pseudorapidity for three different triggers for the multiplicity distribution measured with TOTEM and CMS (analysis II.iii). In the last plot one can see a clear auto-correlation of the trigger at $|\eta| = 5.3$. The fitted univariate spline used for interpolation of the simulated result is shown as a dashed line.

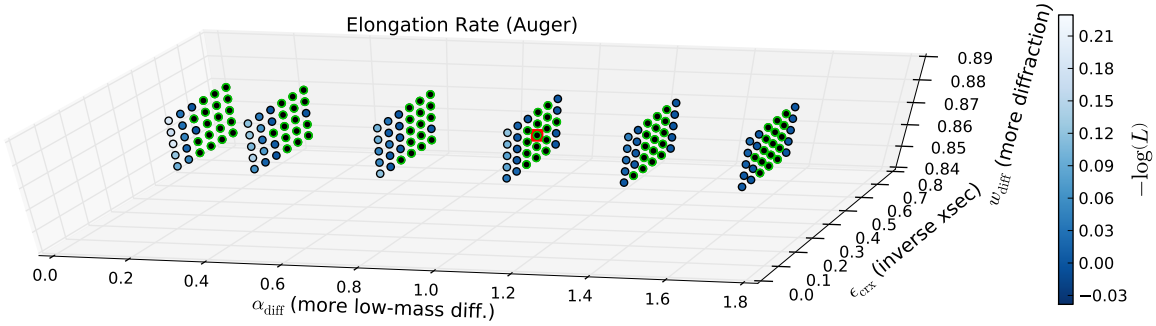


Figure 88: Comparison of the EPOS predictions of all 180 parameter sets to data from the X_{max} measurement of the Pierre Auger Observatory. The colour code of each marker shows the likelihood of data-MC agreement. Darker blue denotes a better agreement whereas the light blue shows stronger disagreement. The parameters with the smallest $-\log(L)$ are marked with a green border and the default parameter set is marked with a red border.

estimator is efficient and it can be shown that when $L(\vec{\theta})$ is maximised, the estimator of $\vec{\theta}$, $\hat{\theta}$, is the best estimator given the data [Met02].

In the following it is preferable to use a function that can possibly be optimised and only has to be summed up when taking different measurements into account. Hence, the values are given in the form of $-\log(L)$.

The likelihood to describe the measurement of X_{max} is for most parameter sets the same (indicated by the multiple green markers in Fig. 88 at $-\log(L) = 0$), since there is only a penalty when the data lies outside the proton or iron predictions. This occurs for the parameter $\epsilon_{\mathrm{crx}} > 0.45$ or $\epsilon_{\mathrm{crx}} < 0.35$. For the latter, this effect is enhanced by a small α_{diff}. Many high-mass diffractive collisions, combined with the large cross section, force the shower to develop more quickly, and, thus, $\langle X_{\mathrm{max}} \rangle$ decreases. It is interesting that w_{diff} has no effect in the chosen range of parameter value.

The proton-proton inelastic cross section is described very well (Fig. 89). This is not at all surprising since the result is considered in the tuning of the EPOS-LHC default parameter set. It is also expected that ϵ_{crx} is the dominating parameter due to its direct connection to the cross section value. From the figure one can deduce that the best parameter set contains the default ϵ_{crx} value but prefers a higher w_{diff}. One necessary remark at this point is that a measurement with the ALFA detector by the ATLAS Collaboration published recently yields a slightly lower inelastic cross section but is not included here.

Figure 90 shows that for the multiplicity measurement the strongest effect comes from ϵ_{crx}. A high value yields a more probable choice for the parameters. For a lower cross section the multiplicity both in the central and forward region decreases for the non-diffractive dominated distributions. The single-diffractive-enhanced multiplicity on the other hand in-

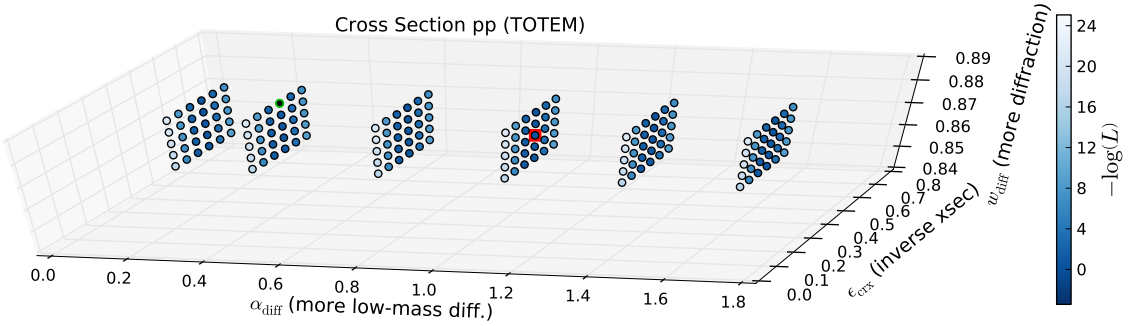


Figure 89: $-\log(L)$ values for the proton-proton cross section measurement. A description of the plot can be found in the caption of Fig. 88.

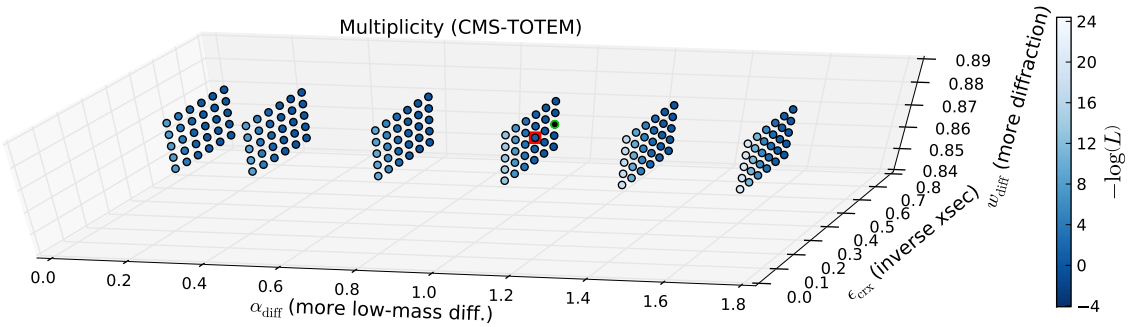


Figure 90: $-\log(L)$ values for the proton-proton cross section measurement. A description of the plot can be found in the caption of Fig. 88.

creases, most likely because the single-diffractive-enhanced sample contains a substantial part of non-diffractive collisions. This effect arises when the multiplicity of the latter contribution decreases and, hence, a greater fraction of the events pass the selection criteria of having no charged particle in one of the two forward acceptance regions.

The spectra of ultra-forward neutrons measured by the LHCf experiment (Fig. 91) show that the likelihood increases for larger ϵ_{crx}. The other parameters have little effect which is again a surprising result. In fact a closer look reveals that, e. g., larger w_{diff} values are preferred in general but the overall best fit is achieved by the default values for α_{diff} and w_{diff} and 0.5 for ϵ_{crx}.

The production cross section in proton-lead collisions (Fig. 92) is for most parameters well described. To introduce the smallest possible bias, the unextrapolated single-arm and double-arm cross sections are compared to the model prediction (cf. Table 12). Only few

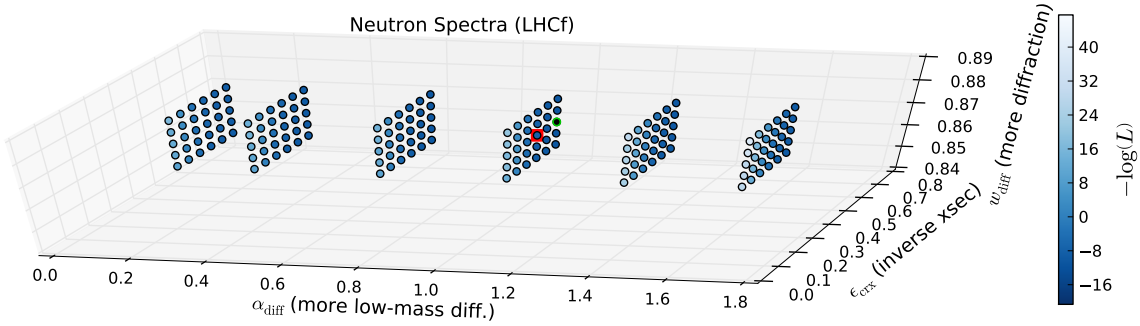


Figure 91: $-\log(L)$ values for the neutron spectra. A description of the plot can be found in the caption of Fig. 88.

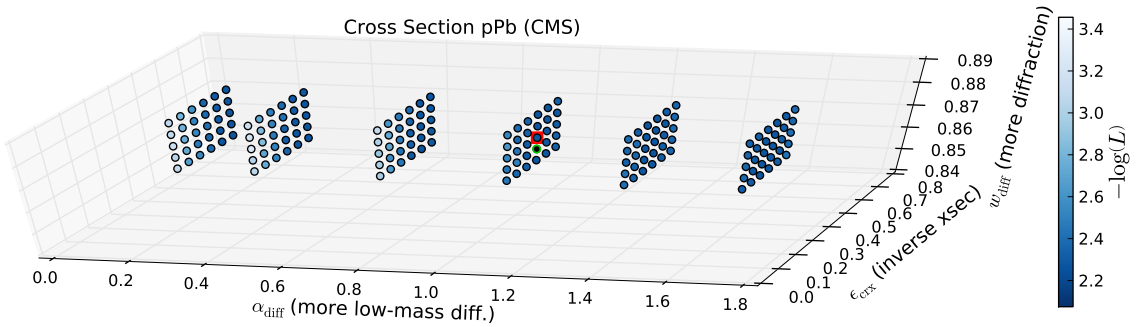


Figure 92: $-\log(L)$ values for the proton-lead cross section measurement. A description of the plot can be found in the caption of Fig. 88.

diffractive events contribute to these selections resulting in a robust measurement towards parameters influencing diffraction. For the most likely parameter set, ϵ_{crx} is very close to the central value of the measurement. Due to large uncertainties on the measured cross sections, deviations from this parameter are still acceptable and change the likelihood only slightly.

Arguably, the most relevant analysis for these parameters are the measured diffractive cross sections by the CMS Collaboration (Fig. 93). Only a small fraction of the parameter set is marked in a dark blue colour, suggesting many of the α_{diff} values that are allowed by the other measurements are excluded here. The preferred α_{diff} is clearly 1.05 but a finer grid could reveal a more exact result. In this plot the single- and double-diffractive cross sections are already combined with the differential cross sections as a function of ξ. The single-diffractive cross section prefers either $\epsilon_{\text{crx}} \simeq 0.5$ or small α_{diff} values of around 0.4. Both are in tension with other measurements not least of which with the ξ-distributions of

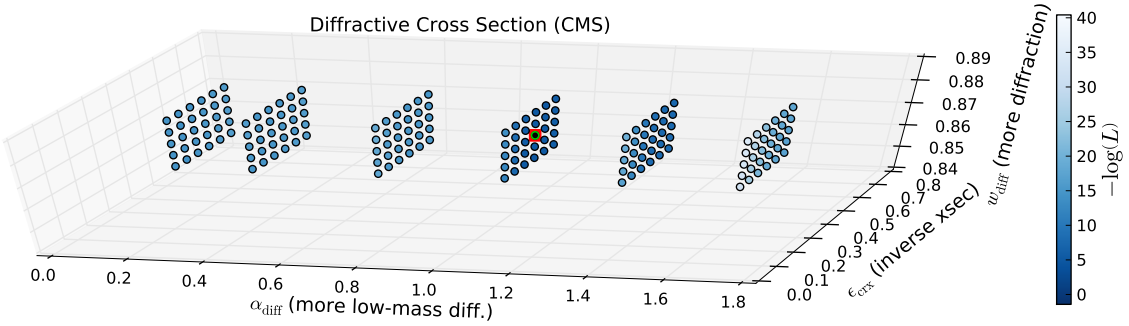


Figure 93: $-\log(L)$ values for the diffractive cross sections and ξ-distributions. A description of the plot can be found in the caption of Fig. 88.

the same analysis. The double-diffractive cross section is, however, too low in the default EPOS-LHC with 2.2 mb compared to 5.2 mb. Only by using a very extreme parameter set ($w_{\mathrm{diff}} = 0.885$, $\alpha_{\mathrm{diff}} = 1.65$, and $\epsilon_{\mathrm{crx}} = 0.25$) a cross section over 5 mb can be reached.

The differential single- and double-diffractive cross sections are also not described perfectly (top panel of Fig. 86). Even though in absolute values the default parameter set is close to the data, the shape does not agree well and cannot match data by any of the tested parameter sets. Influence on the shape might be hidden in parameters not accessible to this study due to simulation time constraints.

Another approach to solve this problem would be to change the underlying physical model of diffraction. A hint for the necessity is given by Appendix A of [Kha+15], which shows that outside of the measurable ξ-range, the differential cross section sees a strong peak at $\lg \xi < -7$, suggesting a possible fault in the model that cannot be refuted by measurements that are limited to a pseudorapidity coverage $|\eta| < 6.6$.

4.3.1.5 *Sensitivity of Model Parameters To LHC and Air Shower Measurements*

In the previous section valuable insights on how different observables project changes in the parameter set of EPOS are gained. For builders of hadronic interaction models the inverted question is also of interest: which analysis can constrain a model parameter in the best possible way?

Let us first define a variable to project the 3-dimensional parameter space into one dimension by studying a single parameter and averaging over the remaining dimensions.

Assuming $\mathcal{P}_{\mathrm{sel}}$ to be one parameter of $[\alpha_{\mathrm{diff}}, \epsilon_{\mathrm{crx}}, w_{\mathrm{diff}}]$ and the remaining two to be $\mathcal{P}_{1,i}$ and $\mathcal{P}_{2,j}$ with i and j representing the simulated points of the parameter space, we can define the average likelihood as

$$\langle -\log L\left(\mathcal{P}_{\mathrm{sel}}\right)\rangle = \frac{1}{N\,M}\sum_{i}^{N}\sum_{j}^{M} -\log\left(L\left(\mathcal{P}_{\mathrm{sel}}, \mathcal{P}_{1,i}\mathcal{P}_{2,j}\right)\right) \tag{98}$$

and the relative likelihood as

$$\text{relative likelihood}\,(\mathcal{P}_{\text{sel}}) = \langle -\log L\,(\mathcal{P}_{\text{sel}})\rangle - \langle -\log L\,(\mathcal{P}_{\text{sel,default}})\rangle\ , \tag{99}$$

where $\mathcal{P}_{\text{sel,default}}$ is the default value of the fixed parameter. In its most general form, this procedure is called "marginalisation". Of course we should keep in mind, that the chosen grid size is currently limited by simulation time. Consequently the average does not cover the full phase-space. The marginalising could be improved by not having fixed parameter points but using a Markov-Chain Monte Carlo algorithm to efficiently sample from the phase-space. This is one possible improvement to further develop this study.

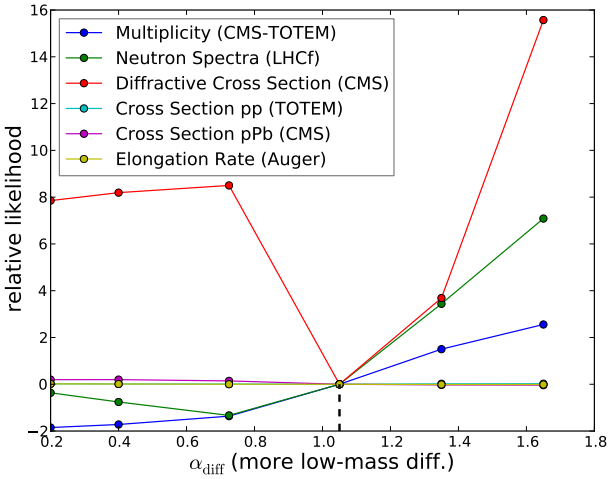
4.3.1.6 *Combining all Analyses*

It is possible to use the marginalising technique to combine the results of many parameters. Figure 94 shows the relative likelihood for each of the three parameters. The size of the influence of a parameter on $\log(L)$ is determined by the slope of the relative likelihood. With arguments from the definition of the likelihood, the slope is determined by the sensitivity of the fixed parameters towards an analysis (represented by lines). Another feature of this representation is that all relative likelihood values below unity stand for an improvement on the default EPOS-LHC setting.

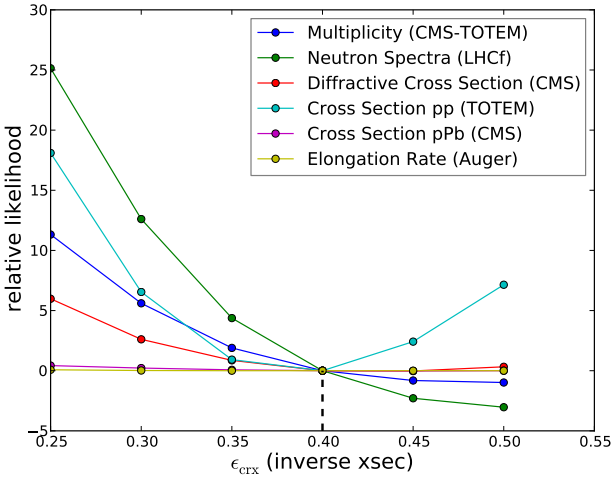
For the different values of α_{diff} the diffractive analysis is the most important one. This plot also shows that a value close to the default value is the preferred choice. Also the multiplicity analysis and neutron spectra can constrain this parameter. However, both prefer a smaller value of α_{diff} in contrast to the diffraction analysis. All mentioned analyses agree that a larger value of α_{diff} is less likely. Other analyses are insensitive to this parameter, mainly because α_{diff} only influences the diffractive mass distribution but not the cross section per se. The constraint derived from the X_{max} measurement is small. On the other hand, for the predecessor of EPOS-LHC, EPOS1.99, the diffractive parameter is $\alpha_{\text{diff}} = 0.45$ for which the diffractive analysis gives a very high relative likelihood value of about 8.

The plot in the middle shows the relative likelihood for changes in ϵ_{crx}. Except for the p-Pb cross section and X_{max} analyses for which no changes are visible over the full parameter range in the relative likelihood; the other analyses agree that values of 0.35 and smaller are unlikely. Higher values than the default of 0.4 are less likely when comparing to data from the cross section p-p measurement and the diffractive cross sections but more likely when comparing to the neutron spectra or the multiplicity measurements. This disagreement cannot easily be explained, but ϵ_{crx} is directly related to the inelastic cross section. It also has been tuned to the measurement of TOTEM at $\sqrt{s} = 7\,\text{TeV}$. It is thus reasonable to rely for this parameter on a direct cross section measurement. Also the p-Pb cross section prefers $\epsilon_{\text{crx}} = 0.4$. The two analyses preferring a higher value of ϵ_{crx} indicate that additional parameters of EPOS must be studied, too.

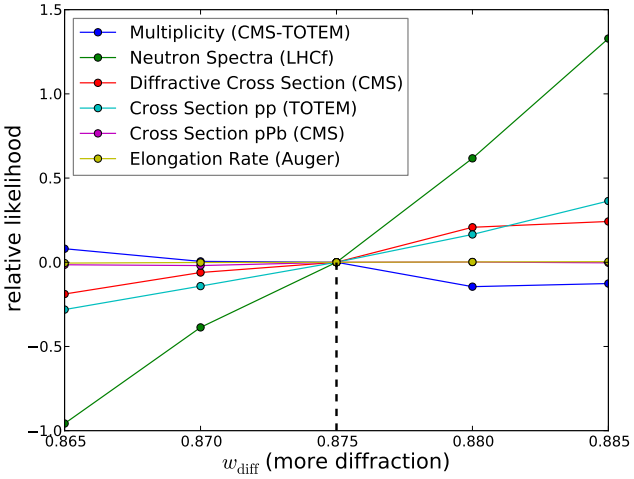
The bottom plot of the figure shows the relative likelihood for changes in w_{diff}. Here the values of the relative likelihood (between -1 and 1.5) suggest, that the parameter has less



(a) Diffractive mass parameter



(b) Saturation scale of the cross section



(c) Probability of diffractive collisions

Figure 94: Sensitivity of the likelihood to changes in one parameter (P_{sel}) while averaging over the parameter grid in the other dimensions as defined in Eq. (99). The y-axis shows the relative likelihood. The default value is marked with a dashed vertical line.

influence on the simulation result. The neutron spectra are influenced the strongest. At 0.865 they have a relative likelihood of -1 which goes up to 1.3 at 0.885. In general the other analyses follow a similar trend but less pronounced. Only the multiplicity measurement find 0.880 and 0.885 to be more likely with a relative likelihood of -0.1. To reach sensitivities comparable to other parameters, w_{diff} would need to be changed to 0.625 or to an unphysical value of 1.125 assuming a linear trend in the neutron spectra measurement. Private communication with one of the authors revealed that neither of these changes are reasonable.

It can be concluded that α_{diff} and ϵ_{crx} are by far more sensitive to the chosen studies. There are clear signs that parts of the parameter space are excluded by the measurements, e. g., smaller ϵ_{crx} and larger α_{diff} than their respective default values.

4.3.2 IMPLICATIONS ON THE PROTON-LEAD CROSS SECTION

The p-Pb cross section is measured first in the visible acceptance, σ_{had}, and is then corrected with simulations to the ideal acceptance, σ_{prod} (Section 4.1.3). The extrapolation from σ_{had} to σ_{prod} amounts to 126 mb (187 mb) for the single-arm (double-arm) selection. The uncertainties related to the extrapolation depend mainly on the modelling of diffractive collisions (Section 4.2). Some models, in particular analytic descriptions that do not generate events on particle level, can only benefit from the fully extrapolated cross section. The uncertainties on the measurements of σ_{had} are smaller compared to σ_{prod} even though the dominating uncertainty of 3.5 % from the luminosity determination remains.

It is interesting to study the change of σ_{prod} with modified parameters. Figure 95 shows its sensitivity to the parameter set and reveals that it is only sensitive to changes in ϵ_{crx}. The dependence on α_{diff} is small. This parameter only changes the characteristics of diffractive collisions propagating as a mild effect to the calculation of σ_{prod} itself. On the other hand, the parameter w_{diff} influences the screening correction but also here only a small effect is visible. This is most likely a result of the insufficient parameter range of w_{diff} that is also observed for other measurements.

One can define the nuclear modification factor

$$R_{\text{p-Pb}} = \frac{\sigma_{\text{prod}}(\text{p-Pb})}{\sigma_{\text{prod}}(\text{p-p})}\,, \tag{100}$$

where the denominator and the nominator are evaluated at $\sqrt{s_{NN}} = 5020\,\text{GeV}$. The factor $R_{\text{p-Pb}}$ isolates the influence of the nuclear effect, in other words, how the model applies the GRT-based calculations. Looking at the dependence of the three parameters, one can determine in which way the parameters enter the calculations. Diffractive parameters are expected to change the magnitude of shadowing and anti-shadowing effects. Again, Fig. 96 shows that only ϵ_{crx} has an influence on $R_{\text{p-Pb}}$.

Lastly, we check how the extrapolation of σ_{had} is influenced (Fig. 97). Here, the effect is as anticipated: changing α_{diff} to smaller values, increases the amount of low-mass compared

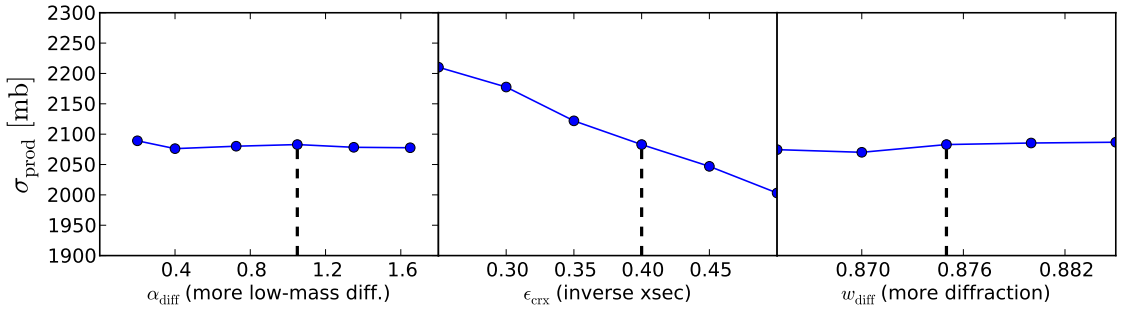


Figure 95: Influence of parameter changes in EPOS-LHC on σ_{prod}. The default values are marked with dashed vertical lines.

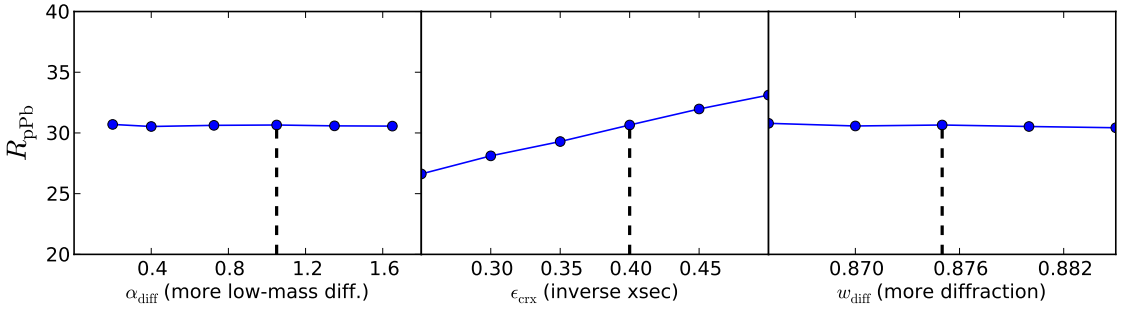


Figure 96: Influence of parameter changes in EPOS-LHC on the nuclear scaling where $R_{\text{p-Pb}}$ is defined as the ratio between the proton-nucleus and proton-proton cross sections. The default values are marked with dashed vertical lines.

to high-mass diffractive states. For low-mass diffractive states, particles are produced with high rapidity and the events have a small probability to be detected in forward detectors. Often the states cannot be distinguished from elastic collisions and will not be identified as an inelastic event by the HF detector. The extrapolated part therefore increases. The necessary extrapolation increases by about 50 mb from 0.25 to 1.65. When ϵ_{crx} is increased the necessary extrapolation decreases. From $\epsilon_{\text{crx}} = 0.25$ to 0.5 it decreases by about 100 mb. Here, the fraction of diffractive events stays roughly the same but when the cross section increases, a—in absolute value—larger part of the cross section also grows. In the range of simulated w_{diff} parameters only a slight trend towards larger necessary extrapolation is visible when the probability of diffractive collision increases. The change is, however, less than 20 mb.

The systematic uncertainty on the measurement of the production cross section is estimated to be 0.9 % to 1.9 % (Table 11), which corresponds to 19 mb to 39 mb. Taking into account the fact that the parameter space especially of ϵ_{crx} and α_{diff} is very constraint by

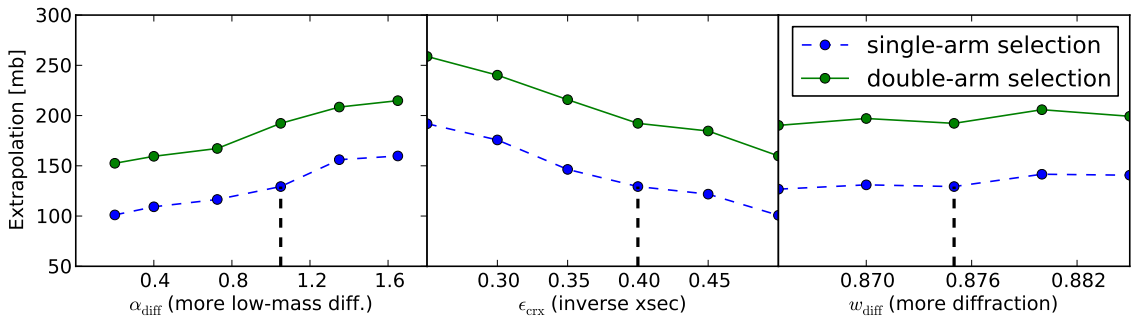


Figure 97: The extrapolation to the full production cross section, $\sigma_{\mathrm{prod}} - \sigma_{\mathrm{vis,had}}$, is shown for the two selections used in Section 4.1. The default values are marked with dashed vertical lines.

data, these values seem to be realistic estimates of the systematic uncertainty of the extrapolation.

5

SUMMARY

> When I was young I was happy to see the pieces of a lengthy calculation cancel and produce a zero result. This told me that I had been smart and had not made a mistake. Only later did I recognise that this was stupid: a good physicist should know a priori that the answer will be zero.
>
> Vladimir Gribov

This thesis is devoted to measurements in the forward phase-space at the LHC and the study of their implications on the modelling of hadronic air showers of ultra-high-energy cosmic rays. To achieve this, hadronic interaction models, as they are used in air shower simulations, are integrated into the workflow of experiments at the LHC. The first step in this mutually beneficial undertaking was the development of the CRMC program. CRMC allows to access all cosmic ray hadronic interaction models in a standardised way. The application of the program helped in many ways:

- It produces different output formats that are commonly used in particle physics and is, thus, easy to integrate in experimental frameworks. This enables the use of the models in detector simulations, making direct comparison to reconstructed quantities possible.
- For the CMS Collaboration, the author of this thesis took over the task of being one of the "heavy ion reconstruction contacts". The contact person is responsible for all event generators related to heavy ion simulation production. This involves updating, validating, and integration of new model parameters for changing LHC run conditions. Within this work, CRMC was integrated into the CMS framework.
- Many simulations have been produced with this tool, which turned out to be invaluable in particular for proton-lead collisions. Besides the models included in CRMC, only one other generator in the CMS framework can describe the proton-lead system.

CRMC is released publicly and is also available in the Generator Service [Heg+14] package with the aim of increasing the visibility to all CERN users. Already now the program is used by almost all experiments at the LHC. A publication in a scientific journal documenting the details of CRMC is planned for the near future.

One of the largest contributing groups to the CASTOR detector operation is from the Karlsruhe Institute of Technology. In this context, many tasks were carried out to improve the detector performance. For the CASTOR physics programme, the proton-lead data are highly valuable. During data taking, the author provided support to the shift team as one of the detector experts. In addition, the alignment of CASTOR has been improved. The ramp-up of the CMS magnet moves the two CASTOR halves by about 1 cm in the transverse direction, which introduces a systematic uncertainty of 16 % on the energy calibration. In this work, a procedure to significantly reduce this uncertainty was developed:

- Contact and non-contact distance sensors that are mounted on the detector and are designed to monitor the installation process, are exploited to extract a position during data taking conditions. The information is combined with a survey measurement that derives the position relative to an external reference point. The CASTOR position has been determined with about 3 mm accuracy, which reduces the uncertainty by about a factor of 3.

- An improvement of the method has been proposed that implies a new calibration technique of the sensors. Presently, this calibration has already been performed as part of another PhD thesis.

- A new detector geometry was implemented in GEANT4 in the CMS framework. This allows to align the detector halves independently. At the same time multiple corrections to the old geometry have been made, e. g., removing overlaps of volumes.

- A bachelor thesis was supervised for which further improvements to the geometry have been implemented. This includes a new geometry of the corners of the active material resulting in a 7 % decrease of the response in the simulation of minimum bias collisions at $\sqrt{s} = 13\,\text{TeV}$.

The CASTOR and the Hadronic Forward (HF) detectors are employed in the analysis of proton-lead collision data. The main part of this thesis is dedicated to the determination of the production cross section in order to test theories describing nuclear effects as they are also relevant to cosmic ray interactions. The problems in deriving such quantity from theory are discussed in detail and advantages of measuring the production cross section for the cosmic ray community are presented. The measurement has been performed as follows:

- All strategies available to the CMS experiment to derive the cross section require precise knowledge of the LHC beam luminosity. Therefore, the luminosity determination using the HF detectors is accurately calibrated by the van-der-Meer scan method. The details of the analysis are outlined, for which the author contributed in many parts: calculation of beam-beam deflection, determination of the charge of the beams, length-scale calibration, and more. The luminosity calibration is determined with an uncertainty of 3.5 % in proton-lead collisions. The calibration is also obtained

for proton-proton collisions recorded in 2013. Both results also are available as a public document [CMS14d].

- The cross section, σ_{prod}, is determined in a counting method that selects events with particle production based on the signal in the HF detector.
- The event counting is corrected for a contribution from electronic noise in the detector. Collisions of a proton with a photon emitted from the nucleus require an additional correction that can be calculated accurately in theory and is explicitly subtracted from the measurement. The final result is presented as a visible hadronic cross section with negligible uncertainties from diffractive processes and as a production cross section fully corrected with simulations. The latter is found to be $\sigma_{\text{prod}} = 2061 \pm 3\,(\text{stat}) \pm 34\,(\text{syst}) \pm 72\,(\text{lumi})\,\text{mb}$.
- The results are compared to theory predictions from air shower models and from a standard Glauber calculation. The DPMJET and QGSJET models overshoot the data slightly, while EPOS-LHC agrees within systematic uncertainties. The Glauber calculation reveals that shadowing and antishadowing effects must be small or cancel each other. A net effect as large as $+10\,\%$ $(-10\,\%)$ as predicted for proton-air collisions can be excluded with a confidence level of 3.5 (1.8)-standard deviations. This important result is available as an analysis summary [CMS13b] and is currently in the last stages of the internal review within the CMS Collaboration, after which a publication as a journal article will follow.
- Furthermore, diffractive events in proton-lead collisions are studied with the CASTOR detector. On one hand this has direct implications on the extrapolation of σ_{prod} and on the other hand this is of immense interest to cosmic ray physics since nuclear effects in diffractive collisions could never before be studied at these energies. The data analysis reveals that no major deviations of the medium-mass diffraction as predicted by the air shower models, EPOS and QGSJET, is identified. For HIJING and DPMJET the results indicate that diffraction is modelled incompletely.

The final part of the thesis is devoted to a novel method of combining air shower and collider measurements to constrain hadronic interaction models. In this global analysis, the hadronic interaction model EPOS is modified by changing three of its parameters that predominately influence the diffractive mass distribution, the probability of diffraction, and the cross section. Simulations are performed for a phase-space of 180 different parameter sets and are then compared to selected measurements. It was possible to derive the sensitivity of the data to changes in the model parameters. Among others, the following results were obtained:

- Each parameter is best constrained by different measurements and its optimal value is derived. In some cases tension between measurements are found.
- Air shower measurements can constrain the model parameters and more analyses need to be included in the future.

- From the constrained parameters the conclusion is drawn that the systematic uncertainties on σ_{prod} are well estimated in this thesis.

Due to the combination of air shower and collider measurements this approach opens a completely new field of study. Many improvements, including Markov-Chain Monte Carlo sampling, are proposed to further clarify the interpretation of the results. However, already in the current state, the immense value of forward phase-space measurements in improving modelling of hadronic interaction physics is demonstrated.

A

APPENDIX

A.1 SAMPLING THE BREIT-WIGNER DISTRIBUTION USING A RANDOM NUMBER GENERATOR

Inverse sampling or cumulative distribution function (CDF) sampling is a commonly used method to obtain a distribution according to the associated probability density function (PDF) from uniformly distributed real numbers $U \in (0,1)$ (see e. g., [Gen03]). The idea of the method is depicted in Fig. 98. Here, the PDF of a Breit-Wigner distribution or Cauchy distribution is used as an example to derive the method. The PDF is given by

$$f(x; x_0, \gamma) = \frac{1}{\pi\gamma \left[1 + \left(\frac{x - x_0}{\gamma}\right)^2\right]} \tag{101}$$

and the cumulative distribution function (CDF) is

$$F(x; x_0, \gamma) = \int_{-\infty}^{x} \mathrm{d}x' \, f(x'; x_0, \gamma) = \frac{1}{\pi} \arctan\left(\frac{x - x_0}{\gamma}\right) + \frac{1}{2}, \tag{102}$$

with two parameters, γ and x_0. The CDF is a monotonic function with the properties of being 0 for $x \to -\infty$ and reaching a maximum of 1 for $x \to \infty$. A relationship between the inverted CDF and the uniform random number U yields

$$F^{-1}(U) = \gamma \tan\left[\pi\left(U - \frac{1}{2}\right)\right] + x_0 \,. \tag{103}$$

The resulting distribution will follow the PDF, $f(x)$. This is shown by the fact that the inverse of the CDF, $F^{-1}(U)$, has the same cumulative distribution function as $f(x)$,

$$P(F^{-1}(U) \leqslant x) = P(U \leqslant F(x)) = F(x) \,, \tag{104}$$

where P denotes probability. Consequently, Eq. (103) can be used to generate a Breit-Wigner distribution.

A.2 ADDITIONAL INFORMATION ABOUT THE LENGTH SCALE EVALUATION FOR VAN-DER-MEER SCANS IN PROTON-LEAD COLLISIONS

At the CMS interaction point, the beam positions controlled by the LHC magnets that are calibrated with data from reconstructed vertices using CMS pixel and silicon strip trackers.

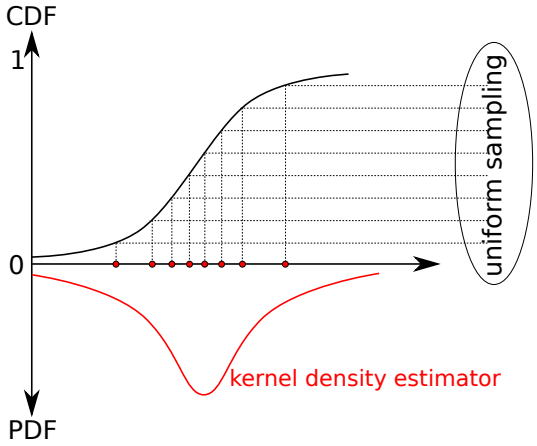


Figure 98: Illustration of the idea behind inverse transform sampling.

While it is true that the vertex reconstruction provides a resolution of the same order of magnitude than the size of the beam shifts in x and y, the relative shifts can be determined very accurately. For this method, referred to as length scale calibration, both projectile and target beams are shifted by the same amount in the X-direction, ΔX, and Y-direction, ΔY, in order to move the average position of the primary collision vertex. Since the relevant beam shifts are of the order of µm, the relative change between the previous vertex position and the current vertex position given derived from tracking is assumed to be exact. Calibration factors are derived by comparing the result to the shifts intended by the magnet setting. The caveat of this method is the limited statistics of around 30 s of data acquisition for each beam position.

Length scale data were taken for p-p and for p-Pb collisions. Here, the analysis of the p-Pb calibration is discussed, for which the data were taken in fill 3503 run 210986. The order of the vertex positions (in µm) for the two directions is given by

$$\Delta X: \; -100,\, -50,\, 0,\, +50,\, +100,\, +50,\, 0,\, -50,\, -100,\, 0 \tag{105}$$

and

$$\Delta Y: \; +100,\, +50,\, 0,\, -50,\, -100,\, -50,\, 0,\, +50,\, +100,\, 0\,. \tag{106}$$

Four identical and unbiased triggers were responsible to distribute the events to four data sets at the maximum output rate: HLT_ZeroBias_part1_v1, HLT_ZeroBias_part2_v1, HLT_ZeroBias_part3_v1, and HLT_ZeroBias_part4_v1. At about 500 Hz per trigger, the number of events available is still low (statistical uncertainty of the same order as systematic uncertainty). Henceforth, the only requirement for the event selection is a reconstructed vertex from tracking.

Each change in position is identified in the data by a drop in the rate of selected events. When changing to a new position the beams are adjusted sequentially in X and Y and as soon as their distance increases, a drop compared to the nominal rate is visible. The drops

are identified by going below 0.5 of the median of the rate over all events. Sections in orbit numbers corresponding to the ΔX and ΔY values are identified by this. An additional time interval of 2 s is removed from the end of each section in order to remove biases due to beam movements. From the front of the section, a larger time period is subtracted to allow the beams to properly adjust to the new position. To determine this interval, the χ^2 of a Gaussian fit to the distribution of vertex position is studied. In general, longer skipped intervals result in a lower χ^2 in each section (see Fig. 99). Not all sections comprise the same total length. To achieve good χ^2-values but still skip less time then the shortest section, the time interval of 10 s is skipped at the beginning of each section. This is illustrated in Fig. 100.

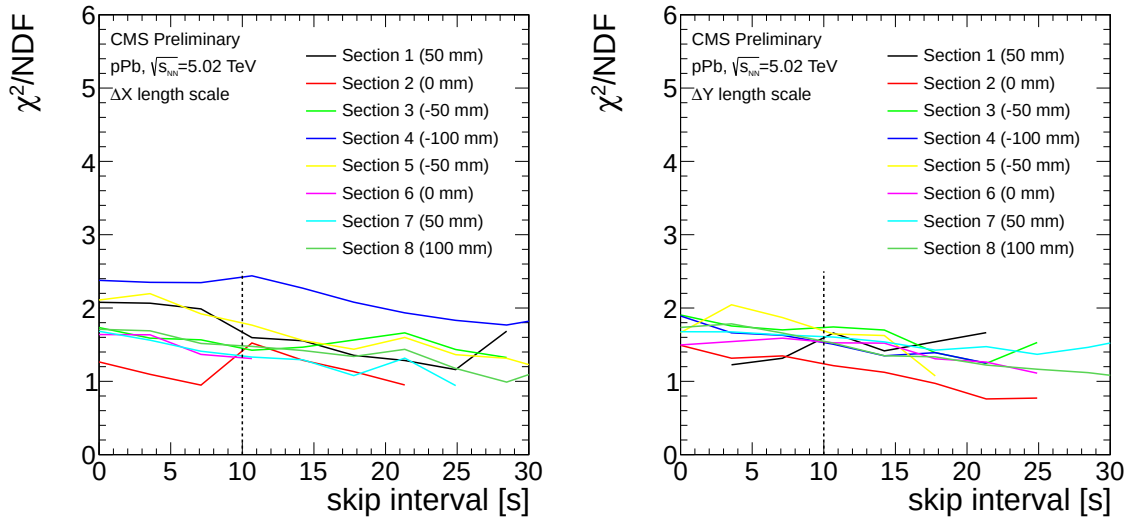


Figure 99: The χ^2 values of the Gaussian fits of different sections as defined in Fig. 100 plotted as a function of the time skipped at the beginning of each section.

In each section, the vertex positions are weighted by $1/\sigma_V$, where σ_V is the uncertainty on the reconstructed vertex position (for more information on the vertex reconstruction, see [Cha+14a]). The distributions of the weighted positions for each section are shown in Fig. 101 and are fitted by a Gaussian function. To avoid hysteresis effects, the two direction going from higher-to-lower and going from lower-to-higher $\Delta X/Y$ positions are treated separately and for each direction the first data point is ignored. Subsequently, the mean and the width of the Gaussian functions are compared to the LHC reference positions, where their relation is well described by a first-order polynomial. The result is shown in Fig. 61 on Page 104.

Table 14 lists the slope values of the fitted polynomials for two directions in X and Y, which are the correction factors that need to be applied to the positional shifts as set by the LHC magnets. The two directions of shifting the beam positions are combined to one value by taking their average.

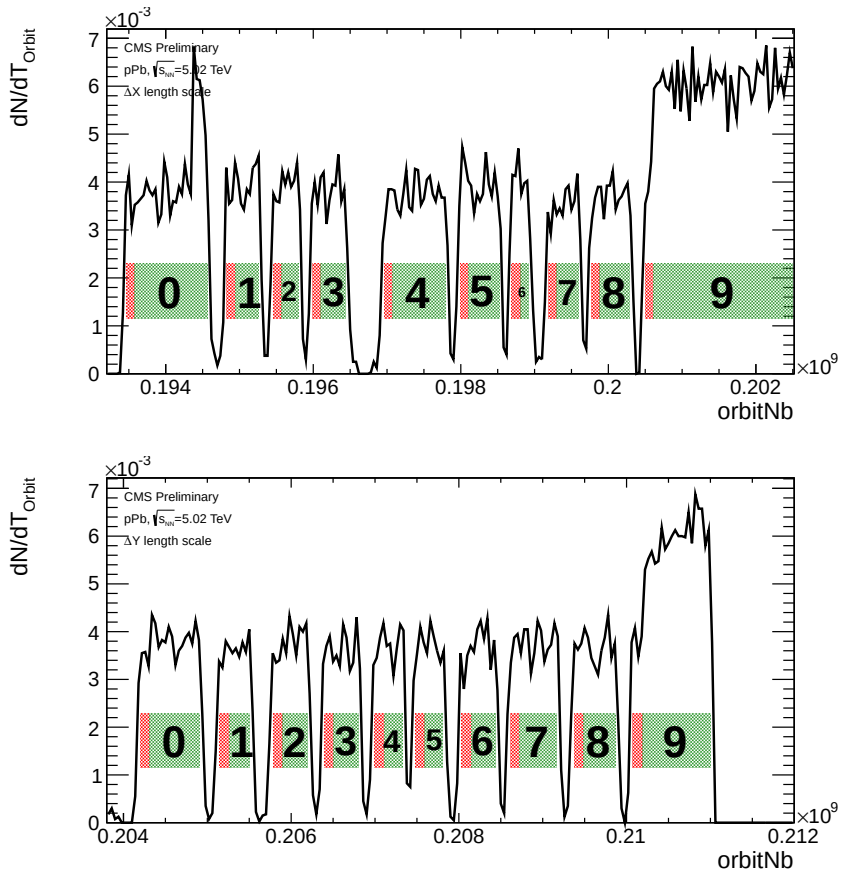


Figure 100: Identified sections for different beam positions in the data for ΔX (top) and ΔY (bottom). Red corresponds to the skipped period whereas green shows the accepted.

Table 14: Correction factors determined by the slope of the linear fits to length scale measurement for the p-Pb. The values are given in units of [µm/µm].

Correction factor	ΔX	ΔY
low-to-high	0.977	0.992
high-to-low	0.992	0.992
average	0.984	0.992

Three different sources of systematic uncertainties are taken into account. Skipping longer time at the beginning of each section reduces the χ^2 and therefore increases the fit quality. The change of the slope values is additionally evaluated with skipping 5 or 15 s. The full spread of the slope values is taken as a systematic uncertainty. Secondly, the

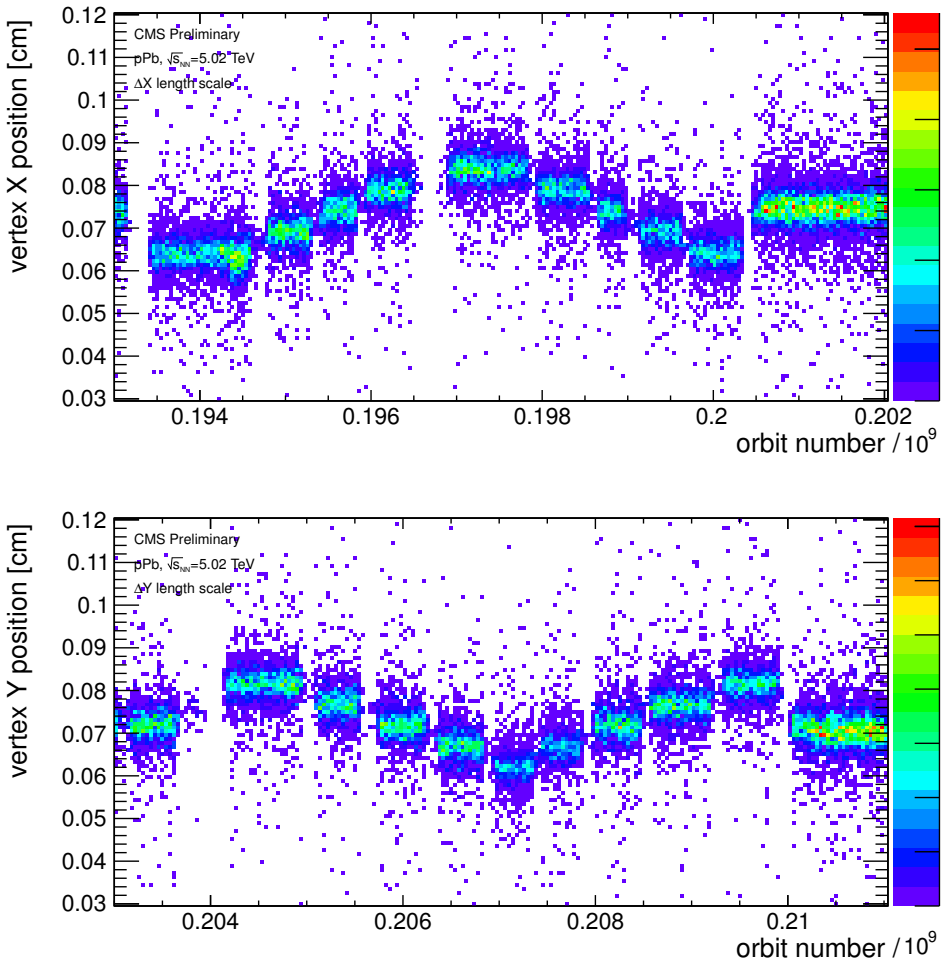


Figure 101: Vertex positions for X (top) and Y (bottom) as a function of orbit number, which can be translated to time by orbit number/11 246 Hz.

distribution of the vertex positions is not completely Gaussian. The effect can be assigned to either the beam density or a bias in tracker data. Here, the full difference to the mean of the distribution of the vertex positions is taken as a systematic uncertainty. Lastly, the difference between the two scan directions is added as an uncertainty to the slope. All systematic uncertainties are listed in Table 14.

The systematic uncertainties of the length scale corrections are added in quadrature and the correction factors are then

$$c_X = 0.984 \pm 0.005\,(\text{stat}) \pm 0.007\,(\text{syst}) \tag{107}$$

$$c_Y = 0.992 \pm 0.006\,(\text{stat}) \pm 0.004\,(\text{syst}). \tag{108}$$

Table 15: Uncertainties on the correction factors. The values are given in units of [µm/µm].

Source of uncertainty	ΔX	ΔY
Statistical	0.005	0.006
Direction	0.007	0.000
Fitting	0.002	0.003
Hysteresis	0.001	0.003

The factors of both directions are close to unity and their systematic uncertainty is comparable to but slightly higher than their statistical uncertainty.

A.3 ADDITIONAL INFORMATION ABOUT THE PROTON-LEAD CROSS SECTION MEASUREMENT

HF RADIATION DAMAGE CORRECTION

The radiation damage correction factors are listed in Table 16 for each ring of the HF detector. For the determination of the production cross section in p-Pb collisions only the factors of column "C_i (MC)" are used. The other factors are derived from calibration with $Z \rightarrow e^- + e^+$ events and comparing the measured energy of 2011 with 2013 for the same centre-of-mass energy. They are in good agreement with the former factors. Nevertheless, they are only valid for the energy deposit in the long fibres due to limited statistics in the 2013 data.

The factors are applied to each ring. The change in the energy of the towers is shown in Figs. 100 to 103. Whereas Fig. 102 shows the distribution of all tower energies, the others show the distribution of the hottest tower energy as defined in the text.

Table 16: Derived correction factors for HF rings, determined in this analysis are shown in column "C_i (MC)". The column labelled as "C_i (pp energy flow)" lists the result of an independent measurement of the radiation damage that is only valid for long fibres.

	Plus-side		Minus-side	
$i\eta$	$1/C_i$ (MC)	C_i (p-p energy flow)	$1/C_i$ (MC)	C_i (p-p energy flow)
± 40	0.76	0.83$\pm$0.10	0.85	0.75$\pm$0.09
± 39	0.65	0.69$\pm$0.06	0.69	0.61$\pm$0.07
± 38	0.72	0.72$\pm$0.04	0.75	0.67$\pm$0.03
± 37	0.73	0.80$\pm$0.03	0.78	0.83$\pm$0.03
± 36	0.81	0.86$\pm$0.03	0.85	0.85$\pm$0.03
± 35	0.85	0.92$\pm$0.03	0.84	0.84$\pm$0.03
± 34	0.86	0.91$\pm$0.03	0.90	0.92$\pm$0.03
± 33	0.90	0.96$\pm$0.03	0.91	0.87$\pm$0.03
± 32	0.97	0.99$\pm$0.03	0.94	0.89$\pm$0.03
± 31	0.98	0.95$\pm$0.03	0.96	0.91$\pm$0.03
± 30	1.00	0.97$\pm$0.04	1.00	0.94$\pm$0.04

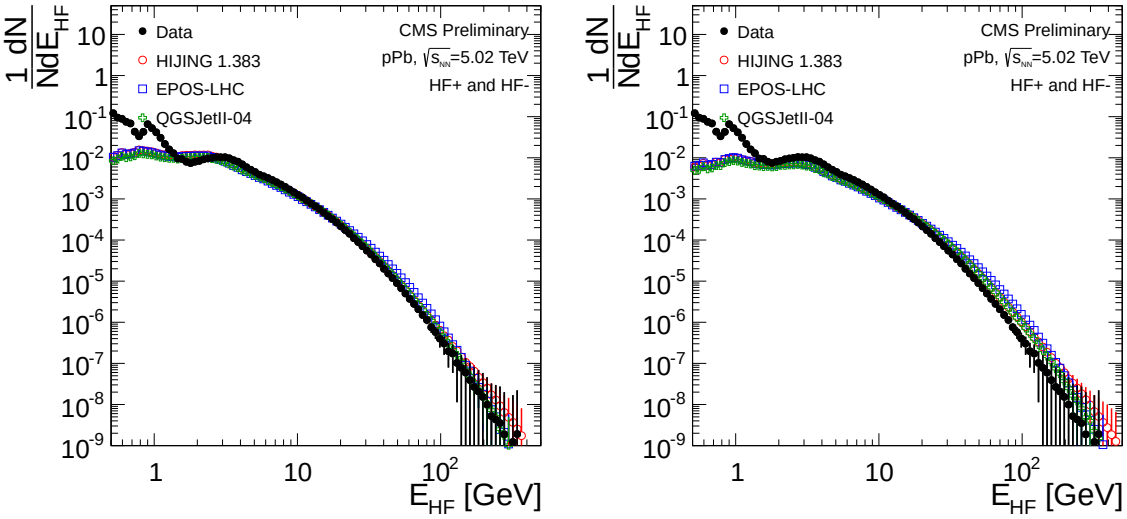


Figure 102: Distribution of tower energies of all towers per event. Left panel: With rescaling applied. Right panel: No rescaling.

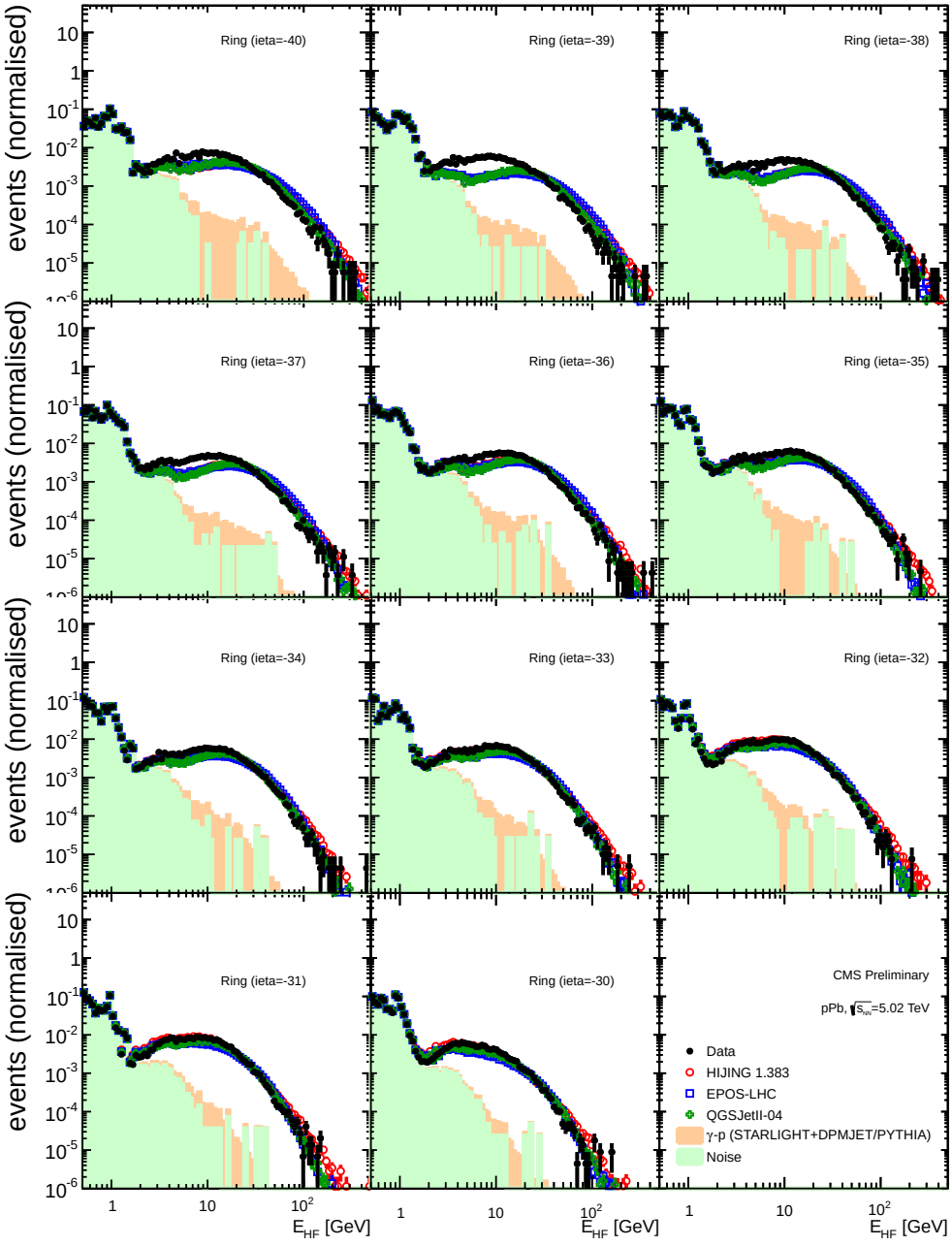


Figure 103: Distribution of E_{HF} with no rescaling applied ($\eta-$ side).

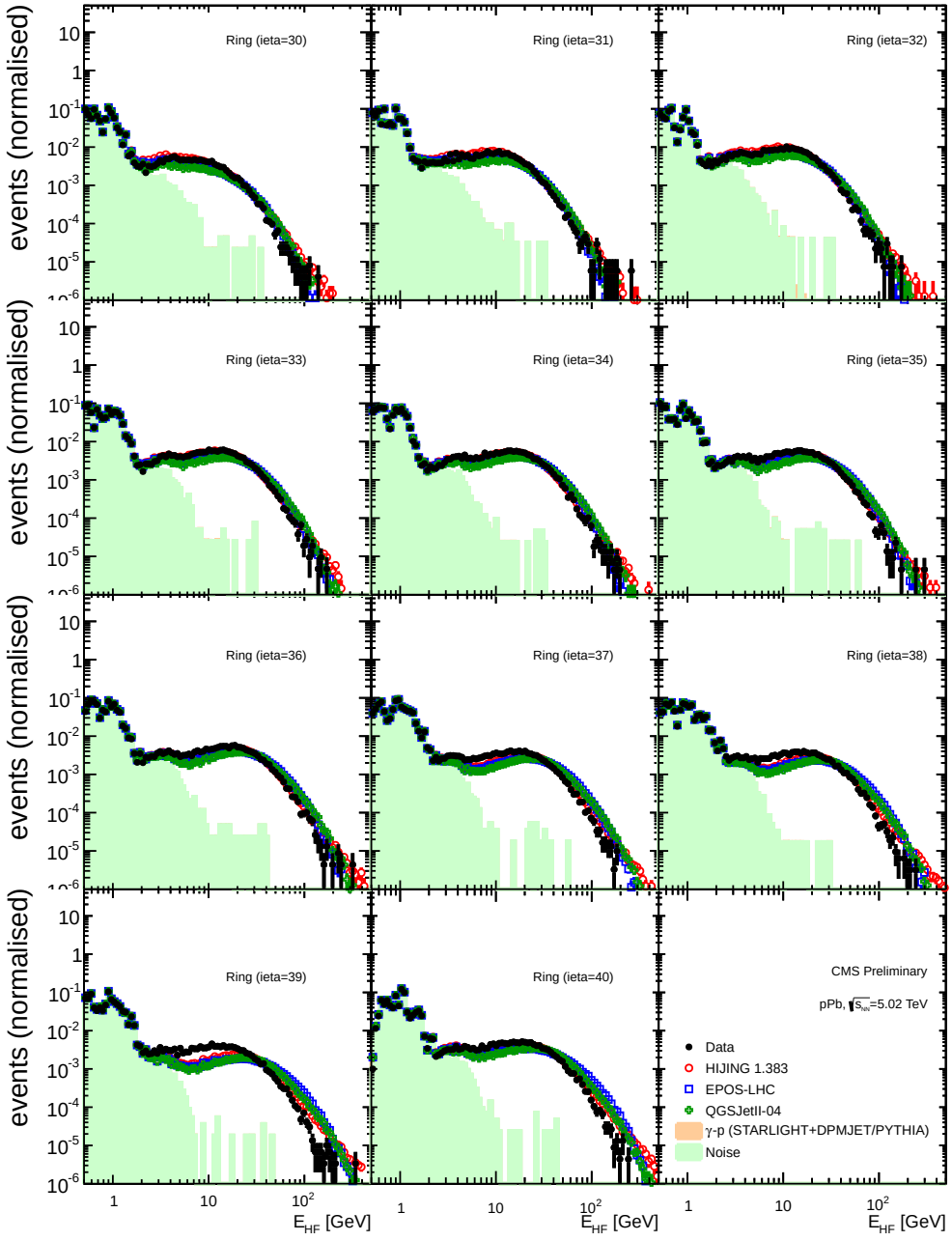


Figure 102: Distribution of E_{HF} with no rescaling applied ($\eta+$ side).

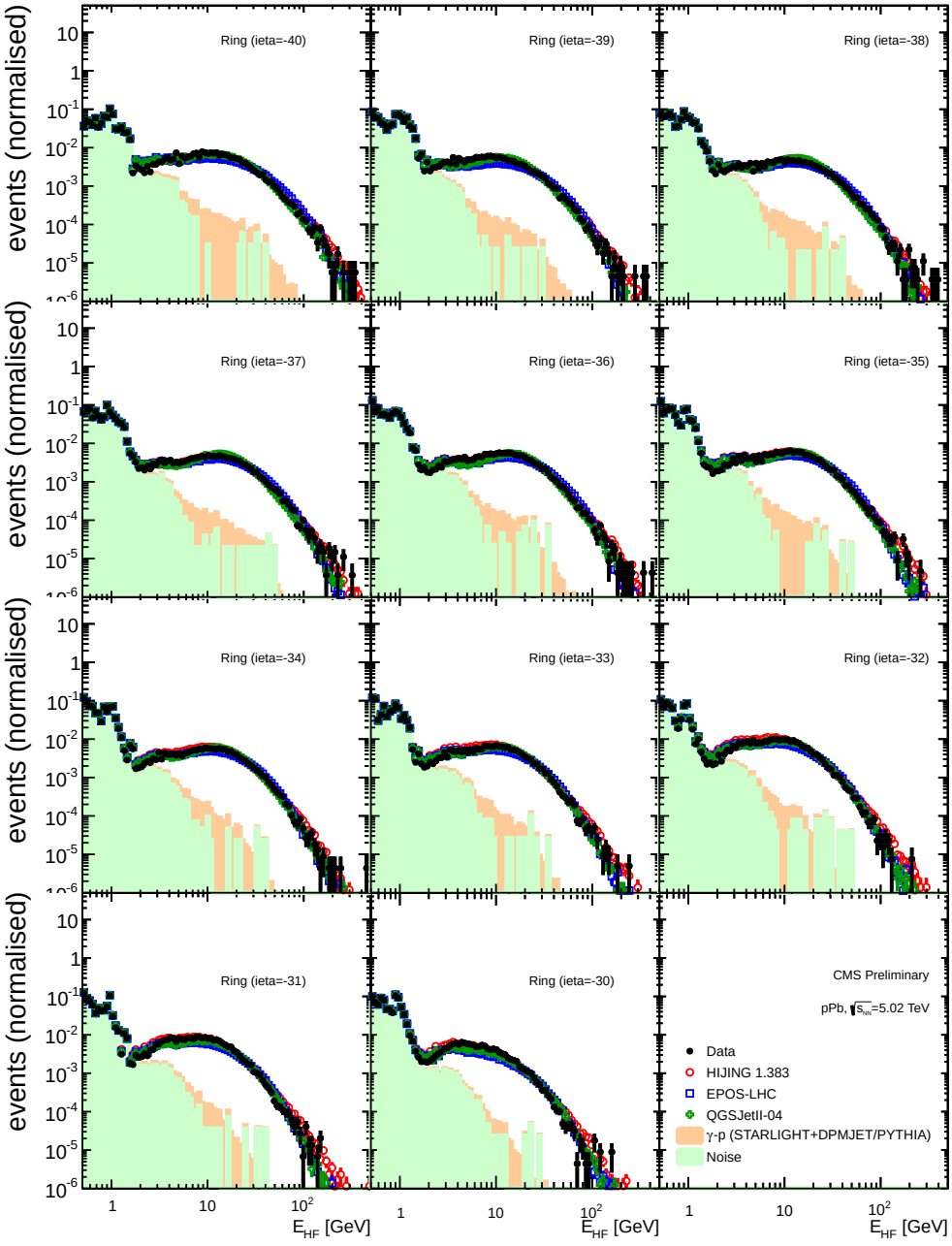


Figure 101: Distribution of E_{HF} with rescaling applied ($\eta-$ side).

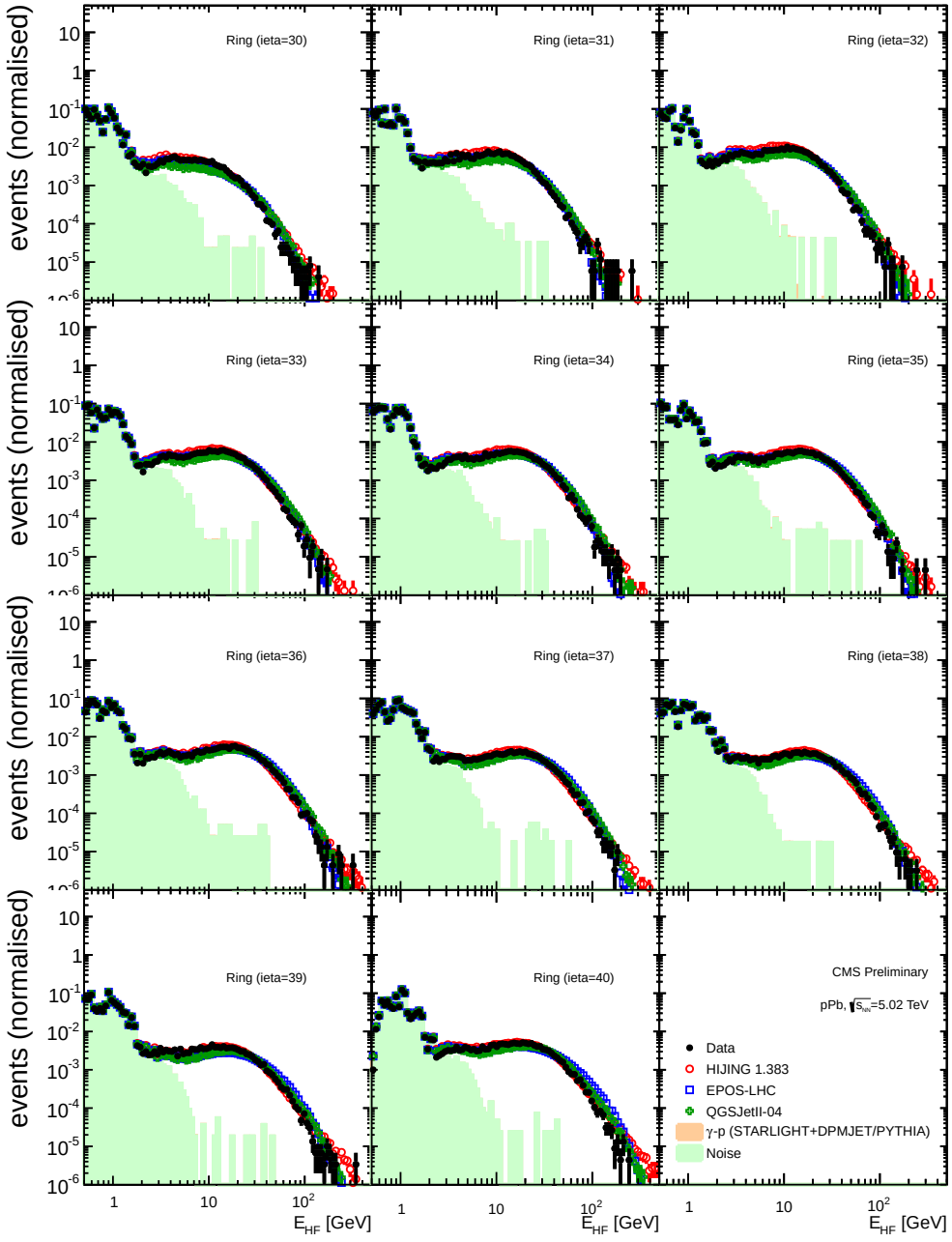


Figure 100: Distribution of E_{HF} with rescaling applied ($\eta+$ side).

GLAUBER CALCULATION

The Glauber calculation to go from the proton-proton cross section to the proton-lead cross section is performed with a MC program. The parameters describing the nuclei are given in Table 17. A systematic uncertainty is assigned to the derived value because not all of the parameters are known with perfect accuracy. Modifying the parameters leads to changes in the p-Pb cross section value as shown in Table 18.

Table 17: Standard parameters for the Glauber calculation to derive the proton-lead cross section, $\sigma_{\text{Glauber,pPb}}(5.02\,\text{TeV})$.

Parameter	Value
A_1	1
Z_1	1
A_2	208
Z_2	82
Woods-Saxon radius lead nucleus	6.624 fm
Skin depth of the lead nucleus	0.546 fm
min impact parameter	0 fm
max impact parameter	20 fm
$\sqrt{s_{NN}}$	5020 GeV
nn-overlap function	black disk
Nucleons with hard core	no

Table 18: The table shows the systematic uncertainty on the cross section calculated by the Glauber model apart from inelastic screening and short-range correlations. Various input parameters are changed and the standard deviation is taken as the uncertainty on the central value determined by the default parameters (first row). "WS" stands for Wood-Saxon nucleon density.

Setting of Glauber Monte Carlo program	σ_{prod} [mb]
default WS, $\sigma_{\text{inel,pp}}$ = 70 mb	2125
default WS, $\sigma_{\text{inel,pp}}$ = 68.5 mb	2116
default WS, $\sigma_{\text{inel,pp}}$ = 71.5 mb	2136
radius: 6.684 fm, skin depth: 0.536 fm, $\sigma_{\text{inel,pp}}$ = 70 mb	2134
radius: 6.564 fm, skin depth: 0.556 fm, $\sigma_{\text{inel,pp}}$ = 70 mb	2112
default WS, hard-core, $\sigma_{\text{inel,pp}}$ = 70 mb	2107
default WS, grey disk, $\sigma_{\text{inel,pp}}$ = 70 mb	2155
default WS, hard-core, grey disk, $\sigma_{\text{inel,pp}}$ = 70 mb	2162
default WS, Gaussian profile for proton, $\sigma_{\text{inel,pp}}$ = 70 mb	2180
Combined	2125 ± 40

BIBLIOGRAPHY

[Aab+13] A. Aab et al. "The Pierre Auger Observatory: Contributions to the 33rd International Cosmic Ray Conference (ICRC 2013)." In: *e-print* (2013). arXiv: 1307.5059 [astro-ph.HE].

[Aab+14a] A. Aab et al. "Depth of Maximum of Air-Shower Profiles at the Pierre Auger Observatory: Measurements at Energies above $10^{17.8}$ eV." In: *Phys. Rev. D* (2014). arXiv: 1409.4809 [astro-ph.HE].

[Aab+14b] A. Aab et al. "Muons in Air Showers at the Pierre Auger Observatory: Measurement of Atmospheric Production Depth." In: *Phys. Rev. D* 90 (2014), p. 012012. DOI: 10.1103/PhysRevD.90.012012. arXiv: 1407.5919 [hep-ex].

[Aab+15a] A. Aab et al. "Muons in Air Showers at the Pierre Auger Observatory: Mean Number in Highly Inclined Events." In: *Phys. Rev. D* 91 (2015), p. 032003. DOI: 10.1103/PhysRevD.91.032003. arXiv: 1408.1421 [astro-ph.HE].

[Aab+15b] A. Aab et al. "The Pierre Auger Cosmic Ray Observatory." In: *Nucl. Instrum. Meth.* (2015). arXiv: 1502.01323 [astro-ph.IM].

[Aad+08] G. Aad et al. "The ATLAS Experiment at the CERN Large Hadron Collider." In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[Aad+12a] G. Aad et al. "Measurements of the Pseudorapidity Dependence of the Total Transverse Energy in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV with ATLAS." In: *JHEP* 1211 (2012), p. 033. DOI: 10.1007/JHEP11(2012)033. arXiv: 1208.6256 [hep-ex].

[Aad+12b] G. Aad et al. "Rapidity Gap Cross Sections Measured with the ATLAS Detector in pp Collisions at $\sqrt{s} = 7$ TeV." In: *Eur. Phys. J. C* 72 (2012), p. 1926. DOI: 10.1140/epjc/s10052-012-1926-0. arXiv: 1201.2808 [hep-ex].

[Aad+12c] G. Aad et al. "Underlying Event Characteristics and their Dependence on Jet Size of Charged-Particle Jet Events in pp Collisions at $\sqrt{s} = 7$ TeV with the ATLAS Detector." In: *Phys. Rev. D* 86 (2012), p. 072004. DOI: 10.1103/PhysRevD.86.072004. arXiv: 1208.0563 [hep-ex].

[Aad+14] G. Aad et al. "Measurement of the Total Cross Section from Elastic Scattering in pp Collisions at $\sqrt{s} = 7$ TeV with the ATLAS Detector." In: *Nucl. Phys. B* 889 (2014), p. 486. DOI: 10.1016/j.nuclphysb.2014.10.019. arXiv: 1408.5778 [hep-ex].

[Aai+13] R. Aaij et al. "Measurement of the Forward Energy Flow in pp Collisions at $\sqrt{s} = 7$ TeV." In: *Eur. Phys. J. C* 73 (2013), p. 2421. DOI: 10.1140/epjc/s10052-013-2421-y. arXiv: 1212.4755 [hep-ex].

[Aai+14] R. Aaij et al. "Measurement of Charged Particle Multiplicities and Densities in pp Collisions at $\sqrt{s} = 7$ TeV in the Forward Region." In: *Eur. Phys. J. C* 74 (2014), p. 2888. DOI: 10.1140/epjc/s10052-014-2888-1. arXiv: 1402.4430 [hep-ex].

[Aam+08] K. Aamodt et al. "The ALICE Experiment at the CERN LHC." In: *JINST* 3 (2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002.

[Aar+10] F. Aaron et al. "Combined Measurement and QCD Analysis of the Inclusive $e^{\pm}$ p Scattering Cross Sections at HERA." In: *JHEP* 1001 (2010), p. 109. DOI: 10.1007/JHEP01(2010)109. arXiv: 0911.0884 [hep-ex].

[Aar+15] M. Aartsen et al. "Atmospheric and Astrophysical Neutrinos above 1 TeV Interacting in IceCube." In: *Phys. Rev. D* 91.2 (2015), p. 022001. DOI: 10.1103/PhysRevD.91.022001. arXiv: 1410.1749 [astro-ph.HE].

[Abb+13] R. Abbasi et al. "All-Particle Cosmic Ray Energy Spectrum Measured with 26 IceTop Stations." In: *Astropart. Phys.* 44 (2013), p. 40. DOI: 10.1016/j.astropartphys.2013.01.016. arXiv: 1202.3039 [astro-ph.HE].

[Abb+14a] R. Abbasi et al. "Indications of Intermediate-Scale Anisotropy of Cosmic Rays with Energy Greater Than 57 EeV in the Northern Sky Measured with the Surface Detector of the Telescope Array Experiment." In: *Astrophys. J.* 790 (2014), p. L21. DOI: 10.1088/2041-8205/790/2/L21. arXiv: 1404.5890 [astro-ph.HE].

[Abb+14b] R. Abbasi et al. "Study of Ultra-High Energy Cosmic Ray Composition Using Telescope Array's Middle Drum Detector and Surface Array in Hybrid Mode." In: *Astropart. Phys.* 64 (2014), p. 49. DOI: 10.1016/j.astropartphys.2014.11.004. arXiv: 1408.1726 [astro-ph.HE].

[Abe+13] B. Abelev et al. "Measurement of Inelastic, Single- and Double-Diffraction Cross Sections in Proton–Proton Collisions at the LHC with ALICE." In: *Eur. Phys. J. C* 73 (2013), p. 2456. DOI: 10.1140/epjc/s10052-013-2456-0. arXiv: 1208.4968 [hep-ex].

[Abe+14a] B. B. Abelev et al. "Measurement of visible cross sections in proton-lead collisions at $\sqrt{s_{NN}}$ = 5.02 TeV in van der Meer scans with the ALICE detector." In: *JINST* 9 (2014), p. 1100. DOI: 10.1088/1748-0221/9/11/P11003. arXiv: 1405.1849 [nucl-ex].

[Abe+14b] B. B. Abelev et al. "Production of Charged Pions, Kaons and Protons at Large Transverse Momenta in pp and Pb-Pb Collisions at $\sqrt{s_{NN}} = 2.76\,TeV$." In: *Phys. Lett.* 736 (2014), p. 196. DOI: 10.1016/j.physletb.2014.07.011. arXiv: 1401.1250 [nucl-ex].

[Abe+94] F. Abe et al. "Measurement of the $\bar{p}p$ total cross-section at $\sqrt{s} = 546$ and 1800 GeV." In: *Phys. Rev. D* 50 (1994), p. 5550. DOI: 10.1103/PhysRevD.50.5550.

[ABG12] R. Aloisio, V. Berezinsky, and A. Gazizov. "Transition From Galactic to Extragalactic Cosmic Rays." In: *Astropart. Phys.* 39-40 (2012), p. 129. DOI: 10.1016/j.astropartphys.2012.09.007. arXiv: 1211.0494 [astro-ph.HE].

[Abr+12] P. Abreu et al. "Measurement of the Proton-Air Cross-Section at $\sqrt{s} = 57\,\text{TeV}$ with the Pierre Auger Observatory." In: *Phys. Rev. Lett.* 109 (2012), p. 062002. DOI: 10.1103/PhysRevLett.109.062002. arXiv: 1208.1520 [hep-ex].

[Abu+12] T. Abu-Zayyad et al. "The Surface Detector Array of the Telescope Array Experiment." In: *Nucl. Instrum. Meth. A* 689 (2012), p. 87. DOI: 10.1016/j.nima.2012.05.079. arXiv: 1201.4964 [astro-ph.IM].

[Abu+13] T. Abu-Zayyad et al. "Pierre Auger Observatory and Telescope Array: Joint Contributions to the 33rd International Cosmic Ray Conference (ICRC 2013)." In: *e-print* (2013). arXiv: 1310.0647 [astro-ph.HE].

[Adr+12] O. Adriani et al. "Measurement of Forward Neutral Pion Transverse Momentum Spectra for $\sqrt{s} = 7\,\text{TeV}$ Proton-Proton Collisions at LHC." In: *Phys. Rev. D* 86 (2012), p. 092001. DOI: 10.1103/PhysRevD.86.092001. arXiv: 1205.4578 [hep-ex].

[AGK73] V. Abramovsky, V. Gribov, and O. Kancheli. "Character of Inclusive Spectra and Fluctuations Produced in Inelastic Processes by Multi-Pomeron Exchange." In: *Yad. Fiz.* 18 (1973), p. 595.

[Agl+04] M. Aglietta et al. "The Cosmic Ray Primary Composition in the 'Knee' Region through the EAS Electromagnetic and Muon Measurements at EAS-TOP." In: *Astropart. Phys.* 21 (2004), p. 583. DOI: 10.1016/j.astropartphys.2004.04.005.

[AGL05] E. Avsar, G. Gustafson, and L. Lonnblad. "Energy conservation and saturation in small-x evolution." In: *JHEP* 07 (2005), p. 062. DOI: 10.1088/1126-6708/2005/07/062. arXiv: hep-ph/0503181 [hep-ph].

[Ahn+09] E.-J. Ahn et al. "Cosmic Ray Interaction Event Generator SIBYLL 2.1." In: *Phys. Rev. D* 80 (2009), p. 094003. DOI: 10.1103/PhysRevD.80.094003. arXiv: 0906.4113 [hep-ph].

[Alb+14] M. Albrow et al. *CMS-TOTEM Precision Proton Spectrometer*. Tech. rep. CERN-LHCC-2014-021. TOTEM-TDR-003. CMS-TDR-13. Geneva: CERN, Sept. 2014.

[Alb11] M. Albrow. "Double Pomeron Exchange: from the ISR to the LHC." In: *AIP Conf. Proc.* 1350 (2011), p. 119. DOI: 10.1063/1.3601389. arXiv: 1010.0625 [hep-ex].

[Ali+12] A. Alici et al. *Study of the LHC Ghost Charge and Satellite Bunches for Luminosity Calibration*. Tech. rep. CERN-ATS-Note-2012-029. http://cds.cern.ch/record/1427728/. Geneva: CERN, Feb. 2012.

[All+13] J. Allen et al. "Air Shower Simulation and Hadronic Interactions." In: *EPJ Web Conf.* 53 (2013), p. 01007. DOI: 10.1051/epjconf/20135301007. arXiv: 1306.6090 [astro-ph.HE].

[Alv+08] J. Alves A. Augusto et al. "The LHCb Detector at the LHC." In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.

[Alv+10] M. Alvioli et al. "Diffraction on Nuclei: Effects of Nucleon Correlations." In: *Phys. Rev. C* 81 (2010), p. 025204. DOI: 10.1103/PhysRevC.81.025204. arXiv: 0911.1382 [nucl-th].

[Alv+13] M. Alvioli et al. "Nucleon Momentum Distributions, their Spin-Isospin Dependence and Short-Range Correlations." In: *Phys. Rev. C* 87 (2013), p. 034603. DOI: 10.1103/PhysRevC.87.034603. arXiv: 1211.0134 [nucl-th].

[Alw+07] J. Alwall et al. "A Standard Format for Les Houches Event Files." In: *Comput. Phys. Commun.* 176 (2007), p. 300. DOI: 10.1016/j.cpc.2006.11.010. arXiv: hep-ph/0609017 [hep-ph].

[Ame+08] M. Amenomori et al. "The All-Particle Spectrum of Primary Cosmic Rays in the Wide Energy Range from 10^{14} eV to 10^{17} eV Observed with the Tibet-III Air-Shower Array." In: *Astrophys. J.* 678 (2008), p. 1165. DOI: 10.1086/529514. arXiv: 0801.1803 [hep-ex].

[And+10] V. Andreev et al. "Performance Studies of a Full-Length Prototype for the CASTOR Forward Calorimeter at the CMS Experiment." In: *Eur. Phys. J. C* 67 (2010), p. 601. DOI: 10.1140/epjc/s10052-010-1316-4.

[Ane+08] G. Anelli et al. "The TOTEM Experiment at the CERN Large Hadron Collider." In: *JINST* 3 (2008), S08007. DOI: 10.1088/1748-0221/3/08/S08007.

[Ang+99] A. Angelis et al. "Formation of Centauro and Strangelets in Nucleus-Nucleus Collisions at the LHC and Their Identification by the ALICE Experiment." In: Proceedings of the 26th International Cosmic Ray Conference (ICRC 1999) (1999), p. 424. arXiv: hep-ph/9908210 [hep-ph].

[Ant+13a] G. Antchev et al. "Luminosity-Independent Measurement of the Proton-Proton Total Cross Section at $\sqrt{s} = 8\,\text{TeV}$." In: *Phys. Rev. Lett.* 111.1 (2013), p. 012001. DOI: 10.1103/PhysRevLett.111.012001.

[Ant+13b] G. Antchev et al. "Luminosity-independent measurements of total, elastic and inelastic cross-sections at $\sqrt{s} = 7\,\text{TeV}$." In: *Europhys. Lett.* 101 (2013), p. 21004. DOI: 10.1209/0295-5075/101/21004.

[Ant+13c] G. Antchev et al. "Measurement of Proton-Proton Inelastic Scattering Cross-Section at $\sqrt{s} = 7\,\text{TeV}$." In: *Europhys. Lett.* 101 (2013), p. 21003. DOI: 10.1209/0295-5075/101/21003.

[Ant15] J. F. R. Antunes. "Study of Novel Sensor Technologies for the Monitoring of Particle Physics Detectors." In prepartation. PhD thesis. Lisbon: Universidade Nova De Lisboa, 2015.

[Ape+11a] W. Apel et al. "Kneelike Structure in the Spectrum of the Heavy Component of Cosmic Rays Observed with KASCADE-Grande." In: *Phys. Rev. Lett.* 107 (2011), p. 171104. DOI: 10.1103/PhysRevLett.107.171104. arXiv: 1107.5885 [astro-ph.HE].

[Ape+11b] W. Apel et al. "Muon Production Height Studies with the Air Shower Experiment KASCADE-Grande." In: *Astropart. Phys.* 34 (2011), p. 476. DOI: 10.1016/j.astropartphys.2010.10.016.

[Ape+13] W. Apel et al. "KASCADE-Grande Measurements of Energy Spectra for Elemental Groups of Cosmic Rays." In: *Astropart. Phys.* 47 (2013), p. 54. DOI: 10.1016/j.astropartphys.2013.06.004. arXiv: 1306.6283 [astro-ph.HE].

[Ava+86] V. V. Avakian et al. "Determining inelastic interaction cross-sections for nucleons and pions incident on carbon and lead nuclei at 0.5 TeV - 5 TeV." In: *Bull. Acad. Sci. USSR, Phys. Ser.* 50.11 (1986), p. 4.

[Bab+74] A. Babaev et al. "The neutron total cross section measurements on protons and nuclei in the energy range of 28-54 GeV." In: *Phys. Lett. B* 51 (1974), p. 501. DOI: 10.1016/0370-2693(74)90321-9.

[Bar+13] E. Barcikowski et al. "Mass Composition Working Group Report at UHECR-2012." In: *EPJ Web Conf.* 53 (2013), p. 01006. DOI: 10.1051/epjconf/20135301006. arXiv: 1306.4430 [astro-ph.HE].

[Bar14] C. Barschel. "Precision Luminosity Measurement at LHCb with Beam-Gas Imaging." "https://cds.cern.ch/record/1693671/files/CERN-THESIS-2013-301.pdf". PhD thesis. Aachen, Germany: RWTH Aachen University, III. Physikalisches Institut A, 2014.

[Bau11] C. Baus. "Study of Hadronic Interactions Using Longitudinal Shower Profiles of the Pierre Auger Observatory." Master thesis. Karlsruhe: Karlsruhe Institute of Technology, 2011.

[Bay+08] G. Bayatian et al. "Design, Performance and Calibration of the CMS Forward Calorimeter Wedges." In: *Eur. Phys. J. C* 53 (2008), p. 139. DOI: 10.1140/epjc/s10052-007-0459-4.

[BCG94] R. Brun, F. Carminati, and S. Giani. *GEANT Detector Description and Simulation Tool*. Tech. rep. CERN-W-5013. Geneva: CERN, 1994.

[BE80] M. Bassetti and G. Erskine. *Closed Expression for the Electrical Field of a Two-dimensional Gaussian Charge*. Tech. rep. CERN-ISR-TH/80-06. Geneva: CERN, 1980.

[BEH09] J. Bluemer, R. Engel, and J. R. Horandel. "Cosmic Rays from the Knee to the Highest Energies." In: *Prog. Part. Nucl. Phys.* 63 (2009), p. 293. DOI: 10.1016/j.ppnp.2009.05.002. arXiv: 0904.0725 [astro-ph.HE].

[Bel+07] D. Belohrad et al. *Mechanical Design of the Intensity Measurement Devices for the LHC*. Tech. rep. CERN-AB-2007-026. Geneva: CERN, 2007, p. 253.

[Bel66] G. Bellettini. "Proton-nuclei cross sections at 20 GeV." In: *Nucl. Phys.* 79 (1966), p. 609. DOI: 10.1016/0029-5582(66)90267-7.

[Ber+07] T. Bergmann et al. "One-Dimensional Hybrid Approach to Extensive Air Shower Simulation." In: *Astropart. Phys.* 26 (2007), p. 420. DOI: 10.1016/j.astropartphys.2006.08.005. arXiv: astro-ph/0606564 [astro-ph].

[BM09] P. M. Bauleo and J. R. Martino. "The Dawn of the Particle Astronomy Era in Ultra-High-Energy Cosmic Rays." In: *Nature* 458N7240 (2009), p. 847. DOI: 10.1038/nature07948.

[Boc+13] A. Boccardi et al. *LHC Luminosity Calibration using the Longitudinal Density Monitor*. Tech. rep. CERN-ATS-Note-2013-034. Geneva: CERN, 2013.

[Bop+08] F. W. Bopp et al. "Antiparticle to Particle Production Ratios in Hadron-Hadron and d-Au Collisions in the DPMJET-III Monte Carlo." In: *Phys. Rev. C* 77 (2008), p. 014904. DOI: 10.1103/PhysRevC.77.014904. eprint: hep-ph/0505035.

[BP02] V. Barone and E. Predazzi. *High-Energy Particle Diffraction*. Berlin, Germany: Springer, 2002.

[BR97] R. Brun and F. Rademakers. "ROOT: An Object Oriented Data Analysis Framework." In: *Nucl. Instrum. Meth. A* 389 (1997), p. 81. DOI: 10.1016/S0168-9002(97)00048-X.

[Brü+04] O. S. Brüning et al. *LHC Design Report*. Geneva: CERN, 2004.

[Buc+10] A. Buckley et al. "Systematic Event Generator Tuning for the LHC." In: *Eur. Phys. J. C* 65 (2010), p. 331. DOI: 10.1140/epjc/s10052-009-1196-7. arXiv: 0907.2973 [hep-ph].

[Buc+13] A. Buckley et al. "Rivet User Manual." In: *Comput. Phys. Commun.* 184 (2013), p. 2803. DOI: 10.1016/j.cpc.2013.05.021. arXiv: 1003.0694 [hep-ph].

[Bud+13] N. Budnev et al. "Tunka-25 Air Shower Cherenkov Array: The Main Results." In: *Astropart. Phys.* 50 (2013), p. 18. DOI: 10.1016/j.astropartphys.2013.09.006.

[BUE15] C. Baus, R. Ulrich, and R. Engel. "Relation between Hadronic Interactions and Ultra-High Energy Extensive Air Showers." In: *EPJ Web of Conferences* 99 (2015), p. 11001. DOI: 10.1051/epjconf/20159911001.

[But09] J. Butterworth. "Starlight: A C++ Object-Oriented Monte Carlo Event Generator for Ultra-Peripheral Collisions." "https://dspace.creighton.edu/xmlui/handle/10504/4543". PhD thesis. Creighton University, 2009.

[Car+79] A. S. Carroll et al. "Absorption Cross-Sections of $\pi^{\pm}$, $K^{\pm}$, p and Anti-p on Nuclei between 60 GeV/c and 280 GeV/c." In: *Phys. Lett. B* 80 (1979), p. 319. DOI: 10.1016/0370-2693(79)90226-0.

[Car13] N. Cartiglia. "Measurement of the Proton-Proton Total Cross Section at 2, 7, 8 and 57 TeV." In: *e-print* (2013), p. 55. arXiv: 1303.2927 [hep-ex].

[CG99] S. R. Coleman and S. L. Glashow. "High-Energy Tests of Lorentz Invariance." In: *Phys. Rev. D* 59 (1999), p. 116008. DOI: 10.1103/PhysRevD.59.116008. arXiv: hep-ph/9812418 [hep-ph].

[Cha+08] S. Chatrchyan et al. "The CMS Experiment at the CERN LHC." In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[Cha+11] S. Chatrchyan et al. "Measurement of Energy Flow at Large Pseudorapidities in pp Collisions at $\sqrt{s} = 0.9$ and 7 TeV." In: *JHEP* 1111 (2011), p. 148. DOI: 10.1007/JHEP11(2011)148, 10.1007/JHEP02(2012)055. arXiv: 1110.0211 [hep-ex].

[Cha+13a] S. Chatrchyan et al. "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 7$ TeV." In: *Phys. Lett. B* 722 (2013), p. 5. DOI: 10.1016/j.physletb.2013.03.024. arXiv: 1210.6718 [hep-ex].

[Cha+13b] S. Chatrchyan et al. "Study of the Underlying Event at Forward Rapidity in pp Collisions at $\sqrt{s} = 0.9$, 2.76, and 7 TeV." In: *JHEP* 1304 (2013), p. 072. DOI: 10.1007/JHEP04(2013)072. arXiv: 1302.2394 [hep-ex].

[Cha+14a] S. Chatrchyan et al. "Description and performance of track and primary-vertex reconstruction with the CMS tracker." In: *JINST* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].

[Cha+14b] S. Chatrchyan et al. "Measurement of Pseudorapidity Distributions of Charged Particles in Proton-Proton Collisions at $\sqrt{s}$ = 8 TeV by the CMS and TOTEM Experiments." In: *Eur. Phys. J. C* 74.10 (2014), p. 3053. DOI: 10.1140/epjc/s10052-014-3053-6. arXiv: 1405.0722 [hep-ex].

[Cio+11] C. Ciofi degli Atti et al. "Number of Collisions in the Glauber Model and Beyond." In: *Phys. Rev. C* 84 (2011), p. 025205. DOI: 10.1103/PhysRevC.84.025205. arXiv: 1105.1080 [nucl-th].

[CMS12] CMS Collaboration. *Measurement of the Pseudorapidity and Centrality Dependence of the Very Forward Energy Density in PbPb Collisions at* $\sqrt{s} = 2.76$ *ATeV*. CMS Physics Analysis Summary. Geneva: CERN, 2012.

[CMS13a] CMS Collaboration. *CASTOR Very Forward Calorimeter Performance in 2013 Data in pPb and pp Collisions*. Tech. rep. CMS-DP-2013-035. "https://cds.cern.ch/record/1632684". Geneva: CERN, Nov. 2013.

[CMS13b] CMS Collaboration. *Measurement of the Inelastic pPb Cross Section at* $\sqrt{s_{NN}} = 5.02$ *TeV*. CMS Physics Analysis Summary CMS-PAS-FSQ-13-006. Geneva: CERN, 2013.

[CMS14a] CMS and TOTEM Collaborations. "Alignment of CASTOR with 2013 pA TOTEM-Data." In: CMS-DP-2014-014 (Mar. 2014). "https://cds.cern.ch/record/1697694".

[CMS14b] CMS Collaboration. "Jet Measurement with the CASTOR Calorimeter." In: CMS-DP-2014-022 (June 2014). "https://cds.cern.ch/record/1968147".

[CMS14c] CMS Collaboration. *Leading Track and Leading Jet Cross Sections at Small Transverse Momenta in pp Collisions at* $\sqrt{s}$ *=8 TeV*. CMS Physics Analysis Summary CMS-PAS-FSQ-12-032. Geneva: CERN, 2014.

[CMS14d] CMS Collaboration. *Luminosity calibration for the 2013 proton-lead and proton-proton data taking*. CMS Physics Analysis Summary CMS-PAS-LUM-13-002. 2014.

[Cud+02] J. Cudell et al. "Benchmarks for the Forward Observables at RHIC, the Tevatron Run II and the LHC." In: *Phys. Rev. Lett.* 89 (2002), p. 201801. DOI: 10.1103/PhysRevLett.89.201801. arXiv: hep-ph/0206172 [hep-ph].

[CVZ05] L. Cazon, R. Vazquez, and E. Zas. "Depth Development of Extensive Air Showers from Muon Time Distributions." In: *Astropart. Phys.* 23 (2005), p. 393. DOI: 10.1016/j.astropartphys.2005.01.009. arXiv: astro-ph/0412338 [astro-ph].

[Dec94] F. J. Decker. "Beam distributions beyond RMS." In: *SLAC-PUB-6684, C94-10-02.1* (1994).

[dEn+11] D. d'Enterria et al. "Constraints From the First Lhc Data on Hadronic Event Generators for Ultra-High Energy Cosmic-Ray Physics." In: *Astropart. Phys.* 35 (2011), p. 98. DOI: 10.1016/j.astropartphys.2011.05.002. arXiv: 1101.5596 [astro-ph.HE].

[Den+73] S. P. Denisov et al. "Absorption cross-sections for pions, kaons, protons and anti-protons on complex nuclei in the 6 to 60 GeV/c momentum range." In: *Nucl. Phys. B* 61 (1973), p. 62. DOI: 10.1016/0550-3213(73)90351-9.

[dEn03] D. G. d'Enterria. "Hard Scattering Cross-Sections at LHC in the Glauber Approach: From pp to pA and AA Collisions." In: *e-print* (2003). arXiv: nucl-ex/0302016 [nucl-ex].

[Der+00] U. Dersch et al. "Total Cross Section Measurements with π^-, Σ^- and Protons on Nuclei and Nucleons around 600 GeV/c." In: *Nucl. Phys. B* 579 (2000), p. 277. DOI: 10.1016/S0550-3213(00)00204-2. arXiv: hep-ex/9910052.

[DH01] M. Dobbs and J. B. Hansen. "The HepMC C++ Monte Carlo Event Record for High Energy Physics." In: *Comput. Phys. Commun.* 134 (2001), p. 41. DOI: 10.1016/S0010-4655(00)00189-2.

[DN12] I. Dremin and V. Nechitailo. "Elastic pp-Scattering at $\sqrt{s} = 7$ TeV with the Genuine Orear Regime and the Dip." In: *Phys. Rev. D* 85 (2012), p. 074009. DOI: 10.1103/PhysRevD.85.074009. arXiv: 1202.2016 [hep-ph].

[Don+02] S. Donnachie et al. *Pomeron Physics and QCD*. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 2002. ISBN: 9781139441391.

[Dre+01] H. Drescher et al. "Parton Based Gribov-Regge Theory." In: *Phys. Rept.* 350 (2001), p. 93. DOI: 10.1016/S0370-1573(00)00122-8. arXiv: hep-ph/0007198 [hep-ph].

[EB08] L. Evans and P. Bryant. "LHC Machine." In: *JINST* 3 (2008). Ed. by L. Evans, S08001. DOI: 10.1088/1748-0221/3/08/S08001.

[Eli15] M. Eliachevitsch. "Revised Simulation of the Response of the CASTOR Very Forward Calorimeter at the CMS Experiment." Bachelor thesis. In preparation. Karlsruhe: Karlsruhe Institute of Technology, 2015.

[Eng+70] J. Engler et al. "Neutron-Nucleus Total Cross-Sections Between 8 GeV/c and 21 GeV/c." In: *Phys. Lett. B* 32 (1970), p. 716. DOI: 10.1016/0370-2693(70)90453-3.

[Eng+98] R. Engel et al. "Proton Proton Cross-Section at $\sqrt{s}$ similar to 30 TeV." In: *Phys. Rev. D* 58 (1998), p. 014019. DOI: 10.1103/PhysRevD.58.014019. arXiv: hep-ph/9802384 [hep-ph].

[ER96] R. Engel and J. Ranft. "Hadronic Photon-Photon Interactions at High-Energies." In: *Phys. Rev. D* 54 (1996), p. 4244. DOI: 10.1103/PhysRevD.54.4244. arXiv: hep-ph/9509373 [hep-ph].

[EU12] R. Engel and R. Ulrich. "Cross-Section Calculation using Glauber Theory with Inelastic Screening." Internal note but the result is used in [Abr+12]. http://www.auger.org:8080/index.php/gap-notes/viewdownload/5-gap-notes-2012/310-gap2012-056. 2012.

[FA13] G. R. Farrar and J. D. Allen. "A New Physical Phenomenon in Ultra-High Energy Collisions." In: *EPJ Web Conf.* 53 (2013), p. 07007. DOI: 10.1051/epjconf/20135307007. arXiv: 1307.2322 [hep-ph].

[FGL11] C. Flensburg, G. Gustafson, and L. Lonnblad. "Inclusive and exclusive observables from dipoles in high energy collisions." In: *JHEP* 08 (2011), p. 103. DOI: 10.1007/JHEP08(2011)103. arXiv: 1103.4321 [hep-ph].

[Fra+93] L. Frankfurt et al. "Evidence for short range correlations from high Q^2 (e, e′) reactions." In: *Phys. Rev. C* 48 (1993), p. 2451. DOI: 10.1103/PhysRevC.48.2451.

[Fro61] M. Froissart. "Asymptotic Behavior and Subtractions in the Mandelstam Representation." In: *Phys. Rev.* 123 (1961), p. 1053. DOI: 10.1103/PhysRev.123.1053.

[Gen03] J. Gentle. *Random Number Generation and Monte Carlo Methods*. Statistics and Computing. https://books.google.de/books?id=ERSSDBDcYOIC. Springer, 2003. ISBN: 9780387001784.

[GK14] P. Grafström and W. Kozanecki. "Luminosity Determination at Proton Colliders." In: *Prog. Part. Nucl. Phys.* (2014). In press. DOI: 10.1016/j.ppnp.2014.11.002.

[GM70] R. Glauber and G. Matthiae. "High-Energy Scattering of Protons by Nuclei." In: *Nucl. Phys. B* 21 (1970), p. 135.

[Got10] P. Gottlicher. "Design and Test Beam Studies for the CASTOR Calorimeter of the CMS Experiment." In: *Nucl. Instrum. Meth. A* 623 (2010), p. 225. DOI: 10.1016/j.nima.2010.02.203.

[Gou83] K. A. Goulianos. "Diffractive Interactions of Hadrons at High-Energies." In: *Phys. Rept.* 101 (1983), p. 169. DOI: 10.1016/0370-1573(83)90010-8.

[GP93] G. Grindhammer and S. Peters. "The Parameterized Simulation of Electromagnetic Showers in Homogeneous and Sampling Calorimeters." In: (1993). arXiv: hep-ex/0001020 [hep-ex].

[Gre66] K. Greisen. "End to the Cosmic-Ray Spectrum?" In: *Phys. Rev. Lett.* 16.17 (1966), p. 748. DOI: 10.1103/PhysRevLett.16.748.

[Gri68] V. Gribov. "A Reggeon Diagram Technique." In: *Sov. Phys. JETP* 26 (1968), p. 414.

[GW60] M. Good and W. Walker. "Diffraction Disssociation of Beam Particles." In: *Phys. Rev.* 120 (1960), p. 1857. DOI: 10.1103/PhysRev.120.1857.

[Hec+98] D. Heck et al. "CORSIKA: A Monte Carlo Code to Simulate Extensive Air Showers." In: *FZKA-6019* (1998).

[Heg+14] B. Hegner et al. "Status and New Developments of the Generator Services." In: *J. Phys. Conf. Ser.* 513 (2014), p. 022029. DOI: 10.1088/1742-6596/513/2/022029.

[Hei50] W. Heitler. *The Quantum Theory of Radiation*. 2nd. Oxford: Oxford Univ. Press, 1950.

[HFB07] A. J. Hamilton, R. A. Fesen, and W. P. Blair. "A High Resolution UV Absorption Spectrum of Supernova ejecta in SN1006." In: *Mon. Not. Roy. Astron. Soc.* 381 (2007), p. 771. DOI: 10.1111/j.1365-2966.2007.12264.x. arXiv: astro-ph/0602553 [astro-ph].

[Hil84] A. Hillas. "The Origin of Ultra-High-Energy Cosmic Rays." In: *Ann. Rev. Astron. Astrophys.* 22.1 (1984), p. 425. DOI: 10.1146/annurev.aa.22.090184.002233.

[ISS01] V. Innocente, L. Silvestris, and D. Stickland. "CMS Software Architecture - Software Framework, Services and Persistency in High Level Trigger, Reconstruction and Analysis." In: *Comput. Phys. Commun.* 140 (2001), p. 31. DOI: 10.1016/S0010-4655(01)00253-3.

[Kaw14] K. Kawade. "Measurement of Neutron Production in the Very Forward Rapidity at LHC $\sqrt{s}$ =7 TeV pp Collision." "http://ir.nul.nagoya-u.ac.jp/jspui/bitstream/2237/20328/1/k10659_%E4%B8%BB%E8%AB%96%E6%96%87.pdf". PhD thesis. Nagoya: Nagoya University, 2014.

[Kha+15] V. Khachatryan et al. "Measurement of Diffraction Dissociation Cross Sections in pp Collisions at $\sqrt{s}$ = 7 TeV." In: *Phys. Rev. D* 92.1 (2015), p. 012003. DOI: 10.1103/PhysRevD.92.012003. arXiv: 1503.08689 [hep-ex].

[Kle05] K. Kleinknecht. *Detektoren für Teilchenstrahlung*. 4th. Wiesbaden: Teubner, 2005. ISBN: 3-8351-0058-0.

[KN99] S. Klein and J. Nystrand. "Exclusive vector meson production in relativistic heavy ion collisions." In: *Phys. Rev. C* 60 (1999), p. 014903. DOI: 10.1103/PhysRevC.60.014903. arXiv: hep-ph/9902259 [hep-ph].

[Kna97] J. Knapp. "Comparison of Hadronic Interaction Models Used in Simulations of Extensive Cosmic Ray Air Showers." In: *FZKA-5970* (1997). "bibliothek.fzk.de/zb/berichte/FZKA5970.pdf".

[KO11] K. Kotera and A. V. Olinto. "The Astrophysics of Ultrahigh Energy Cosmic Rays." In: *Ann. Rev. Astron. Astrophys.* 49 (2011), p. 119. DOI: 10.1146/annurev-astro-081710-102620. arXiv: 1101.4256 [astro-ph.HE].

[KP06] P. Katsas and A. Panagiotou. *Simulation of Energy Response Linearity, Resolution and e^-/π Ratio for the CASTOR Calorimeter at CMS*. Tech. rep. CERN-CMS-NOTE-2006-147. Geneva: CERN, 2006.

[Kuz12] E. Kuznetsova. "Performance and Calibration of CASTOR Calorimeter at CMS." In: *Phys. Procedia* 37 (2012), p. 356. DOI: 10.1016/j.phpro.2012.02.386.

[LHC12] LHCb Collaboration. *First analysis of the pPb pilot run data with LHCb*. LHCb Note LHCb-CONF-2012-034. Dec. 2012.

[LHC15] LHC Working Group on Forward Physics and Diffraction. *Yellow Report*. In prepartation. Geneva: CERN, 2015.

[Lib13] S. Liberati. "Tests of Lorentz Invariance: A 2013 Update." In: *Class. Quant. Grav.* 30 (2013), p. 133001. DOI: 10.1088/0264-9381/30/13/133001. arXiv: 1304.5795 [gr-qc].

[Lip12] P. Lipari. "Hadronic Interactions and Shower Development for UHE Cosmic Rays." Talk given at Workshop on Particle Physics at Auger. Lisbon, Jan. 2012.

[Lun+04] R. Luna et al. "Influence of Diffractive Interactions on Cosmic Ray Air Showers." In: *Phys. Rev. D* 70 (2004), p. 114034. DOI: 10.1103/PhysRevD.70.114034. arXiv: hep-ph/0408303 [hep-ph].

[Man13] D. Manglunki. "Prospects for Light-Ion LHC Running." Talk given at Workshop on Results and prospects of forward physics at the LHC. https://indico.cern.ch/event/223562/session/1/contribution/5/material/slides/1.pdf. Geneva, Sitzerland, Feb. 2013.

[Mar13] E. Martynov. "Proton (Antiproton) Elastic Scattering at Energies from FNAL to the LHC in the Tripole Pomeron-Odderon Model." In: *Phys. Rev. D* 87.11 (2013), p. 114018. DOI: 10.1103/PhysRevD.87.114018. arXiv: 1305.3093 [hep-ph].

[Mat05] J. Matthews. "A Heitler Model of Extensive Air Showers." In: *Astropart. Phys.* 22.5-6 (2005), p. 387. ISSN: 0927-6505. DOI: DOI:10.1016/j.astropartphys.2004.09.003.

[Mee68] S. van der Meer. *Calibration of the effective beam height in the ISR*. Tech. rep. CERN-ISR-PO-68-31. ISR-PO-68-31. Geneva: CERN, 1968.

[Met02] W. J. Metzger. "Statistical Methods in Data Analysis." Nijmegen, 2002.

[MH11] J. Mans and J. Hanson. *Calibration of the HF Absolute Energy Scale with Z di-electron Events Using 1/fb of Data in 2011*. Tech. rep. CMS-DN-11/012. http://cms.cern.ch/iCMS/jsp/openfile.jsp?type=DN&year=2011&files=DN2011_012.pdf. Geneva: CERN, 2011.

[Mic69] R. Mickens. "High-Energy Relations for Scattering of Particles and Antiparticles." In: *Lett. Nuovo Cim.* 2S1 (1969), p. 231. DOI: 10.1007/BF02754365.

[Mil+07] M. L. Miller et al. "Glauber modeling in high energy nuclear collisions." In: *Ann. Rev. Nucl. Part. Sci.* 57 (2007), p. 205. DOI: 10.1146/annurev.nucl.57.090506.123020. arXiv: nucl-ex/0701025 [nucl-ex].

[MS13a] M. Menon and P. Silva. "A Study on Analytic Parametrizations for the Proton-Proton Cross-Sections and Asymptotia." In: *e-print* (2013). arXiv: 1305.2947 [hep-ph].

[MS13b] M. Menon and P. Silva. "An Updated Analysis on the Rise of the Hadronic Total Cross-Section at the LHC Energy Region." In: *Int. J. Mod. Phys. A* 28 (2013), p. 1350099. DOI: 10.1142/S0217751X13500991. arXiv: 1212.5096 [hep-ph].

[Mur+75] R. Murthy et al. "Neutron Total Cross-Sections on Nuclei at Fermilab Energies." In: *Nucl. Phys. B* 92 (1975), p. 269.

[Nak+10] K. Nakamura et al. "Review of Particle Physics." In: *J. Phys. G* 37 (2010), p. 075021. DOI: 10.1088/0954-3899/37/7A/075021.

[OLT09] P. Odier, M. Ludwig, and S. Thoulet. *The DCCT for the LHC Beam Intensity Measurement*. Tech. rep. CERN-BE-2009-019. Geneva: CERN, May 2009.

[Ost10] S. Ostapchenko. "Total and Diffractive Cross Sections in Enhanced Pomeron Scheme." In: *Phys. Rev. D* 81 (2010), p. 114028. DOI: 10.1103/PhysRevD.81.114028. arXiv: 1003.0196 [hep-ph].

[Ost11] S. Ostapchenko. "Monte Carlo Treatment of Hadronic Interactions in Enhanced Pomeron Scheme: I. QGSJET-II model." In: *Phys. Rev. D* 83 (2011), p. 014018. DOI: 10.1103/PhysRevD.83.014018. arXiv: 1010.1869 [hep-ph].

[Par+11] R. Parsons et al. "Systematic uncertainties in air shower measurements from high-energy hadronic interaction models." In: *Astropart. Phys.* 34 (2011), p. 832. DOI: 10.1016/j.astropartphys.2011.02.007. arXiv: 1102.4603 [astro-ph.HE].

[Pie+11] Pierre Auger Collaboration et al. "Search for First Harmonic Modulation in the Right Ascension Distribution of Cosmic Rays Detected at the Pierre Auger Observatory." In: *Astropart. Phys.* 34 (Mar. 2011), p. 627. DOI: 10.1016/j.astropartphys.2010.12.007. arXiv: 1103.2721 [astro-ph.HE].

[Pie+13] T. Pierog et al. "EPOS LHC: Test of Collective Hadronization with LHC Data." In: *e-print* (2013). arXiv: 1306.0121 [hep-ph].

[Pie13] T. Pierog. "LHC Data and Extensive Air Showers." In: *EPJ Web Conf.* 52 (2013), p. 03001. DOI: 10.1051/epjconf/20125203001.

[Por14] A. Porcelli. "Measurement of the Depth of Shower Maximum in the Transition Region between Galactic and Extragalactic Cosmic Rays with the Pierre Auger Observatory." "http://www.auger.org:8080/index.php/internal/document-centre/viewdownload/26-gap-notes-2014/1954-gap2014-058". PhD thesis. Karlsruhe, Germany: Karlsruhe Institute of Technology, 2014.

[Pro+14] V. Prosin et al. "Tunka-133: Results of 3 year operation." In: *Nucl. Instrum. Meth. A* 756 (2014), p. 94. DOI: 10.1016/j.nima.2013.09.018.

[RK12] R. Rahmat and R. Kroeger. "HF GFlash." In: *Phys. Procedia* 37 (2012), p. 340. DOI: 10.1016/j.phpro.2012.02.385.

[SB87] T. Sjostrand and M. Bengtsson. "The Lund Monte Carlo for Jet Fragmentation and e^+ e^- Physics. Jetset Version 6.3: An Update." In: *Comput. Phys. Commun.* 43 (1987), p. 367. DOI: 10.1016/0010-4655(87)90054-3.

[Shn+07] R. Shneor et al. "Investigation of Proton-Proton Short-Range Correlations via the C-12(e, e′ pp) reaction." In: *Phys. Rev. Lett.* 99 (2007), p. 072501. DOI: 10.1103/PhysRevLett.99.072501. arXiv: nucl-ex/0703023 [nucl-ex].

[Sjo88] T. Sjostrand. "Status of Fragmentation Models." In: *Int. J. Mod. Phys. A* 3 (1988), p. 751. DOI: 10.1142/S0217751X88000345.

[Smi+13] M. Smith et al. "The Astrophysical Multimessenger Observatory Network (AMON)." In: *Astropart. Phys.* 45 (2013), p. 56. DOI: 10.1016/j.astropartphys.2013.03.003. arXiv: 1211.5602 [astro-ph.HE].

[SMS08] T. Sjostrand, S. Mrenna, and P. Z. Skands. "A Brief Introduction To PYTHIA 8.1." In: *Comput. Phys. Commun.* 178 (2008), p. 852. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].

[SV14] G. Sigl and A. van Vliet. "PeV Neutrinos from Ultra-High-Energy Cosmic Rays." In: *e-print* (2014), p. 177. arXiv: 1407.6577 [astro-ph.HE].

[Tok+12] H. Tokuno et al. "New Air Fluorescence Detectors Employed in the Telescope Array Experiment." In: *Nucl. Instrum. Meth. A* 676 (2012), p. 54. DOI: 10.1016/j.nima.2012.02.044. arXiv: 1201.0002 [astro-ph.IM].

[UEU11] R. Ulrich, R. Engel, and M. Unger. "Hadronic multiparticle production at ultra-high energies and extensive air showers." In: *Phys. Rev. D* 83 (2011), p. 054026. DOI: 10.1103/PhysRevD.83.054026. arXiv: 1010.4310 [hep-ph].

[WBP10] S. M. White, H. Burkhardt, and P. Puzo. "Determination of the Absolute Luminosity at the LHC." "http://cds.cern.ch/record/1308187/". PhD thesis. Orsay, France: Orsay, Université Paris-Sud 11, 2010.

[Wei34] C. von Weizsäcker. "Radiation Emitted in Collisions of Very Fast Electrons." In: *Z. Phys. A* 88 (1934), p. 612. DOI: 10.1007/BF01333110.

[Wer93] K. Werner. "Strings, Pomerons, and the Venus Model of Hadronic Interactions at Ultrarelativistic Energies." In: *Phys. Rept.* 232 (1993), p. 87. DOI: 10.1016/0370-1573(93)90078-R.

[WG91] X.-N. Wang and M. Gyulassy. "HIJING: A Monte Carlo Model for Multiple Jet Production in pp, pA and AA Collisions." In: *Phys. Rev. D* 44 (1991), p. 3501. DOI: 10.1103/PhysRevD.44.3501.

[Wil74] K. G. Wilson. "Confinement of Quarks." In: *Phys. Rev. D* 10 (1974), p. 2445. DOI: 10.1103/PhysRevD.10.2445.

[WLP06] K. Werner, F.-M. Liu, and T. Pierog. "Parton Ladder Splitting and the Rapidity Dependence of Transverse Momentum Spectra in Deuteron-Gold Collisions at RHIC." In: *Phys. Rev. C* 74 (2006), p. 044902. DOI: 10.1103/PhysRevC.74.044902. arXiv: hep-ph/0506232 [hep-ph].

[Zen12] T. Zenis. "Measurement of the inelastic proton-proton cross section with the ATLAS detector." In: *AIP Conf. Proc.* 1523 (2012), p. 26. DOI: 10.1063/1.4802108.

[ZK66] G. Zatsepin and V. Kuzmin. "Upper limit of the spectrum of cosmic rays." In: *JETP Lett.* 4 (1966), p. 78.

ACKNOWLEDGMENTS

Foremost I want to thank Prof. J. Blümer for providing the possibility and the environment to write this thesis and to graduate. I also want to thank Prof. G. Quast for acting as co-referee for this work. At the same time I am grateful to the extensive support of the doctoral fellow program, KSETA, which I have been part of and which is organised by my referees along with many others.

I am much obliged to my direct supervisor, Dr. Ralf Ulrich. He leads a Helmholtz Young Investigator group and, within this frame, offered me endless advice and support, day and night. Due to his sophisticated leadership the group could make a name for itself in the large CMS Collaboration. In this context, I also want to thank my brave collaborators Dr. Igor Katkov and Hauke Wöhrman for joining in this amazing effort. I am also grateful to our (slightly) newer group members, Melike Akbiyik, Sebastian Baur, and Michael Eliachevitch with whom it was always pleasant to work with. Same goes to all members of our little coffee breaks who made the hard work bearable and even more so to Daniela Mockler, Anatoli Fedynitch, and Felix Riehn for their special support in hard times.

At CERN, many people helped with the CMS-based analysis and were providing invaluable work to keep the detectors running. I do not even dare to mention the full list, yet, some deserve extra credit. As chair of the review committee Prof. G. Veres vitally improved this analysis in many parts and was a very inspirational source. The same applies to another member of the committee, Dr. N. Cartiglia who is most likely unaware that his material on high-energy cross sections made me realise how important this subject is. Finally, I am grateful to the luminosity group for obtaining, after some initial struggles, a much lower uncertainty than originally estimated.

Thanks goes also to all those who helped me preparing this document. Foremost, Dr. Ralf Ulrich has improved the logic of the text in many places and helped me to make this a well-written scientific document. Also Dr. Darko Veberic, Ariel Bridgeman, and David Schmidt have read the document extensively and I learned a lot from their comments. Furthermore, Dr. Tanguy Pierog and Dr. Igor Katkov helped to improve parts of the thesis.

Last but not least I wish to thank my family for their never-ending support and my fianceé, Dr. Lu Lu, with whom I will be forever linked in life and science.

DECLARATION

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Karlsruhe, April 2015

Colin Baus