# Measurement of the very forward energy in proton-proton collisions at 13 TeV with the CMS experiment and implications for hadronic interaction models

Zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften
von der Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)

genehmigte Dissertation

von
Sebastian Baur
aus Memmingen

Tag der mündlichen Prüfung: 10.11.2017
Referent: Prof. Dr. Dr. h.c. J. Blümer
Korreferent: Prof. Dr. Th. Müller
Betreuer: Dr. R. Ulrich

Soli Deo Gloria.

# Abstract

The theoretical understanding of soft interactions in hadronic collisions remains one of the major open questions of particle physics. Such processes form the underlying event in proton-proton collisions and dominate the development of extensive air showers produced by high energy cosmic rays. The very forward calorimeter of the CMS experiment CASTOR measures the largest energy densities accessible at the LHC and is a unique tool to study soft multiparticle production relevant for the underlying event and air shower development. Within this work, an important contribution is made to ensure the best possible data quality and detector performance. To achieve this, a dedicated calibration of infrared distance sensors is performed and the currently most precise determination of the position of CASTOR is achieved. The very forward collision energy is measured in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ and fully corrected to the stable particle level. The results are given in terms of the differential production cross section $\mathrm{d}\sigma/\mathrm{d}E$ and as a function of the charged particle multiplicity at central rapidity, $\langle E \rangle\,(N_{\mathrm{ch}})$. The unique data are used to derive implications on hadronic interaction models. New parameter tunes of PYTHIA 8 are obtained by optimizing the model predictions to the data. The predictions of air shower maxima and muon numbers by SIBYLL 2.3, QGSJETII.04 and EPOS LHC are reviewed in comparison to the data.

# Zusammenfassung

Das theoretische Verständnis weicher Wechselwirkungen mit geringem Impulsübertrag in Hadron-Kollisionen ist eines der ungelösten Probleme der Teilchenphysik. Diese Prozesse bilden den dominierenden Untergrund in Proton-Proton Kollisisonen und bestimmen die Entwicklung von ausgedehnten Luftschauern, die von Teilchen der kosmischen Strahlung erzeugt werden. Das Vorwärtskalorimeter CASTOR des CMS Experiments misst die größten Energiedichten am LHC und ist dadurch ein einzigartiger Detektor um die für den Untergrund und für die Luftschauerentwicklung relevante Teilchenproduktion zu studieren. Im Rahmen dieser Arbeit wurde ein wichtiger Beitrag zur Sicherstellung der Datenqualität und Leistungsfähigkeit des Detektors während der Datennahme geleistet. Dazu wurde eine spezielle Kalibrierung der Infrarot-Abstandssensoren entwickelt und die gegenwärtig präziseste Positionsbestimmung von CASTOR erreicht. Die Kollisionsenergie in der Akzeptanz von CASTOR wurde in Proton-Proton Kollisionen bei $\sqrt{s} = 13\,\text{TeV}$ gemessen und für Detektoreffekte korrigiert. Die Ergebnisse werden als differentieller Wirkungsquerschnitt $\mathrm{d}\sigma/\mathrm{d}E$ und als Funktion der Anzahl geladener Teilchen im Zentralbereich der Kollision $\langle E \rangle\,(N_{\mathrm{ch}})$ präsentiert. Anhand der Daten wurden Konsequenzen für hadronische Wechselwirkungsmodelle abgeleitet. Neue Parameterwerte für PYTHIA 8 wurden bestimmt. Vorhersagen über Luftschauermaxima und Myonzahlen von SIBYLL 2.3, QGSJETII.04 und EPOS LHC wurden anhand der Daten überprüft.

# Contents

# 1 Introduction

The era of hadron colliders – with the LHC [1] as current flagship – is characterized by an enormous improvement of the fundamental understanding of particle physics. The discoveries of the top quark [2, 3] and the Higgs boson [4, 5] completed the picture of elementary particles drawn by the standard model. The theory of strong interaction, quantum chromodynamics (QCD), was experimentally established and is studied to great detail. Moreover, the limits of current understanding are explored and searches for phenomena beyond the standard model are undertaken (see for example [6] for a review). This includes new elementary particles, possible particles forming dark matter, or new states of matter such as a quark-gluon plasma (QGP) [7].

The experiments at the LHC [8–13] operate particle detectors at the edge of technological possibilities, recording collision data at unprecedented rate and with remarkable accuracy. In a domain where partons interact with large momentum transfer, the experimental methods are so well understood that levels of precision similar to those of current perturbative calculations are reached. Huge efforts are therefore undertaken to calculate hadronic scattering processes at ever higher orders, including more and more diagrams and possible loop formations [14]. This allows to determine collision processes to good accuracy and searches for deviations from the standard model predictions become more powerful. There is, on the other hand, still a domain of QCD, characterized by small momentum transfer between scattering partons, that is so far theoretically not well understood. These soft interactions can not be calculated from first principles and need do be modeled phenomenologically. The corresponding models are implemented in hadronic event generators and need to be thoroughly validated against data. At the LHC, soft processes represent a huge background and constitute an important uncertainty in precision measurement and searches for new phenomena.

Even before hadron collider at GeV energies became feasible, hadronic interactions were observed in the Earth's atmosphere. Cosmic ray particles with high energies, impinging on the atmosphere, create particle cascades, so-called extensive air showers. It was in these air showers that the positron (as the first evidence of antimatter), the muon, and the pion were discovered [15–17]. Despite the fact that cosmic rays are studied since more than 100 years, some of the main questions concerning for example their origin and their mass composition are still unsolved [18]. Large observatories such as the

Pierre Auger Observatory [19, 20] have been constructed in order to study cosmic ray induced air showers at the highest energies. The development of these air showers in the atmosphere is dominated by hadronic particle production in directions close to the one of the incoming projectile, the so-called forward phase space. This is related to the fact that the majority of the available energy is directed there. These production mechanisms are not well constrained theoretically but can be studied with very forward detectors at the LHC [21]. The interpretation of air shower measurements depends in a crucial way on phenomenological descriptions provided by hadronic interaction models. Some of the main observations of air showers, such as the number of muons and their production depth, are still poorly described by the models [22, 23]. This leads to large ambiguities in the determination of the primary mass composition of cosmic rays [24–26].

This thesis is dedicated to perform unique measurements of multiparticle production in dominantly soft processes in proton-proton collisions at the LHC. These are studied in the very forward direction, close to the direction of the incoming proton beams. The CASTOR calorimeter of the CMS experiment provides an ideal tool to study the energies carried by the particles in this phase space. With an acceptance of $-6.6 < \eta < -5.2$ it is the most forward calorimeter at the LHC, that is sensitive to both charged and neutral particles. It therefore measures the highest energy densities at the LHC, which makes the measurements very powerful to constrain hadronic interaction models used to describe extensive air showers.

A major effort was made to record a dedicated low luminosity dataset of proton-proton collisions at a center-of-mass energy of 13 TeV. This was the very first data of LHC Run 2 (recorded in June 2015) at the current record energy of the LHC. Within the scope of this work, the CASTOR calorimeter was installed and commissioned in the CMS experiment. It was ensured that the quality of the recorded data is excellent and that the detector properties are well understood in order to perform physics analysis. In preparation of the data taking, the alignment system of CASTOR was revised and improved. A novel two-dimensional calibration method of infrared distance sensors was developed and applied to obtain the position of CASTOR within the CMS reference frame at best possible precision. The measured position was afterwards implemented in Monte Carlo simulations and validated with the data.

The collision energy measured with CASTOR is evaluated inclusively for all collision events seen by the detector. In a second step, the energy is correlated to the activity at central rapidities, determined by the number of charged particles. The potential of CASTOR is for the first time completely exploited by also separating the contributions of electromagnetic and hadronic particles to the total energy. The results are fully corrected for detector effects and presented on the level of stable particles. The results can therefore be directly compared to various model predictions.

In a final step, the obtained results are evaluated towards their implications for hadronic event generators. New parameter tunes of the PYTHIA 8 [27] model are obtained and constrained with the measurements performed with CASTOR. A better estimation of the soft interaction background to collisions with large momentum transfer is achieved. The new tune CP1 of PYTHIA 8 is about to become published by CMS and will be one of the main model references for future analyses of 13 TeV proton-proton data in the CMS collaboration. Event generators used to describe cosmic ray induced air showers are also constrained with the obtained results. The data are combined with a simulation-based method presented in Ref. [28]. The discrepancies between the data and model predictions are quantified and extrapolated to cosmic ray energies. The corresponding impact on the shower maximum depth and muon numbers are calculated and their implications discussed. Furthermore, a novel way to observe QGP-like collective effects in proton-proton collisions is found.

# 2 Interactions of hadronic particles

The standard model (SM) of particle physics, established in the 1970s, has proven to give accurate predictions of the properties of elementary particles within an elegant mathematical framework. It claims that all matter consists of elementary particles with spin $1/2$, quarks and leptons, organized in three generations. There are three leptons (electron, muon, tau) together with the corresponding neutrinos and six quarks (up, down, strange, charm, top, bottom). The SM not only includes the elementary particles but also a description of the forces between those particles. These forces are mediated through gauge bosons. The electromagnetic force is mediated via the photon, the weak force via the $W^{\pm}$ and $Z$ bosons and the strong force via the gluon. The latest success of the SM was marked in 2012 when the previously predicted Higgs boson was discovered by the ATLAS [4] and CMS [5] experiments at the CERN LHC. An illustration of the particles contained in the SM is shown in Fig. 2.1.

## 2.1 Fundamental properties of quantum chromodynamics

The strong force is formulated in the relativistic quantum field theory of *quantum chromodynamics* (QCD). Name-giving feature is the introduction of the color charge (red, green, blue, and the corresponding anti-colors) as an additional quantum number. Only particles with a color charge couple to gluons and thus participate in the strong interaction. The combination of the three different colors (anti-colors) or a pair of the same color and anti-color correspond to a neutral (white) color. Another main feature of QCD is the fact that the strong coupling constant $\alpha_{\mathrm{S}}$ decreases with the scale $Q^2$ of the interaction. The latest measurement of the running of $\alpha_{\mathrm{S}}$ by CMS [30] is shown in Fig. 2.2. This implies that at high energy scales (or small distances) the interaction becomes very weak and the quarks can be treated as free particles (*asymptotic freedom*). At low energies (or large distances) though, the coupling becomes strong and the quarks and gluons are bound together. Therefore, unlike leptons, isolated quarks can not be observed. They are *confined* within colorless composite particles known as *hadrons*.

A clear experimental evidence for the compositeness of hadrons came from deep inelastic scattering of electrons on protons [31, 32]. In these scattering processes the electron scatters of one of the constituents of the proton (generically called *partons*) and the latter breaks up and forms new hadrons. The angle at which the electron is scattered off the

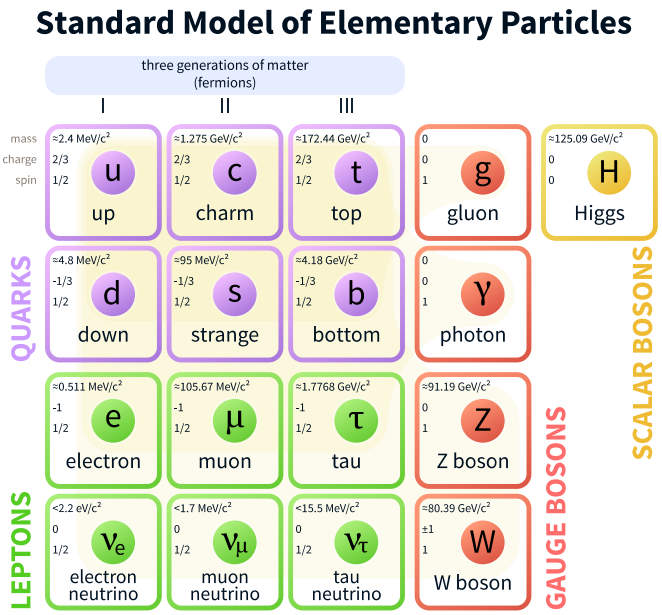



**Fig. 2.1.:** Summary of the particles included in the standard model of particle physics. Taken from Ref. [29].

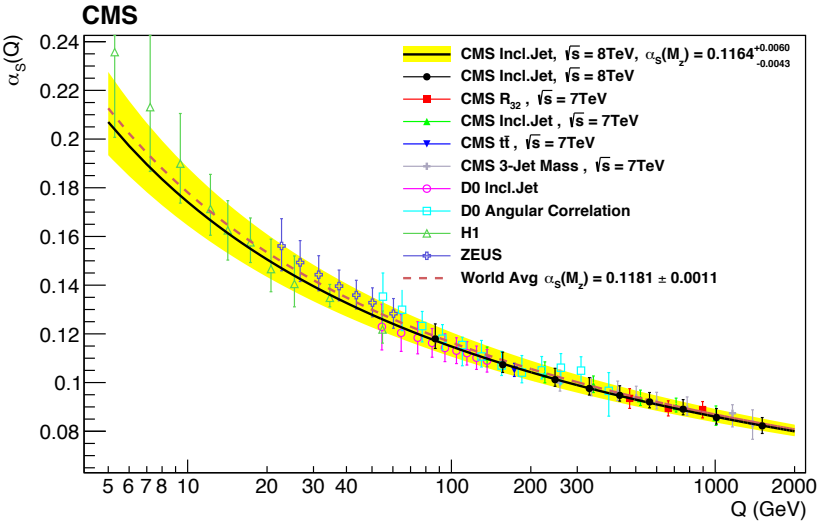



**Fig. 2.2.:** The running $\alpha_S(Q)$ as a function of the scale $Q$ is shown with a solid line and yellow uncertainty band. The dashed line represents the evolution of the world average. Results from various measurements of CMS, ATLAS, D0, H1, and ZEUS are superimposed. Taken from Ref. [30].

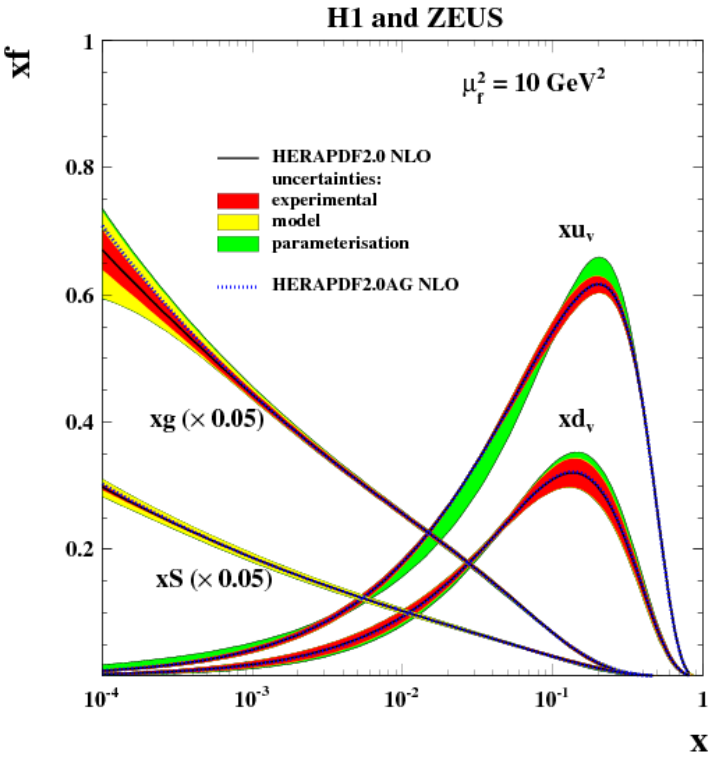



**Fig. 2.3.:** The parton distribution functions of the proton valence quarks $xu_v$, $xd_v$, the sea quarks $xS$ and the gluon $xg$ of HERAPDF2.0 NLO at $\mu_f^2 = Q^2 = 10\,\text{GeV}^2$. The gluon and sea distributions are scaled down by a factor of 20. The experimental, model, and parameterization uncertainties are shown. Taken from Ref. [33].

proton is directly linked to the fraction of the total proton momentum that is carried by the parton. This value is usually called the *Bjorken variable* $x$. Ever since the first measurement of deep inelastic scattering, the interactions of partons are characterized in terms of the scale (momentum transfer) of the interaction $Q^2$ and the Bjorken value $x$ of the participating parton(s). The structure of the proton is also evaluated in terms of $Q^2$ and $x$. These so-called parton distribution functions (PDFs) define the probability to find a parton with a specific value of $x$ when the proton is probed at a certain scale $Q^2$. As an example, the proton PDF at the scale $Q^2 = 10\,\text{GeV}^2$ as fitted to data of the H1 and ZEUS experiments is shown in Fig. 2.3. It is found that at large values of $x$ three quarks dominate the proton structure. Those quarks determine the macroscopic properties of the proton and are called *valence quarks*. They are surrounded by a so-called *sea* of quarks, anti-quarks and gluons that carry only very small momentum fractions of the proton and are thus dominant at low values of $x$.

The theory of QCD is able to predict a large set of phenomena in particle physics related to the strong interactions between quarks and gluons. This is especially important for hadron collisions such as at the LHC. Due to the large energy of the proton beams the partons of the protons can be considered unbound and the hard interactions between individual

partons can be calculated with a perturbative framework. Processes can be written as series expansions in orders of $\alpha_S$ and at large scales, when $\alpha_S \ll 1$, perturbative calculations can be performed. In soft interactions on the other side, at scales below about 1 GeV, this is not the case anymore and non-perturbative models based on phenomenological assumptions are necessary in order to understand and model these soft mechanisms. The basic idea of the *factorization theorem* of QCD is that hard processes and softer components are independent from each other. As a consequence, it is for example possible to write the total cross section for proton-proton interactions with a given final state $X$, $pp \to X$, as a convolution of the non-perturbative PDFs $f(x, \mu)$ (one for each parton) and the perturbative matrix element of the parton-parton scattering:

$$\sigma_{pp\to X} = \sum_{i,j} \int_0^1 \mathrm{d}x_1 \int_0^1 \mathrm{d}x_2 \int \mathrm{d}\Phi_X f_i(x_1, \mu) f_j(x_2, \mu) \frac{\mathrm{d}\sigma_{ij\to X}}{\mathrm{d}x_i \mathrm{d}x_j \mathrm{d}\Phi_X} , \tag{2.1}$$

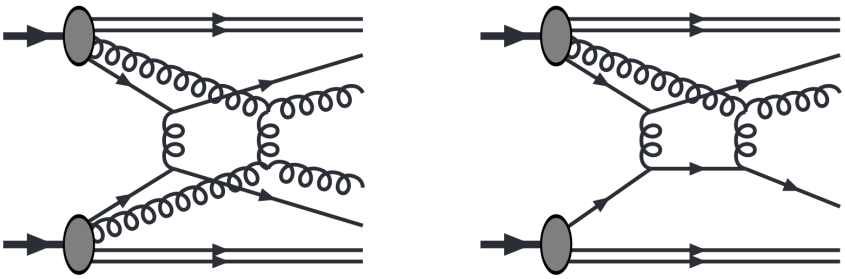
where $x_1$,$x_2$ are the Bjorken values of the two interacting partons, $i$,$j$ are the flavors of all available partons, and $\Phi_X$ the available phase space of the final state.

Detailed descriptions of the history and theory of QCD is given in the literature, for example in Ref. [34]. A phenomenological overview of some of the processes contributing to proton-proton collisions at the LHC shall be given in the following. Descriptions of how these are implemented and modeled by different hadronic interaction models commonly used in high energy physics is given afterwards.

## 2.2 The underlying event in proton-proton collisions at the LHC

At the LHC a large variety of processes contribute to the final state of proton-proton collisions. A hadronic collision can be described as a single *hard scattering* of two independent partons that exchange a large transverse momentum. This type of interaction is in general well described by perturbative QCD (pQCD). Additionally, the two colliding partons are accompanied by other quarks and gluons. Those are likely to interact as well and may emit particles before or after the interaction and create additional hadronic activity. All these processes are summarized in the so-called *Underlying Event*.

**Beam remnants (BR)**
The spectator partons to the hard scattering are usually referred to as the beam remnants. In the QCD picture, the former colorless protons break up and free partons are created that carry color charge. These need to rearrange again after the collision and form new colorless objects. The beam remnants mainly travel along the initial direction of the incoming beam particles. For 13 TeV proton-proton collisions this is at pseudorapidities of $\eta \approx 9$.



(a) (b)

**Fig. 2.4.:** (a) Two $2 \to 2$ scatterings, (b) a $2 \to 2$ scattering followed by a second scattering of a final state parton. Taken from Ref. [35].

**Initial and final state radiation (ISR and FSR)**

Partons may emit quarks or gluons in the initial or final state, i.e. before or after the hard scattering. These emissions are referred to as initial or final state radiation. The emitted partons hadronize and contribute significantly to particle multiplicities in the underlying event.

**Multiparton interactions (MPI)**

Next to the hard scattering, multiple interactions can take place between other partons. The scale of these additional parton-parton interactions is by definition smaller than the one from the hardest scattering. These additional interactions may occur between partons from different protons as well as between partons of the same proton after the hard scattering. Furthermore, scenarios have been discussed in which final state partons after the first interaction interact again with partons in the proton remnants [35, 36]. This is illustrated in Fig. 2.4. Proposals have been made to calculate multiple hard scatterings in the framework of pQCD ([37, 38]. Still, the differential partonic QCD cross section, $\mathrm{d}\hat{\sigma}/\mathrm{d}p_{\mathrm{T}}^2$ contains a term $\propto 1/p_{\mathrm{T}}^4$ and is therefore divergent for $p_{\mathrm{T}} \to 0$. This leads to an infinite cross section and the breakdown of pQCD. The contribution from soft MPI is therefore subject to phenomenological models [39, 40] and many experimental studies (see for example [41]).

In realistic hadronic collisions, the underlying event can not be separated from the hard scattering since the particles overlap in the final state and can hardly be distinguished in the detector. This is visualized in Fig. 2.5. A good understanding of all these mechanisms is therefore of great importance for any measurement at the LHC where the properties of hard scatterings are studied. Unfortunately, these processes mostly happen at small scales of $Q^2$. Perturbative calculations are not possible from first principles because of the large value of $\alpha_{\mathrm{S}}$ and phenomenological descriptions have to be tested with dedicated measurements defined to be sensitive to the underlying event.

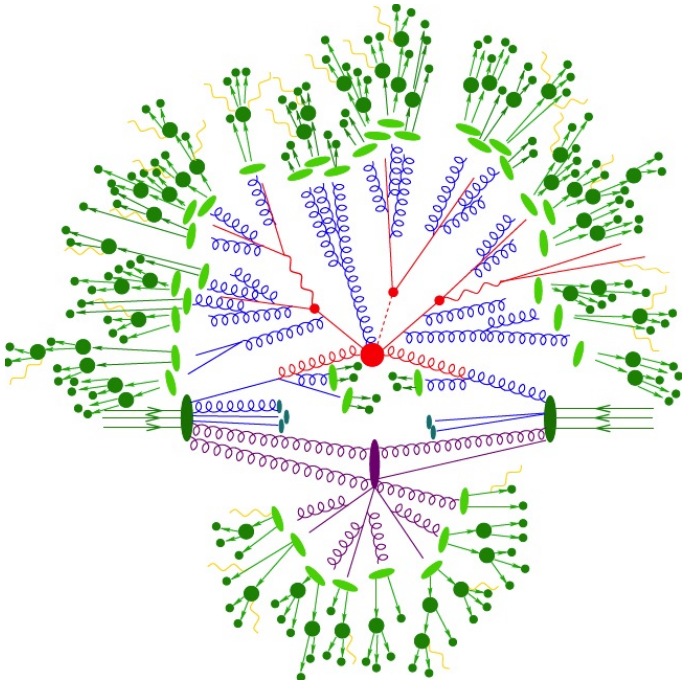


**Fig. 2.5.:** Illustration of a proton-proton collision. The colors indicate different processes in the collision: hard scattering (red), ISR and FSR (blue), MPI (purple), BR (cyan), parton hadronization (green), hadron decays (dark green), and QED Bremsstrahlung (yellow). Taken from Ref. [42].

The most common way to quantify the activity of the underlying event is to study the average charged particle density and average sum of the transverse momentum $p_{\mathrm{T}}$ as a function of the $p_{\mathrm{T}}$ of the hardest object in the event. In the case of hadronic final states, this is usually the leading particle or jet or dimuon pair in the case of a Drell-Yan event. This measurement was introduced by the CDF collaboration at the Tevatron [43–45] and was then also adapted by other experiments and colliders (for example [46–48]). The direction of the hardest object $\Phi_{\mathrm{hard}}$ in the transverse plane defines four regions in phase space that are studied independently, see Fig. 2.6. These regions are commonly defined as

- Towards region: $|\Phi - \Phi_{\mathrm{hard}}| < \pi/3$
- Transverse region: $\pi/3 < |\Phi - \Phi_{\mathrm{hard}}| < 2\pi/3$
- Away region: $|\Phi - \Phi_{\mathrm{hard}}| > 2\pi/3$

The transverse region can further by separated into the so-called TransMax and TransMin regions. Here the transverse region with the largest (smallest) activity is called the TransMax (TransMin) region. Additionally, the difference between those two regions (TransDiff) and the average can be studied (TransAvg). The differentiation between those regions allows to some degree to experimentally disentangle different contributions to the underlying event. The particles originating in the hard scattering fall into the towards and their recoil objects in the away regions, accompanied by hard radiation. The

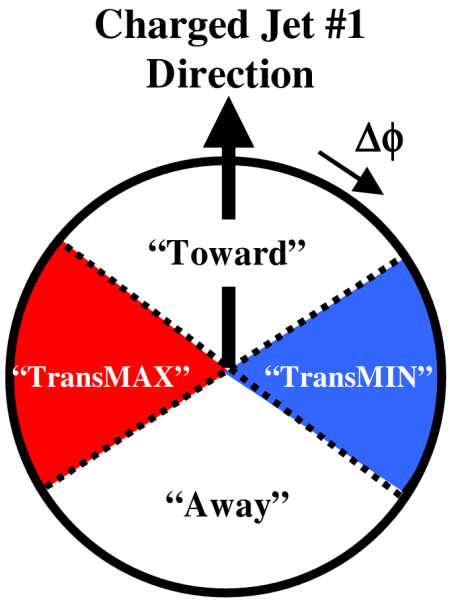



**Fig. 2.6.:** Illustration of the different regions in the azimuthal angle $\Delta\Phi$ relative to the direction of the leading charged jet in the event. The Toward, Away, TransMax, and TransMin regions are defined. Taken from Ref. [49].

transverse region on the other side contains mostly particles from MPI and the BR along with additional soft radiation. Moreover the TransMax region is expected to have more contribution from parton radiation than the TransMin region. By studying the TransDiff region it is possible to separate the contribution from radiation and from MPI [49]. The measurement of the average charged particle multiplicity as a function of the leading particle $p_{\mathrm{T}}$ by CMS at 13 TeV is shown as an example in Fig. 2.7. The typical shape of the curves with a steep rise at low $p_{\mathrm{T}}$ followed by a plateau region at higher $p_{\mathrm{T}}$ is seen.

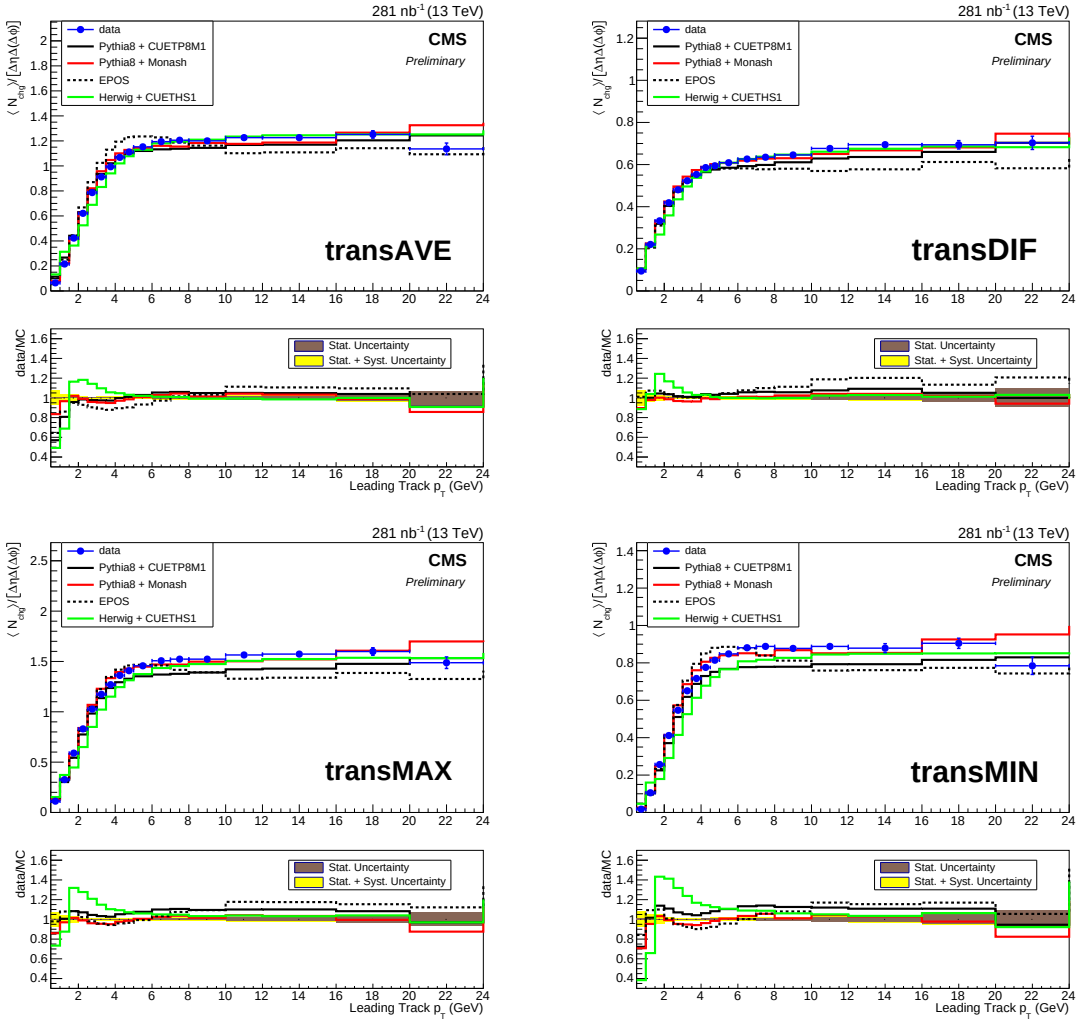



**Fig. 2.7.:** Comparisons of the average particle densities in the TransAvg (top left), TransDiff (top right), TransMax (bottom left), and TransMin (bottom right) regions with various simulations as a function of $p_{\mathrm{T}}$. Taken from Ref. [47].

## 2.3 Gribov-Regge theory

An elegant and widely used alternative way to describe hadronic interactions and especially soft process is the Gribov-Regge theory (GRT). As mentioned, soft processes are dominant in the underlying event of proton-proton collisions. A phenomenological ansatz is used in Regge theory and extended by Gribov in the language of quantum field theory [50]. Interactions are described in terms of an exchange of composite particles, so-called *reggeons*. Reggeons are groups of particles whose mass and spin are related in a common way. One example is the $\rho$-reggeon, which includes the ground state as well as excited states $[\rho(770), \rho_3(1690), \dots]$. In this picture, the scattering of $p + \pi^0 \to n\pi^+$, for example, can not only be described by an exchange of a $\rho$ meson but by all particles of the $\rho$-reggeon. The relation of the spin $\alpha$ to the squared mass of the particles $t$ can be approximated with linear parameterizations, referred to as *regge trajectories* which take the form

$$\alpha(t) = \alpha_0 + \alpha_1 \times t \; . \tag{2.2}$$

Examples are shown in Fig. 2.8. Physical particles are defined as *regge-poles* at values of $t$ where $\alpha(t)$ takes integer values.

An important application of GRT is the calculation of the total cross sections of hadron collisions [52]. From the quantum mechanical description of wave scattering, one obtains that the scattering amplitude $A(s,t)$ depends on the squared center-of-mass energy $s$ and the squared mass $t$ and spin $\alpha$ of the exchanged particle like

$$A\left(s,t\right) \propto s^{\alpha(t)} \; . \tag{2.3}$$

Using the optical theorem which connects the total scattering cross section to the forward scattering amplitude at $t = 0$, one obtains

$$\sigma_{\text{tot}} \propto \frac{1}{s}\text{Im}\left[A\left(s, t=0\right)\right] \propto s^{\alpha_0 - 1} \; . \tag{2.4}$$

The regge-trajectories of all known particles have in common that their intercept is $\alpha_0 < 1/2$ at $t = 0$. This results in a decreasing total cross section as the center-of-mass energy increases. This contradicts measurements at high center-of-mass energies (for the first time discovered in 1973 at the CERN ISR [53, 54]). A possible solution is the introduction of a new reggeon trajectory with $\alpha_0 \geq 1$ (see also Fig. 2.8), the so-called *pomeron* [55] which does not correspond to a well defined physical particle state. Most of the current models assume $\alpha_0 = 1.08$ which is obtained from experimental data of the total proton-proton cross section at high energies according to Eq. (2.4), see Fig. 2.9.

The interaction of two protons (and generally hadrons) at high energies is therefore expressed in the picture of GRT by the exchange of one or more pomerons between the participating partons. A schematic picture is shown in Fig. 2.10. This approach includes

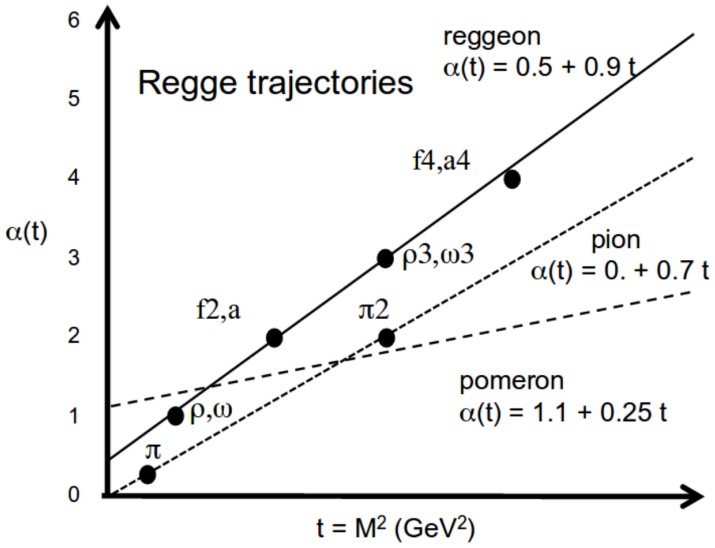



**Fig. 2.8.:** Example of regge trajectories of light mesons (solid lines) in the spin-$M^2$ plane. A possible pomeron trajectory is drawn as dashed line. Modified from Ref. [51].

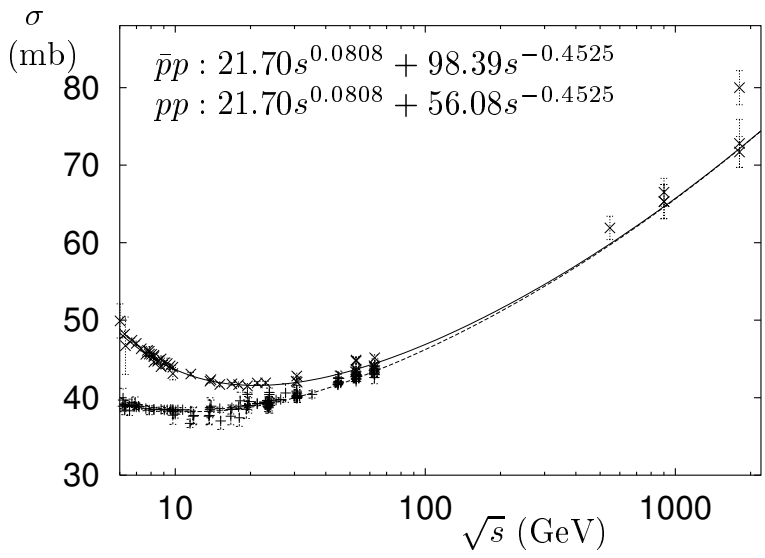



**Fig. 2.9.:** Total proton-proton cross section as a function of the center-of-mass energy. Experimental data is compared and the energy dependence is fitted. Modified from Ref. [52].

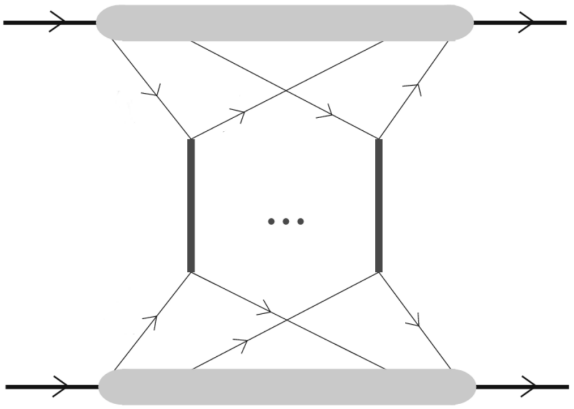


**Fig. 2.10.:** Schematic drawing of a hadron-hadron interaction in the picture of Gribov-Regge-Theory. Multiple exchanges of pomerons are allowed between the initial hadrons. Modified from Ref. [56].

by construction the treatment of multiple interactions. The GRT was extended to the *Parton-based Gribov-Regge Theory* in Ref. [56]. In this approach, pomeron exchanges are treated differently in three categories: Soft processes are described with soft pomeron exchanges as in GRT; hard interactions at scales where perturbative QCD is applicable are realized in forms of parton ladders that are evolved according to QCD rules; and semi-hard processes combine both by coupling soft pomerons to hard parton ladders. This model therefore proposes a complete formalism to treat all parts of hadronic interactions in a consistent way. There is no strict separation between the hard parton scattering and the mechanisms of the underlying event as it is formulated is QCD factorization. Still, the phenomenological effects such as MPI, parton showers or beam remnants are still there. Moreover, the GRT can directly handle any kind of hadronic interactions including those of light mesons as well as heavy nuclei. Event generators based on GRT are therefore widely used to model heavy ion collisions as well as extensive air showers induced by cosmic rays.

## 2.4 Physics in the very forward region

The very forward region ($|\eta| > 5$) provides a very special environment to investigate hadronic collisions. This phase space is different to what is accessible with detectors at central rapidities at the LHC due to simple kinematics. Figure 2.11 illustrates this by showing common acceptances as a function of $Q^2$ and $x$. One can see that, for example with the CASTOR calorimeter of CMS, regions with $Q^2$ as low as 9 GeV $^2$ and $x$ even below $10^{-6}$ can be reached. This allows to probe the proton structure and in particular the gluon density in a region that no other experiment has reached before. Moreover, regions are

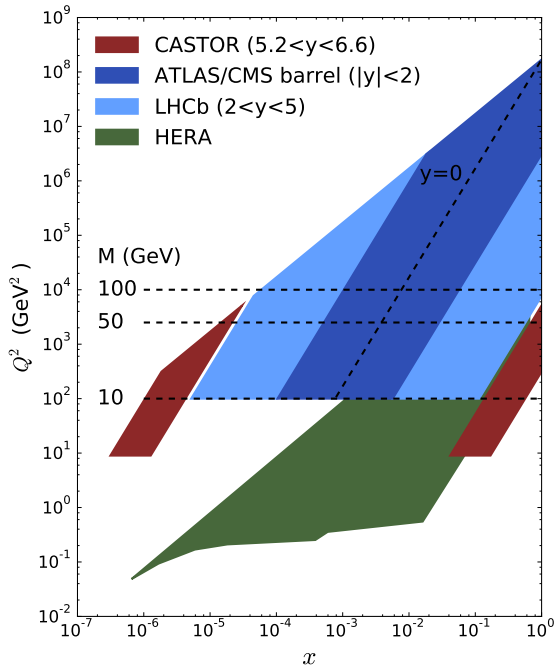



**Fig. 2.11.:** Kinematic acceptances as a function of $x$ and $Q^2$. The acceptance of the CASTOR calorimeter is shown by a red area, while the acceptances of other LHC experiments (blue) and the HERA collider (green) are given as a reference. The LHC acceptance is calculated for a center-of-mass energy of 13 TeV and optimistic experimental limits are used.

being reached where the perturbative QCD is likely to become imprecise and deviations appear.

At the same time, the very forward region is strongly affected by the mechanisms of the underlying event. This has been shown with the CASTOR calorimeter at center-of-mass energies of 0.9, 2.76, and 7 TeV [57]. Figure 2.12 emphasizes this by showing the average energy density per collision $\mathrm{d}E/\mathrm{d}\eta$ for PYTHIA 8 tune CUETP8M1 [58]. Different mechanisms of the underlying event have been switched off for illustration purposes. It can be observed that the energy carried by particles in the very forward phase space is significantly influenced by these mechanisms. When MPI and ISR are switched off in PYTHIA 8, the energy distribution becomes much narrower at $\eta \approx 9$. While the energy density significantly drops at $\eta \approx 6$. The effect of FSR being switched off becomes – in this particular observable – significantly visible only at $\eta > 7$.

Measurements in the very forward direction therefore have the great potential to study hard scatterings at low scales and lowest values of Bjorken $x$ as well as to study the soft particle production of the underlying event in a complementary way to common underlying event measurements. Furthermore, these measurements are a benchmark test to study soft and semi-hard pomeron interactions in the GRT picture. An extensive overview of the physics potential in the forward region at the LHC can be found in Ref. [21].

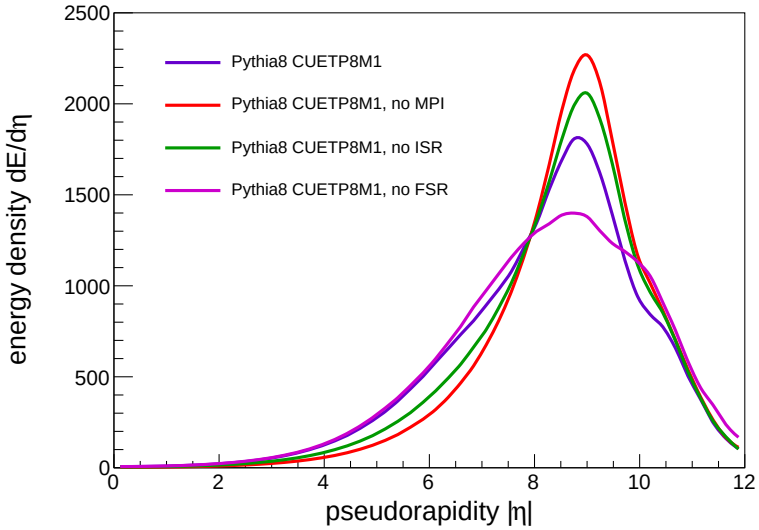



**Fig. 2.12.:** Average energy density $\mathrm{d}E/\mathrm{d}\eta$ as a function of pseudorapidity for PYTHIA 8 tune CUETP8M1. The purple line shows the default tune CUETP8M1, the red, green, and pink line tune CUETP8M1 where the modeling of multiparton interactions (MPI), initial-state radiation (IRS), or final-state radiation (FSR) are switched off respectively. Significant changes in the energy densities at $\eta \approx 6$ can be observed.

## 2.5 The physics of extensive air showers

Cosmic ray particles reach the Earth from galactic and extragalactic sources with very high energy. The highest primary energies ever measured are as large as $10^{20}$ eV and originate most likely in astrophysical objects outside the Galaxy [59]. The exact sources and acceleration mechanisms, though, are still unknown. Due to the steep decrease of the particle flux (see Fig. 2.13), direct measurements of primary cosmic rays with energies exceeding $10^{15}$ eV are not possible. Therefore, indirect measurements of the particle cascades produced by the incoming primary particles in the earth's atmosphere are needed in order to answer open questions such as their origin or mass composition. These *extensive air showers* are systematically studied for example at the Pierre Auger Observatory [19, 20] or the Telescope Array [60, 61]. For a detailed review of the field see for example [18, 62].

The development of extensive air showers as well as the importance of hadronic interactions is extensively described in the literature – see for example [63] – and only the basic principles shall be discussed here. The first interaction of the primary cosmic ray particle and air nuclei can be described as a nucleon-nucleon interaction producing a large number of secondary particles. These are mainly mesons and light baryons that in the following

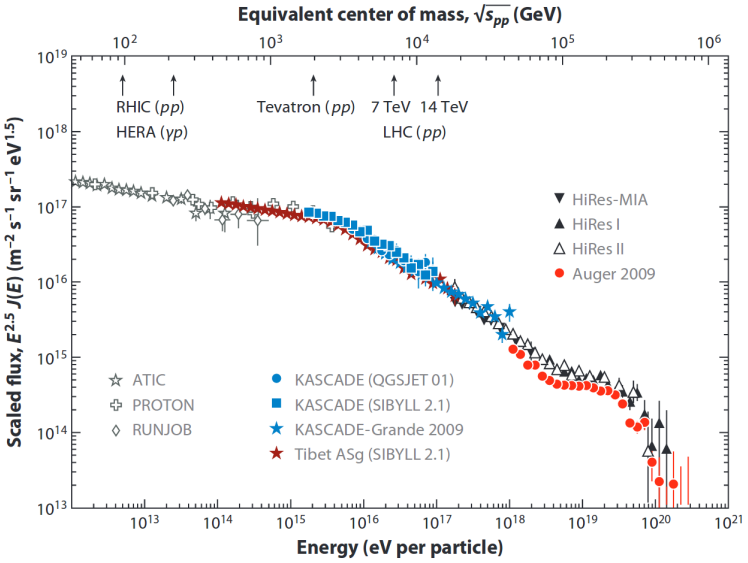



**Fig. 2.13.:** Flux of cosmic rays arriving at the Earth. The flux is scaled with the primary energy in order to emphasize structures in the spectrum. The equivalent center-of-mass energy for protons as cosmic ray particles is also shown. Taken from Ref. [63].

decay or interact again in the atmosphere. At the subsequent stages interactions of mesons with air nuclei become very important:

$$\text{First interaction: } p + N \rightarrow \pi^{\pm}, \pi^0, K^{\pm}, K^0, p, n, \ldots \tag{2.5}$$

$$\text{Later interactions: } (\pi^{\pm}, p, \ldots) + N \rightarrow \pi^{\pm}, \pi^0, K^{\pm}, K^0, p, n, \ldots \tag{2.6}$$

Three main components that are at least partially decoupled from each other develop in the shower, the electromagnetic, muonic and hadronic component, as it is illustrated in Fig. 2.14. Given their short lifetime ($c\tau \approx 25\,\text{nm}$), neutral pions almost immediately decay into two photons without any further interaction. The photons on the other hand are likely to produce electron-positron pairs which on the other hand emit again photons through bremsstrahlung. This simple evolution of the electromagnetic shower can be parameterized in the *Heitler model*. The splitting as well as the ongoing production of $\pi^0$ lead to a fast increase of the number of electrons, positrons, and photons in the shower until their respective energies are low enough so that ionization losses dominate and they are stopped. A similar approximation can be made for the muonic component of the shower in the Heitler-Matthews model [64]. Stable hadrons interact with the air nuclei and produce new generations of hadrons until their energy is low enough for them to decay before any further interaction. This leads to an increasing amount of muons and a decreasing amount of hadrons in the later shower stages. Those muons have high enough energy to reach the ground. An illustration of this shower development can be found in Fig. 2.15. This formalism can be generalized to also account for different primary nuclei with mass $A$, by treating the nucleons of one primary independently. This leads to

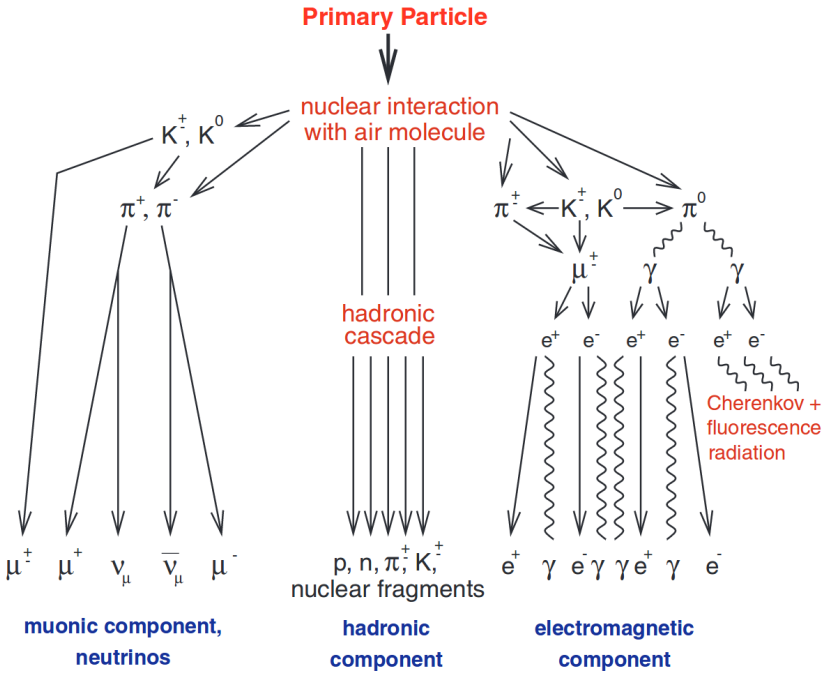



**Fig. 2.14.:** Overview of the interaction products and shower components of cosmic ray induced an extensive air shower. The dominant production and decay mechanisms are illustrated leading to the electromagnetic, muonic and hadronic shower components. Taken from Ref. [65].

$A$ subshowers that share the primary energy $E_0$ equally. Although this picture is rather simplistic, some main properties of air showers can already be derived:

### Depth of the shower maximum

The depth of the shower maximum $X_{\max}$ is referred to as the integrated matter density column that a shower traversed in the atmosphere when the maximum number of charged particles in the shower is reached. As one can see in Fig. 2.15, around $X_{\max}$ the cascade is dominated by electromagnetic particles. The shower maximum depth depends on some of the properties of the primary particle. In the picture of the Heitler-Matthews model, one obtains a dependence of

$$X_{\max} \propto \ln\left(E_0/A\right) \ . \tag{2.7}$$

The value of $X_{\max}$ provides therefore a possible measure for the primary mass if the total energy of the shower is known. Due to strong shower-to-shower fluctuations, the experiments commonly measure the average $\langle X_{\max} \rangle$ and the strength of the fluctuations $\sigma\left(X_{\max}\right)$ as a function of the shower energy in order to calculate the average mass composition of cosmic rays at given energies [24].

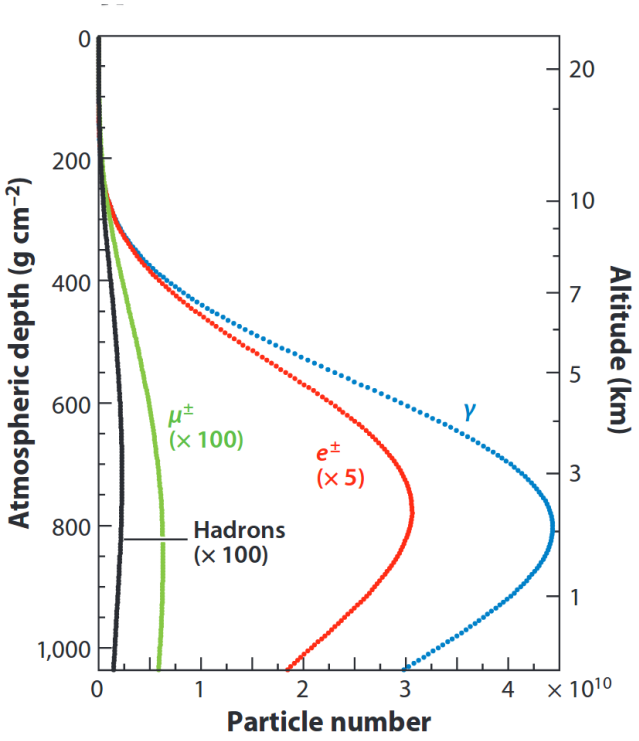



**Fig. 2.15.:** Illustration of the longitudinal shower profiles for vertical, proton-induced showers at $10^{19}$ eV. The particle numbers are given as a function of the atmospheric depth and the equivalent height above sea level. Different particle species in the shower are illustrated with different color. Modified from Ref. [63].

### Electron and muon numbers

In the Heitler-Matthews model, the total number of electrons in the shower maximum is independent of the primary mass and linear proportional on the primary energy:

$$N_{\mathrm{e,max}} \propto E_0 \ . \tag{2.8}$$

This is however not true for the number of muons. Under the assumption that in every hadronic collision $n_{\mathrm{tot}}$ secondary particles are produced and out of theses, $n_{\mathrm{ch}}$ are charged hadrons that each produce one muon after they decay, the total number of muons in an air shower can be estimated to be

$$N_\mu \propto A^{1-\alpha} E_0^\alpha \quad \text{with} \quad \alpha = \frac{\ln n_{\mathrm{ch}}}{\ln n_{\mathrm{tot}}} \ . \tag{2.9}$$

This implies, that the electron number in the shower maximum is a good measure for the shower energy and the ratio to the muon number provides a complementary way to estimate the primary mass.

In reality, the development of air showers can not be easily approximated. The formation of air showers is governed by all secondary particles produced in high- and low-energy collision. The most important secondary particles, though, are the ones with highest

energy because they will drive the development of the air shower. Moving from the frame of the air shower (where the target air nuclei are in rest) to the center-of-mass frame as it is observed in collider experiments, it becomes clear that the collision mechanisms that dominate particle production in the very forward region are of significant importance for air shower development. It is in this phase space that the energy density is highest (see also Fig. 4.1). Unfortunately, these processes are the ones that are the least well theoretically understood, as outlined before. Hadronic event generators are therefore required to interpret the properties of the measured air showers. This is a major limitation. Differences is the phenomenological description of the relevant processes lead to ambiguities and uncertainties in the explanations and interpretations of air showers. This is especially true for the highest energies, where no reference data from hadron colliders is available to constrain the models.

The relation between particle production mechanisms and air showers can be studied in the picture of the Heitler-Matthews model for example with the two parameters, elasticity $\kappa_{\mathrm{el}}$ and pion charge ratio $c$ of hadronic collisions [28].
The elasticity is defined by the fraction of the projectile energy that is carried by the most energetic particle after the collision, in the following referred to as the leading particle:

$$\kappa_{\mathrm{el}} = E_{\mathrm{leading}}/E_0 \ . \tag{2.10}$$

While the Heitler-Matthews model assumes an equal share of the available energy among the secondary particles, a higher elasticity will significantly effect the depth of the shower maximum. Leading particles carry a large fraction of the total energy deeper into the atmosphere and the overlapping showers of different interaction generations will produce larger $\langle X_{\mathrm{max}} \rangle$.
The pion charge ratio, can be defined as the number of neutral pions divided by the number of all pions that are produced in a collision. This is equivalent to

$$c = \frac{n_{\pi^0}}{n_{\pi}} \ . \tag{2.11}$$

The pion charge ratio has direct impact on the parameter $\alpha$ in Eq. (2.9) and therefore on the number of muons in an air shower. This is even more interesting since it was observed that the measured number of muons in inclined air showers exceeds those of model predictions [23]. This could for example be explained by an increase of charged pion production, which corresponds to a reduced pion charge ratio $c$ [66].

It has been shown that the data recorded by the LHC experiments provide great opportunities to validate and improve existing models [67]. Moreover, hadronic interaction models have been tuned to the provided data at $\sqrt{s}$ = 0.9, 7, and 8 TeV and subsequently theoretical uncertainties on cosmic ray data have been reduced. Still, there is significant room for improvements. The tension between the measured and predicted muon numbers

is still unsolved. The models therefore need to be continuously validated as more data are analyzed and the center-of-mass energy at the LHC rises.

## 2.6 Hadronic event generators

Hadronic event generators play an important role in analyzing and interpreting hadronic collisions – in cosmic ray physics as well as at the LHC. In order to perform standard model precision measurements as well as searches for physics beyond the standard model, the LHC experiments face the challenge to understand the full final states of proton-proton collisions as well as possible. This becomes even more important as luminosities at the LHC increase and more collisions happen simultaneously. The processes forming the underlying event remain one of the major challenges. Multipurpose hadronic event generators are an important tool in order to estimate the influence of the underlying event to the studied observables and to subtract the corresponding background contributions. Several generators are commonly used by the LHC experiments, among which are PYTHIA and EPOS LHC. These models have different approaches on how to model hard and soft interactions and shall be briefly described here. Most attention will be given to their modeling of soft multiparticle production.

Beyond the LHC, EPOS LHC, together with QGSJETII and SIBYLL, are also widely used to interpret cosmic ray induces air showers. Short descriptions of these models are provided since their performance will be investigated in this work.

### 2.6.1 PYTHIA 8

The history of the PYTHIA event generator starts in the late 1970s, emerging from the JETSET program that implemented the *Lund Model* of jet fragmentation [68]. PYTHIA itself was introduced in Ref. [69]. The focus was to describe high $p_{\mathrm{T}}$ physics in hadronic collisions. Later the program was continuously extended to include initial and final state radiation as well as beam remnant treatment and multiparton interactions [40, 70, 71]. Thus it was one of the first general purpose Monte Carlo generators for high energy hadronic collisions that included perturbative calculations of the hard scattering as well as an extensive phenomenological description of the soft and semi-hard processes. PYTHIA 6 [72] was for a long time the benchmark event generator for high energy physics. In recent years it is widely replaced by an improved C++ reimplementation in version 8 [27]. For a complete description, see [27, 72].

**Hard scattering**
A large variety of hard scattering processes are available within PYTHIA 8. Standard parton-parton scatterings in $2 \to 2$ QCD processes are possible as well as electroweak processes and the production of top quarks and the Higgs boson. Various sets of parton

distribution functions (PDFs) are implemented inside PYTHIA 8 and additional PDFs can be added via an interface to the LHAPDF [73] software. The matrix element of the hard parton scattering can be calculated inside PYTHIA but also be interfaced to more powerful tools such as MADGRAPH [74] or POWHEG [75]. Standard model processes are extended with processes predicted by models describing beyond the standard model physics such as supersymmetry.

**Hadronization: Lund fragmentation**

The hadronization of colored quarks or gluons into color-neutral hadrons is done via the Lund fragmentation (or string fragmentation) method. The Lund model is a phenomenological approach that was developed in the 1980s [76–78] and was the foundation of the PYTHIA development. The basic principle is best explained in the example of two outgoing quarks. The dipole color field between the quarks is represented by a one-dimensional string that connects the two quarks.[1] The color field potential between the quark pair is assumed to be linear in the distance $r$, with a field strength of about 1 GeV/fm. As the two ends move away from each other, the potential increases until the energy of the string reaches the order of hadron masses. At this stage, a pair of quarks (or di-quarks) is created and the initial string is split up into two strings connecting the newly created quarks and the original quarks. These splittings continue until the invariant mass of a string is small enough to form on-shell hadrons. In this picture, each final state hadron corresponds to a small piece of the initial string. The splitting of the strings is parameterized in the Lund model with a splitting function of the form

$$f(z) = \frac{(1-z)^a}{z} \times e^{-bm_{\mathrm{T}}/z} \tag{2.12}$$

where $m_{\mathrm{T}} = \sqrt{m^2 + p_{\mathrm{T}}^2}$ is the transverse mass of the hadron and $z$ the momentum fraction of the original string [77]. In this formalism, $a$ and $b$ are free parameters that need to be chosen in agreement with data. The most powerful data in this respect are clean di-jet event at the LEP electron-positron collider. The current values in PYTHIA 8.2 are $a = 0.68$ and $b = 0.98$.

**Multiparton interactions**

The divergence of the partonic cross section for $p_{\mathrm{T}} \to 0$ is regularized in PYTHIA by introducing a dampening parameter $p_{\mathrm{T}0}$. The partonic cross section $\hat{\sigma}$ then changes to

$$\frac{\mathrm{d}\hat{\sigma}}{\mathrm{d}p_{\mathrm{T}}^2} \propto \frac{\alpha_S\left(p_{\mathrm{T}}^2\right)}{p_{\mathrm{T}}^4} \to \frac{\alpha_S\left(p_{\mathrm{T}0}^2 + p_{\mathrm{T}}^2\right)}{\left(p_{\mathrm{T}0}^2 + p_{\mathrm{T}}^2\right)^2} \,. \tag{2.13}$$

In addition, the parameter $p_{\mathrm{T}0}$ is set to be energy dependent in the form

$$p_{\mathrm{T}0}\left(E_{CM}\right) = p_{\mathrm{T}0}^{\mathrm{ref}} \times \left(\frac{E_{\mathrm{CM}}}{E_{\mathrm{CM}}^{\mathrm{ref}}}\right)^{E_{\mathrm{CM}}^{\mathrm{pow}}} \,. \tag{2.14}$$

[1] Strings might also connect to pairs of quark and anti-quark.

In PYTHIA 8, the default reference energy $E_{\mathrm{CM}}^{\mathrm{ref}}$ is 7 TeV and the scaling factor $E_{\mathrm{CM}}^{\mathrm{pow}}$ and $p_{\mathrm{T}0}^{\mathrm{ref}}$ are free parameters. While in previous versions of PYTHIA, the simultaneous interactions are assumed to be independent of each other, rescattering is added as of PYTHIA 8.1. Final state particles created in one MPI may interact again with other partons from the original proton (see Fig. 2.4). Since protons are extended objects, the impact parameter of the colliding protons will also affect the amount of MPI. Various parameterizations are implemented in PYTHIA 8 to scale the MPI probabilities. The most commonly used parameterization is a double-gaussian distribution which assumes that a fraction $\beta$ of the parton content is located within an inner radius $r_\beta$ of the proton. This leads to a radial dependent MPI probability

$$P(r) \propto \frac{1-\beta}{r_{\mathrm{tot}}} e^{\left(-r^2/r_{\mathrm{tot}}^2\right)} + \frac{\beta}{r_x} e^{-r^2/r_\beta^2} \tag{2.15}$$

which can be rewritten with $\hat{r} = r/r_{\mathrm{tot}}$ and $\hat{r}_\beta = r_x/r_{\mathrm{tot}}$ to

$$P(\hat{r}) \propto (1-\beta)\, e^{\left(-\hat{r}^2\right)} + \frac{\beta}{\hat{r}_\beta} e^{-\hat{r}^2/\hat{r}_\beta^2} \,. \tag{2.16}$$

This leaves two free parameters for the scaling of MPI according to the density profile of the protons. The additional interactions of MPI are added ad-hoc to the final state after the hard interaction is calculated. The momentum transfer of these additional interactions is therefore by definition smaller that the one from the hard scattering. After all MPI are defined in the generator, the ISR and FSR are added to every single interaction.

**Beam remnants and color reconnection**

As a consequence of the above, a hadron-hadron collision is becoming very complex. In the view of dipole color fields in the Lund model, many of those fields are created by the final state particles of the hard scattering, the MPI, parton showers, and the remaining partons in the beam remnant. It seems natural to assume that some of those fields overlap. Therefore, color reconnection is introduced as a way to describe these interference effects. The description of this field correlation is somewhat arbitrary [79] but manages to describe the effect seen in data [27, 40]. The default PYTHIA model for color reconnection assigns to every system of the final state a probability to reconnect with the next hardest system:

$$P = \frac{{p_{\mathrm{TRec}}}^2}{({p_{\mathrm{TRec}}}^2 + p_{\mathrm{T}}^2)} \text{ with } p_{\mathrm{TRec}} = R \times p_{\mathrm{T}0} \,. \tag{2.17}$$

Here $p_{\mathrm{T}0}$ is the MPI dampening parameter from Eq. (2.14). According to this probability, the color strings are merged in a way to create the shortest total string length. Together with this model, PYTHIA 8 also features new implementations of color reconnection, a more QCD-like model and a model called *gluon move model*. They increase the potential of PYTHIA but are not relevant for this work.

Throughout this work, PYTHIA is used in version 8.2 only. Special attention will be payed to the parameter tunes CUETP8M1 [58] and tune 4C combined with the MBR model for diffraction [80].

### 2.6.2 EPOS LHC

The EPOS [81, 82] event generator relies on the parton-based GRT to simulate the final states of hadronic collisions. EPOS 1.99 was released in 2009 [83] and included improvements based on measurement at the Tevatron and RHIC colliders as well as from air shower measurements. It received an update in 2015 called EPOS LHC [84]. This update included some adjustments to free model parameters as well as a redefinition of collective effects in small systems and will mainly be used in his work. The acronym represents the main features that are explained in some detail. Some accompanying illustrations are shown in Fig. 2.16.

**Energy conserving quantum mechanical multiple scattering approach**
In EPOS, all scatterings between the colliding hadrons are treated simultaneously in one consistent picture. This is a major difference to other models used in high energy physics, as for example PYTHIA, where hard scatterings, parton showers and MPI are treated independently. The multiple interactions are connected in EPOS via the mechanisms of energy sharing. It is ensured that the overall collision energy is conserved and shared among the individual interactions.

**Partons (parton ladders)**
As proposed by high energy GRT, the exchange of pomerons in the scattering processes are realized with parton ladders between the interacting partons (see Fig. 2.16, top). These parton ladders include the emission of space-like (ISR) and time-like (FSR) parton showers. Each parton ladder is then translated into two color strings similar to the Lund model. The strings fragment into hadrons in a purely phenomenological way. Also here LEP data are used to constrain the hadronization model.

**Off-shell remnants**
In addition to the parton scatterings, the projectile and target remnants are treated in the same consistent picture of energy sharing. After the construction of the parton ladders, the remnants are treated as colorless but excited (off-shell) states that also hadronize in the later evolution.

**Splitting of parton ladders**
The splitting of parton ladders describes rescattering of final state partons with partons from the original hadrons. The probability of parton rescattering is increasing with the density of the collision, for example with the mass number of the colliding hadrons or nuclei. In the EPOS picture, these ladder splittings create areas of phase space where

an increased density of parton ladders and subsequently colored strings is created, see Fig. 2.16. These dense regions are treated in a collective hadronization. The string segments in the dense region are referred to as the *core* region. The core is made of clusters in bins of $\eta$ to conserve the local energy density distribution. Each cluster afterwards hadronizes collectively with additional radial flow. In EPOS 2 (currently not publicly available) these collective hadronizations are calculated in a full hydrodynamical approach assuming a fluid quark-gluon-plasma. The exact mathematical implementation is beyond the scope of this work and can be found in Ref. [81, 84, 85]. Collective phenomena are discovered to be present in proton-proton collisions at the LHC [86], which demonstrates the importance of their implementation in EPOS LHC.

### 2.6.3 QGSJetII and Sibyll

Two other hadronic event generators that are widely used in cosmic ray physics are QGSJETII [87] and SIBYLL [88]. Both are based on GRT but differ in some aspects.

**QGSJETII**

The QGSJET family of event generators is based on the Quark-Gluon String model. The main feature of this model is a fully consistent implementation of pomeron exchanges at all order in the frame of GRT. The latest version QGSJETII.04 includes additional corrections for pomeron loops, which makes it unique among GRT-based models. In contrast to EPOS LHC though, no collective effects are implemented.

**SIBYLL**

The version SIBYLL 2.1 [88] was released in 1999. It was designed to study cosmic ray air showers and has been very successful therein. Furthermore, also many observables at colliders have been reproduced well. A new version SIBYLL 2.3 was released in 2016 [89].[2] The new version incorporates tuning to LHC data as well as an improved description of bream remnants and charmed particle production. The SIBYLL models are based on the Dual Parton Model (DPM), combining concepts from GRT, a minijet model and Lund fragmentation. In SIBYLL, there are also no collective effects implemented.

**CRMC**

The CRMC software [90, 91] is an interface tool for hadronic event generators used in cosmic ray physics. It can calculate hadronic collisions with specific energies, both in the center-of-mass or laboratory frame and provides the output in common formats such as ROOT [92] or HEPMC [93]. It therefore provides the interface to systematically study the predictions of these models also for proton-proton collisions at LHC energies. All predictions of EPOS LHC, QGSJETII, and SIBYLL used in this work are obtained with CRMC version 1.6.0, despite SIBYLL 2.1 which is only available in version 1.5.7.

[2] Shortly before the submission of this thesis, an updated, currently not publicly available, version SIBYLL 2.3c was provided by the authors. For consistency reasons, all results and conclusions in this work are provided for SIBYLL 2.3 only. A comparison to SIBYLL.23c is given in appendix E.

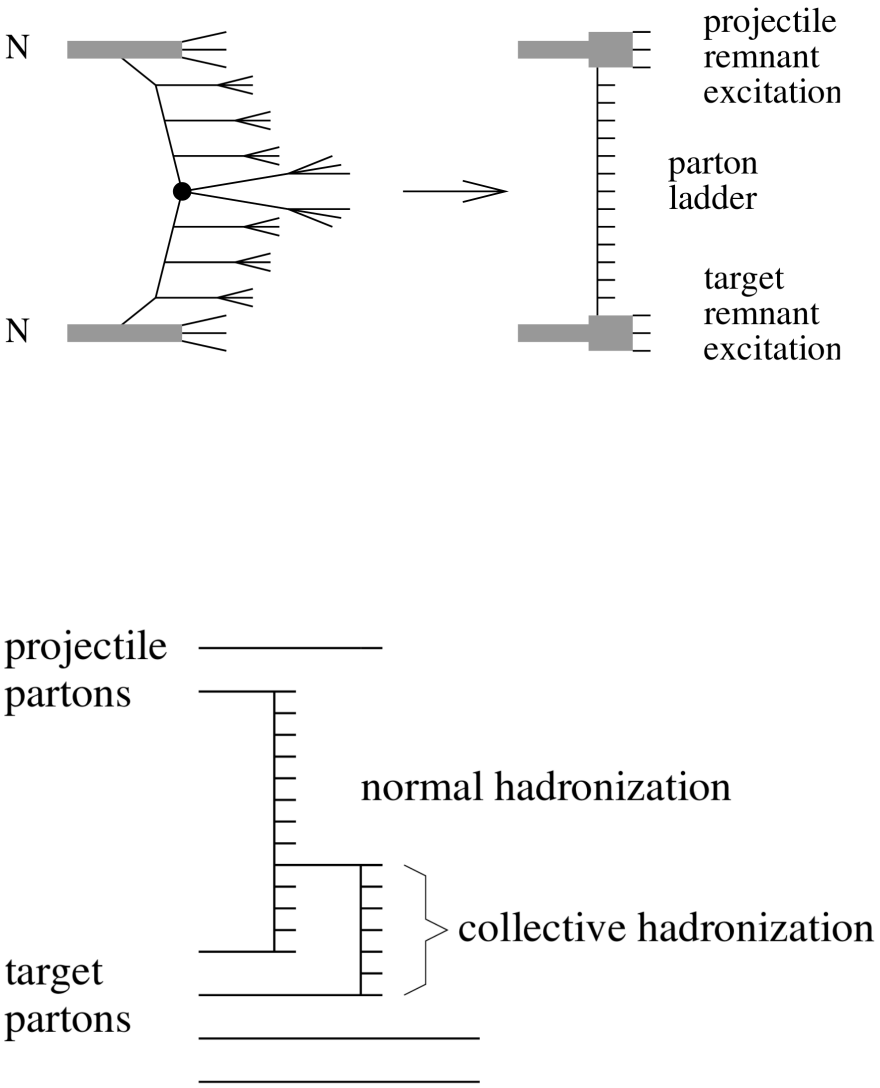



**Fig. 2.16.:** Illustration of hadronic collisions as they are implemented in EPOS. The top panel shows the transition from the parton view to the representation in parton ladders, the bottom panel the implementation of ladder splitting and the resulting collective hadronization. Modified from Ref. [81].

# 3 Experimental setup

> *An experiment is a question which science poses to Nature, and a measurement is the recording of Nature's answer.*
>
> **— Max Planck**
> (Physicist, Nobel Laureate)

The possibilities to build a large hadron collider at the European Organization for Nuclear Research (CERN) have been studied and discussed as early as in 1984 [94]. At that time, the predecessor electron-positron collider LEP was under construction at CERN. Still the potential of studying high energy hadron collisions was considered worth initiating long discussions as well as efforts in technological research and development. Today, the Large Hadron Collider (LHC) is the largest and most powerful particle accelerator ever built. It was constructed in the LEP tunnel after the shutdown in 2000. It has a circumference of roughly 27 km and uses superconducting magnets providing fields of up to 8.33 T in order to accelerate protons to an energy of currently 6.5 TeV. An overview of the LHC accelerator complex is found in Fig. 3.1. A detailed description of the LHC accelerator can be found in Ref. [1]. After the first operations in 2008 protons and lead ions have been collided at center-of-mass energies of 900 GeV, 2.76, 5.02, 7, and 8 TeV. Since June 2015, proton-proton collisions at the currently largest available center-of-mass energy of 13 TeV can be achieved. The LHC thus provides a unique environment to investigate hadronic collisions at the highest energies. These are studied by six experiments: ALICE [8], ATLAS [9], CMS [10], LHCb [11], LHCf [12], and TOTEM [13]. They all have different detector designs optimized to fulfill their major respective science goals.

In this work experimental data of proton-proton collisions recorded with the CMS experiment at the beginning of LHC Run 2 with $\sqrt{s} = 13\,\text{TeV}$ are studied. While the main goal for this run period is to provide high luminosities to the experiments, a special data-taking period took place in the very first weeks providing proton-proton collisions with low beam intensities. Thus, the average number of interactions per bunch crossing (*pile-up*) ranged between 5 and 30 %. These conditions are excellent to perform physics analyses in the very forward direction with the CASTOR calorimeter. Due to problems with the cryogenic system of the CMS solenoid, the latter was not turned on during this time. The lack of

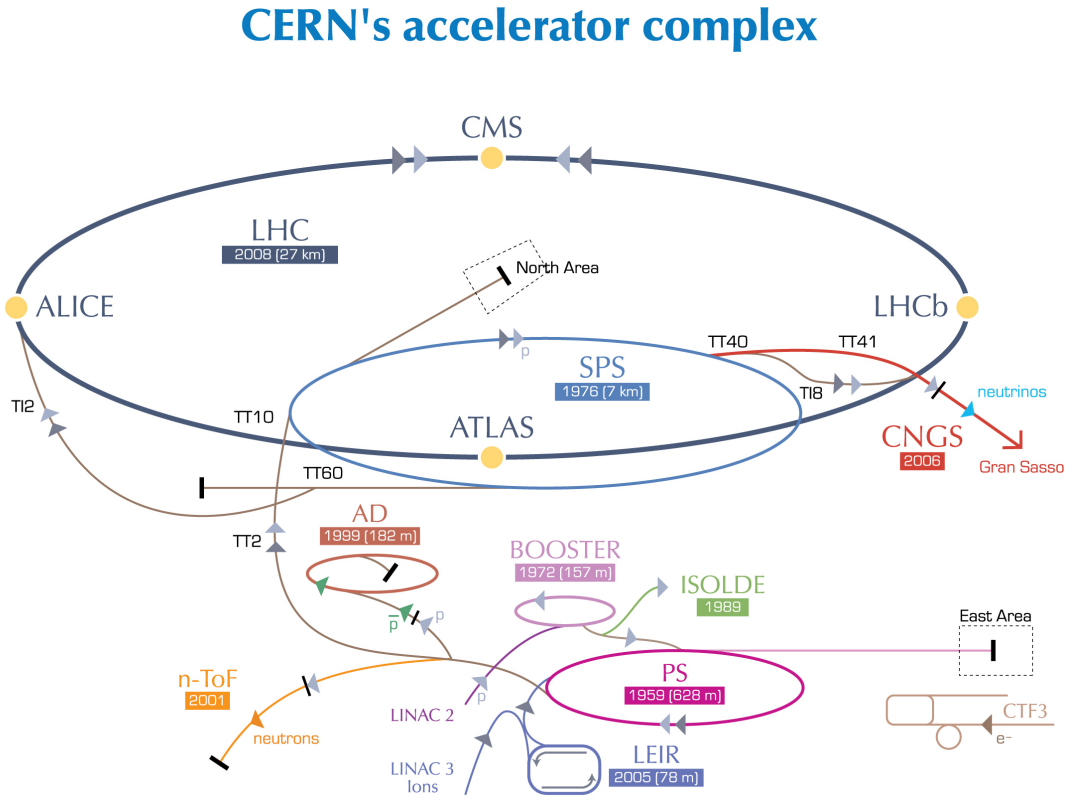



**Fig. 3.1.:** Illustration of the LHC accelerator complex and experiments at CERN. In particular, the LHC and its whole injection and pre-accelerator chain is shown. Picture modified from Ref. [96].

a magnetic field turns out to be an advantage for measurements with CASTOR. Since normally the response of parts of the calorimeter is significantly reduced by stray magnetic fields, the performance of CASTOR was limited in the past [57, 95]. On the other hand, the lack of a magnetic field reduces the capabilities of other subdetectors of CMS. Since charged particles are not bend within the detector volume, their transverse momentum can not be measured by the tracker. This also restricts for example the calibration of jets. Still, this set of data at low luminosities provides unique opportunities.

The main features of the CMS detector are described in the following chapter. Most attention will be given to the subsystems that are of explicit importance to this work, namely the inner pixel tracker, the hadron forward, and the CASTOR calorimeter. A complete description of the CMS experiment and all subsystems can be found in Ref. [10].

## 3.1 The CMS experiment at the CERN LHC

With a length of about 29 m and a diameter of about 15 m, the Compact Muon Solenoid (CMS) experiment is one of the largest experiments at the CERN LHC. A drawing of the

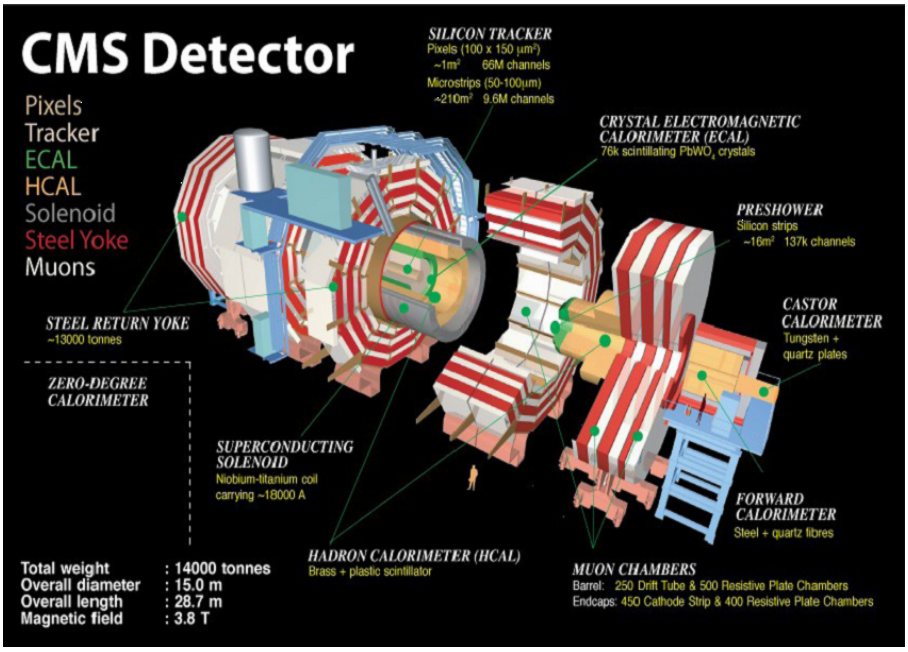



**Fig. 3.2.:** Schematic drawing of the CMS detector. All subsystems are shown and annotated.

CMS detector with all subsystems is also shown in Fig. 3.2. Name giving features are a precise muon detection system and a superconducting solenoid magnet providing a magnetic field of 3.8 T. The CMS coordinate system is defined such that the $x$-axis points south towards the center of the LHC ring, the $y$-axis points vertically upward and the $z$-axis in the direction of the beam to the west. The angle $\phi$ is defined in the $x-y$-plane and the pseudorapidity $\eta$ in a way that positive $\eta$ is towards positive $z$ according to

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \ . \tag{3.1}$$

The innermost component, nearest to the luminous region (also called interaction point, IP), is the silicon pixel tracker. It consists of three cylindrical layers of hybrid silicon pixels at radii of 4.4, 7.3 and 10.2 cm and two forward pixel discs covering radii from about 6–15 cm. The pixel tracker covers the pseudorapidity range $-2.5 < \eta < 2.5$ and provides three hit points over almost the full pseudorapidity range. Figure 3.3 shows the geometrical structure of the pixel layers as well as the hit coverage as a function of pseudorapidity. The pixel tracker is extended by the silicon strip tracker which consists of 10 cylindrical layers and 12 discs. It has the same acceptance in pseudorapidity as the pixel tracker and extends in radius up to 116 cm. Under normal running conditions, charged particles are bend inside the tracker volume and the charge and transverse momentum of the particles can be identified. Since the solenoid magnet was not operational in the beginning of the 2015 data taking period, the standard reconstruction (which assumes curved tracks) is not possible. Still, the pixel tracker provides information of the location of single hits in the different layers. Thus, straight lines can be reconstructed originating

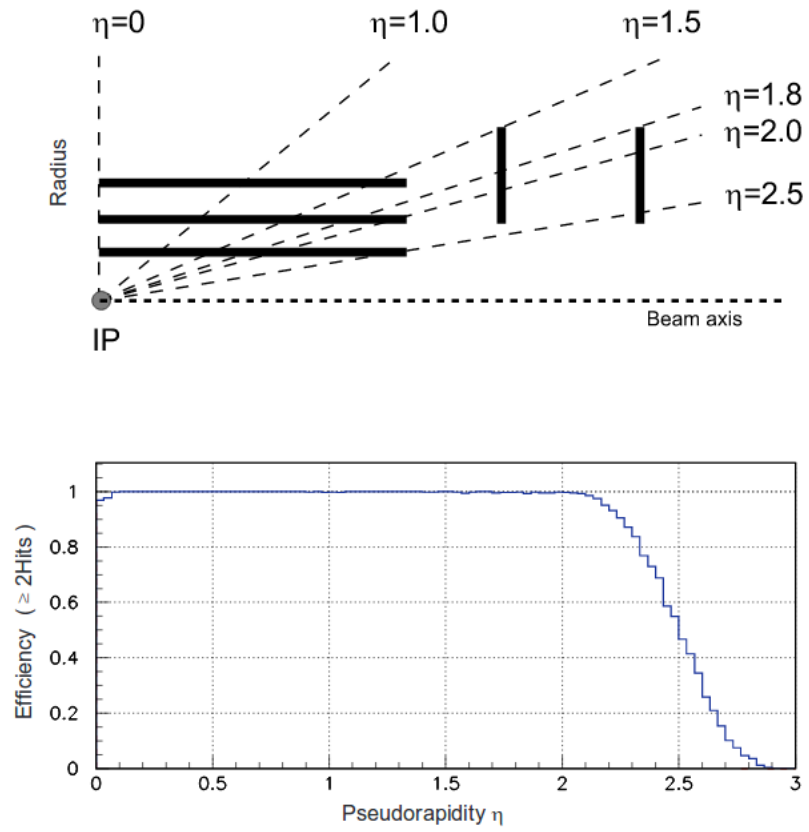



**Fig. 3.3.:** Geometrical structure of the CMS pixel detector (top) and hit coverage as function of pseudorapidity (bottom). Taken from Ref. [10].

from charged particles traversing the detector. A detailed description of this modified tracking method used for this thesis is given in chapter 6.2.1.

Extensive electromagnetic and hadronic calorimeters are also located within the volume of the solenoid magnet covering a pseudorapidity range of $|\eta| < 3.0$. The Electromagnetic Calorimeter (ECAL) consists of a barrel and endcap region. Highly transparent lead-tungstate crystals are used as simultaneous absorber and active material. The scintillation light produced in the crystals is detected using avalanche photodiods (barrel) and vacuum phototriodes (endcap). The total depth of the ECAL is 25.8 radiation lengths. The Hadronic Calorimeter (HCAL) with a depth of 5.8 – 10.6 hadronic interaction lengths is a sampling calorimeter using brass as absorber and plastic scintillator to detect the showering particles. The last shell of the onion-layered design is the muon tracking system outside the solenoid. It covers the pseudorapidity region $|\eta| < 2.4$ and consists of three detectors with different technologies: aluminum drift tubes (DT) in the barrel region, cathode strip chambers (CSC) in the endcap region, and restive plate chambers (RPCs) both in the barrel and endcap regions.

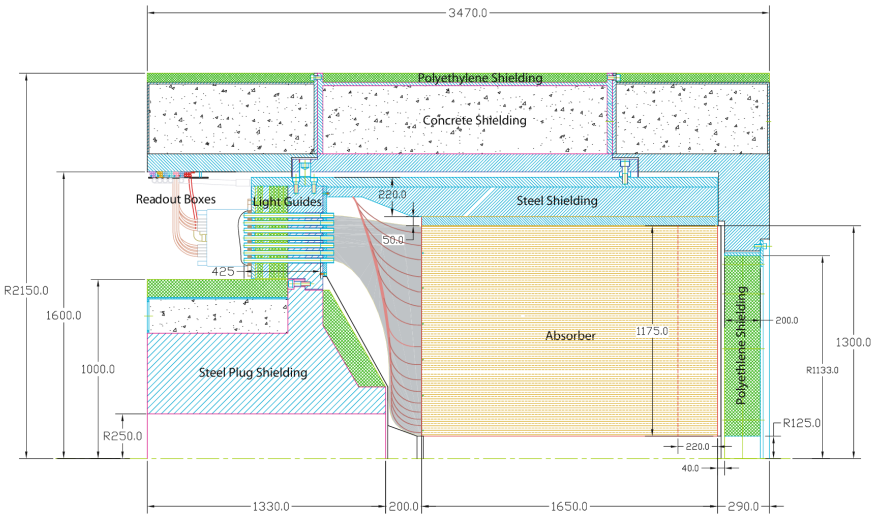



**Fig. 3.4.:** Schematic drawing of the Hadron Forward (HF) calorimeter of CMS. The active area is shown in orange color. The interaction point is at 11.15 m from the front of the calorimeter to the right. Taken from Ref. [97].

The Hadron Forward calorimeter (HF) extends the coverage of the CMS detector towards the forward region at $3 < |\eta| < 5.2$. It is composed of 5 mm thick grooved steel plates. Inside these grooves are fibers of quartz that collect Cherenkov light produced by charged particles from showers in the steel matrix. The fibers are bundled to form readout units (called towers) with a segmentation of $\Delta\phi \times \Delta\eta = 0.175 \times 0.175$. This results in 18 azimuthal segments and 13 segments in pseudorapidity on both of the two sides of CMS. The signals from the fibers are collected with air-cone light guides and propagated to photomultipliers. A cross section drawing of the HF calorimeters is shown in Fig. 3.4, more details can be found in Ref. [97]. The calibration of HF is performed with dedicated test beam measurements using beams of electrons and pions at fixed energy. This calibration is corrected for a slow degradation of the PMTs due to radiation damages. Also collision data is used with events where $Z$-boson decays to two electrons [98]. One of them is required to be reconstructed within the acceptance of the central ECAL in order to apply strong isolation criteria with ECAL and the tracker. The second electron is reconstructed in HF and the corresponding energy is given by

$$E_{\mathrm{HF}} = \frac{m_Z^2 \cosh \eta_{\mathrm{E}} \cosh \eta_{\mathrm{H}}}{2E_{\mathrm{E}} \left(\cosh\left(\eta_{\mathrm{E}} - \eta_{\mathrm{H}}\right) - \cos\left(\phi_{\mathrm{E}} - \phi_{\mathrm{H}}\right)\right)} \tag{3.2}$$

where $E_{\mathrm{E}}, \eta_{\mathrm{E}}, \phi_{\mathrm{E}}$ is the energy and position of the electron in the ECAL and $\eta_{\mathrm{H}}, \phi_{\mathrm{H}}$ the position of the electron candidate in HF. The associated total uncertainty on the energy calibration of HF is found to be 10%. On the negative z-side of CMS, the CASTOR calorimeter joins the HF calorimeter covering $-6.6 < \eta < -5.2$. A detailed description of CASTOR is given below.

The trigger system of CMS is in detail described in Ref. [99]. It contains hardware based triggers (called Level 1 triggers) and software based high-level triggers (HLT triggers). All those triggers select events according to characteristic collision properties. In the scope of this work, only the so-called *zero bias* trigger is used. This is an unbiased trigger that only requires the presence of proton bunches in the interaction region of CMS determined with Beam Pick-up Timing eXperiment (BPTX) devices. Two of those are located at a distance of about 175 m on both sides of the CMS interaction point (BPTX+ and BPTX-). The trigger selects bunch crossings with the digitized BPTX signal by requiring a coincidence of the signals from the detectors on both sides, BPTX+ and BPTX-. The influence of electronic noise and beam gas backgrounds is studied using dedicated triggers that require no beam presence (no BPTX signal) or the presence of only one of the two beams (only BPTX+ or only BPTX-).

## 3.2 The CASTOR calorimeter

The CASTOR calorimeter of CMS is a unique detector among all LHC experiments. Being located at about 14 m distance from the interaction point at the negative z-side of CMS, it covers a pseudorapidity region of $-6.6 < \eta < -5.2$. It is therefore the most forward calorimeter sensitive to both charged and neutral particles. A schematic drawing of the detector as well as a photo of its location within the CMS forward region are shown in Fig. 3.5.

**Detection principle**

CASTOR is a sampling calorimeter of alternating plates of tungsten and quartz. The incident collision particles hit the calorimeter approximately parallel to the beam line. They interact in the tungsten layers and develop particle showers. Charged shower particles traverse the quartz and emit Cherenkov light at the Cherenkov angle of about 45°. The light is guided within the quartz away from the beam pipe towards air-cone light guides that are connected to fine mesh photomultipliers (PMTs). This detection technique is more sensitive to the showers of electromagnetic primary particles, reducing the light yield of hadronic primary particles. This is a common effect in calorimetry and is called non-compensation. In order to optimize the amount of Cherenkov light being transported to the PMTs, the tungsten and quartz plates are tilted by an angle of 45° with respect to the beam direction. The PMTs are tilted by 30° in order to minimize the influence of stray magnetic fields on the PMT gains. The signals of five tungsten/quartz layers are combined onto one PMT and read out together. A sketch of the detection principle can be found in Fig. 3.6.

**Geometrical details**

In order to be able to be removed and re-installed for different data taking periods, CASTOR is composed of two separated half cylinders. When installed in ideal position,

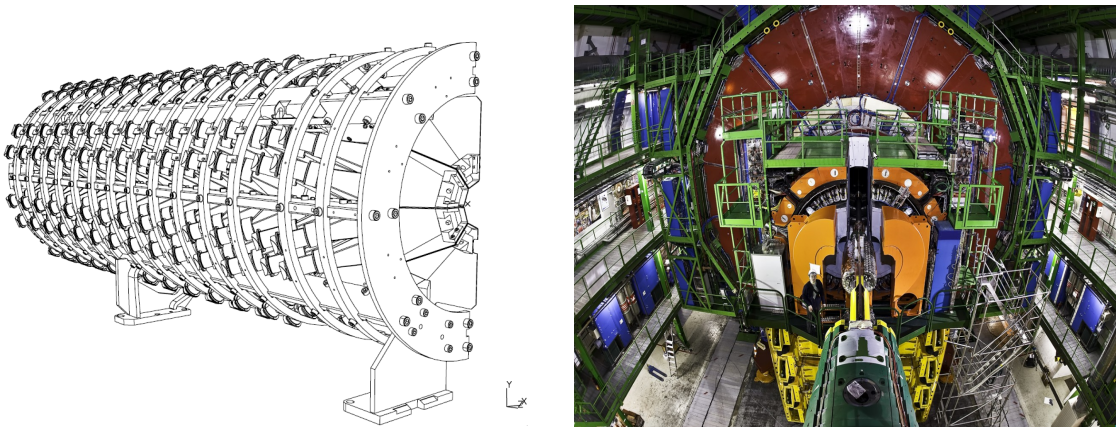


**Fig. 3.5.:** Schematic drawing of one half cylinder of CASTOR the CASTOR calorimeter (left) and a photograph of its location in the forward region of CMS (right) together with a member of the KIT-group, I. Katkov, next to CASTOR.

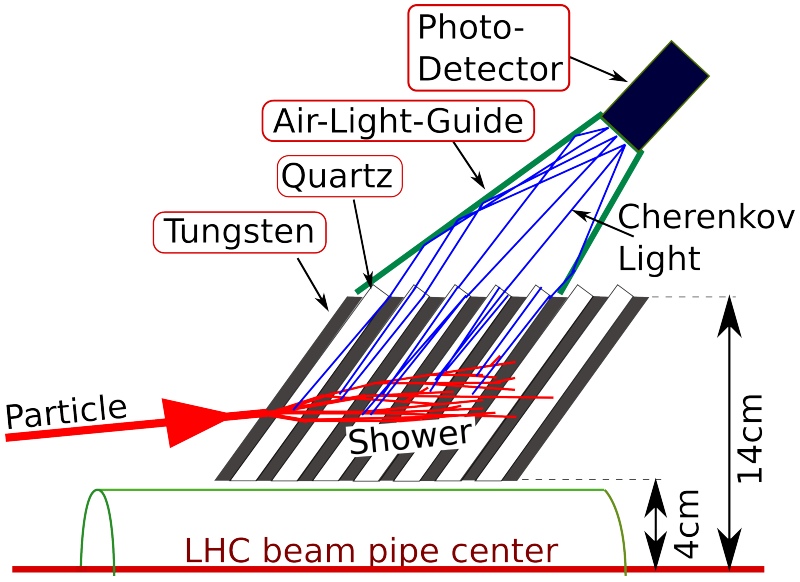



**Fig. 3.6.:** Detection principle of the CASTOR calorimeter. Incoming particles produce showers in the detector volume. The charged component of the shower produces Cherenkov light that is guided to photo multiplier tubes.

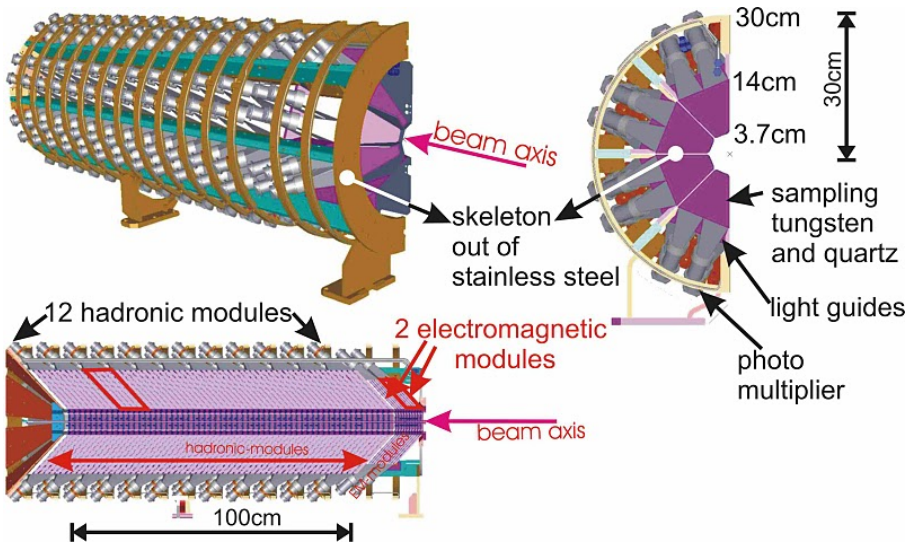



**Fig. 3.7.:** Drawing of the geometrical structure of one half cylinder of the CASTOR calorimeter.

both halves close and form a hollow cylinder with a inner radius of about 4 cm and a typical distance of about 1 cm of the inner detector edge to the LHC beam pipe. The outer radius of the calorimeter is approximately 30 cm. Each half is constructed with a stainless steel skeleton that supports the detector material as well as the PMTs. The half cylinder located at the side towards the center of the LHC ring is called the *near side*, the one on the other side the *far side*. A schematic drawing of the calorimeter is shown in Fig. 3.7. In the x-y-plane the tungsten and quartz plates form an octagon that is separated into 16 units, called sectors. In the longitudinal direction CASTOR consists of 14 segments (with 5 tungsten/quartz layers each), called modules. This adds up to a total of 224 readout channels. The two modules closest to the interaction point are half as thick (5 mm tungsten and 2 mm quartz) as the other 12 modules (10 and 4 mm thick layers). These two modules are mostly sensitive to purely electromagnetic showers initiated by electrons and photons, while hadronic showers extend further into the back of CASTOR. The first two modules are therefore called the electromagnetic section, while the rest is called the hadronic section. The total depth of the calorimeter adds up to 10 nuclear interaction lengths $\lambda_i$.

**Readout and electronics**

As mentioned, the Cherenkov light is collected in the quartz plates and guided to fine mesh PMTs of type Hamamatsu R7494. These have been chosen because of their radiation tolerance in the highly irradiated location where CASTOR is positioned. The charge collected in a PMT is integrated in time intervals of 50 ns by QIE cards [100]. Those consist of ASICs to integrate the charge and a non-linear Flash Analog-to-Digital Converter FADC with a dynamic range of $10^4$. These QIE cards are also used for the HF calorimeter. The digitized signals are transferred via optical fibers to the CMS service cavern where they are integrated in the global CMS trigger and data acquisition (DAQ) system.

**Reconstruction**

The integrated charges of two consecutive time intervals (also called time slices) in units of fC are combined for every readout channel. The timing of the readout is optimized in a way that the majority of the signal is located in the first time interval. If the digitization limit of the ADC is reached, a saturation correction is applied. Based on the signal of the second time slice, the saturated signal in the first time slice is corrected. The combined signal for one readout channel is called *RecHit*. All RecHits within one $\phi$-sector are furthermore added up and form one calorimeter tower.

**Calibration**

For every readout channel, an individual calibration coefficient to convert the charge in units of fC into an energy deposit in units of GeV is determined. These coefficients $C_{\mathrm{i}}$ are composed of different sources.

A relative intercalibration coefficient $C_{\mathrm{rel,i}}$ is determined for every readout channel separately. The goal is to equalize gains of the different PMTs with respect to a reference channel. For this reason special data is taken during times when there are beams in the LHC rings but while they are not yet brought into collisions. In this scenario the particles in the beam scatter with each other and sometimes also surrounding material and produce so-called beam halo muons. Those muons travel approximately parallel to the beam and can be considered minimal ionizing particles (MIPs). Thus, they traverse the whole length of the calorimeter only depositing small energies. These events are recorded with a special trigger. A dedicated event selection is applied and as a result the spectrum of charge deposits for these muon events is obtained for every individual readout channel. An example of this spectrum is shown in Fig. 3.8. The average charge deposit is then calculated and related to the reference channel [Module 4, Sector 9]. The intercalibration coefficient $C_{\mathrm{rel,i}}$ is given by

$$C_{\mathrm{rel,i}} = \frac{\langle Q_{\mathrm{ref}} \rangle}{\langle Q_{\mathrm{i}} \rangle} \,. \tag{3.3}$$

Since the muon signal is very close to the noise background, the PMTs are operated at the highest possible gain during circulating beam periods. During collisions, the high voltage is reduced to optimize the dynamic range of the readout electronics. The change of the signal gain behaves differently for every PMT and therefore, a correction factor has to be determined, that corrects the intercalibration factors as they were determined at high gain to the high voltage during collisions. These factors are determined with LED pulses that can be introduced externally into the PMTs. The signal strength of these pulses is evaluated at different high voltage settings and then parameterized. The correction $C_{\mathrm{HV}}$ is defined as the ratio of the gains $G$ at the high voltage used to record the muon signals $\mathrm{HV}_{\mathrm{muon}}$ and at the high voltage during physics collisions $\mathrm{HV}_{\mathrm{phys}}$:

$$C_{\mathrm{HV}} = \frac{G\left(\mathrm{HV}_{\mathrm{phys}}\right)}{G\left(\mathrm{HV}_{\mathrm{muon}}\right)} \,. \tag{3.4}$$

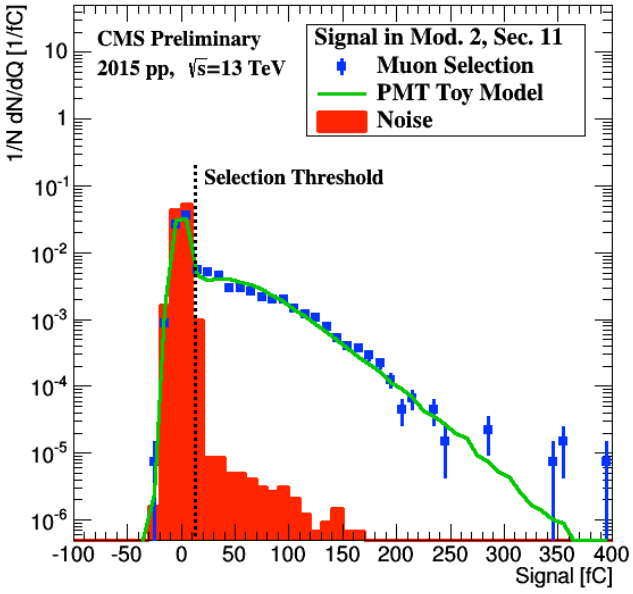



**Fig. 3.8.:** Charge spectrum for a typical CASTOR channel after an offline isolated muon event selection. The overlaid noise distribution is measured from non-colliding bunch data. The model line corresponds to a mesh-type PMT (CASTOR uses the Hamamatsu R5505 and R7494) with an average number of 0.5 photo electrons. The selection threshold used to identify channels above noise is shown as vertical line. [101]

The absolute energy calibration $C_{\mathrm{abs}}$ in units of GeV/fC was studied in a test beam with a full-length prototype of CASTOR. Beams of electrons and pions at fixed energy were used and the calorimeter response was studied [102]. It was found that the average charge per energy was 0.0150 fC/GeV for electrons and about half for pions. The measured energy resolutions are about 5% and 20% for electrons and pions, respectively. A second method to determine the energy scale was developed using proton-proton collision data at a center-of-mass energy of 7 TeV. The $\eta$-dependent average energy density $\mathrm{d}E/\mathrm{d}\eta$ was measured with the HF calorimeters in the region $3 < |\eta| < 5$ [103]. These data were extrapolated to the acceptance of CASTOR using a large set of different hadronic interaction models. This is shown in Fig. 3.9. The interaction models were normalized to the data and the average expected energy in CASTOR was found to be 665 GeV on hadron level and $\langle E_{\mathrm{exp}} \rangle = 336\,\mathrm{GeV}$ on detector level. Given the measured average signal charge in CASTOR of $\langle Q_{\mathrm{meas}} \rangle = 20960\,\mathrm{fC}$ a calibration coefficient is determined with

$$C_{\mathrm{abs}} = \frac{\langle E_{\mathrm{exp}} \rangle}{\langle Q_{\mathrm{meas}} \rangle} \,. \tag{3.5}$$

This factor is equivalent with a response to electrons of 0.0160 GeV/fC. This value is consistent with the test beam results within the systematic uncertainties. This data driven method to determine the energy scale was chosen for further data analysis and reconstruction because it allows to assign realistic and straightforward uncertainties.

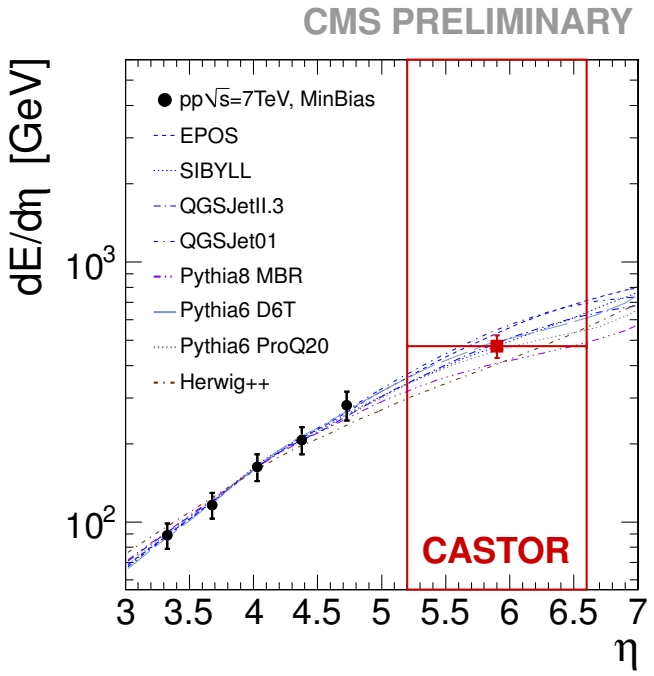



**Fig. 3.9.:** Illustration of the extrapolation method used to determine the CASTOR energy scale. Model predictions of the energy density are normalized to the HF data from 7 TeV pp collisions. The average extrapolated energy in the CASTOR acceptance is indicated. [95]

Systematic uncertainties arise from the absolute calibration of HF (which is known up to 10%), the model dependence of the extrapolation (10%), the implementation of the non-compensation in the detector simulation (5%). These contributions add up in quadrature to a total systematic uncertainty of 15%. The procedure is in more detail explained in Ref. [95].

Combining Eqs. (3.3)–(3.4) one obtains the individual calibration coefficients:

$$C_{\mathrm{i}} = C_{\mathrm{abs}} \times C_{\mathrm{rel}} \times C_{\mathrm{HV}} = \frac{\langle E_{\mathrm{exp}} \rangle}{\langle Q_{\mathrm{meas}} \rangle} \times \frac{\langle Q_{\mathrm{ref}} \rangle}{\langle Q_{\mathrm{i}} \rangle} \times \frac{G\left(\mathrm{HV}_{\mathrm{phys}}\right)}{G\left(\mathrm{HV}_{\mathrm{muon}}\right)} \,. \tag{3.6}$$

**Event Examples**

Figure 3.10 shows examples of three collision events in CASTOR. An event with large activity in the whole calorimeter and a total reconstructed energy of about 3.2 TeV is shown in the top panel, a single hadronic shower of about 1 TeV starting deep in the calorimeter is shown in middle, and an event with about 4 TeV mostly contained in the electromagnetic section in the bottom panel. The geometrical structure of CASTOR can be seen as well as energy deposits emphasized with red areas whose size is proportional to the energy in the corresponding readout channel. These events were recorded in June 2015 and are part of the dataset used for the physics analysis of this thesis.


**Fig. 3.10.:** Examples of three collision events recorded in CASTOR in June 2015. The energy deposited in every readout channel is shown with red area, proportional to the energy.

# 4 Improved alignment of CASTOR

> *The more precise the measurement of position, the more imprecise the measurement of momentum, and vice versa.*
>
> **— Werner Heisenberg**
> (Physicist, Nobel Laureate)

A precise knowledge of the detector position with respect to the global reference frame as well as the location of the collision is of great importance for every subdetector of CMS (see for example [104, 105]). This is especially true at very forward rapidities where the CASTOR calorimeter is located. One reason is that the average energy produced per proton-proton collision rises sharply with pseudorapidity in the acceptance of CASTOR (see Fig. 4.1). Changes in the position of CASTOR lead to a changed acceptance in pseudorapidity and therefore also result in significant changes in the expected average energy deposited per collision. Figure 4.2 illustrates the change of the average energy in CASTOR as a function of a shifted acceptance in pseudorapidity. The corresponding shift of CASTOR in the radial direction in units of cm is also given as a reference. It can be observed, that position changes of the order of 1 cm influence the average energy by about 10% for proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$. As described earlier, the calibration of CASTOR relies on the extrapolation of the average energy measured with the HF calorimeters at $3 < |\eta| < 5$. Uncertainties in the position measurement of CASTOR thus directly affect the absolute calibration and lead to significant uncertainties. During LHC Run 1, the contribution of the position uncertainty to the overall energy scale uncertainty was 16% [95] and therefore the largest single contribution. With a more precise knowledge of the calorimeter position this uncertainty can be significantly reduced. Since CASTOR is removed and installed again for special data taking periods, the position has to be evaluated for every installation period independently.

In order to study the effect of a given detector position on the calorimeter performance, the position needs to be implemented in Monte Carlo simulations. Theoretically, any orientation of the two CASTOR halves can be described with four parameters, a displacement in $x$ and $y$ of the global CMS coordinate system and two angular rotations $\phi$ and $\rho$, assuming a fixed position in $z$. In the ideal scenario all of them are equal to zero. The rotations $\phi$

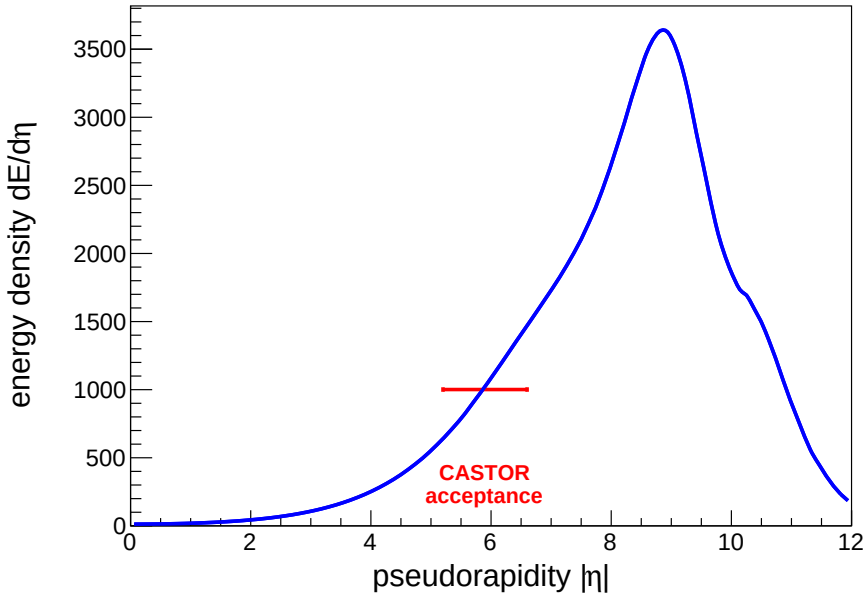



**Fig. 4.1.:** Average energy per unit of pseudorapidity simulated for proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with PYTHIA 8 tune CUETP8M1. The bulk of the available collision energy is produced in the direction of the incoming beam around $\eta \approx 9.5$. The acceptance of CASTOR (indicated by a red line) covers the rising part of this distribution.

and $\rho$ are currently not implemented in the description of CASTOR in the CMS detector simulation based on GEANT4. Studies in the past have shown that these rotations are small and their influence can be neglected. Therefore, the two half-cylinders of CASTOR have each two free coordinates $x$ and $y$ describing the position of their respective centers in the CMS reference frame.

Due to the delicate location close to the LHC beam pipe, the installation of CASTOR can not be reproduced precisely. Changes in the position between different installation periods are therefore not negligible. Another complication arises from magnetic forces on the supporting structures of CASTOR. While CASTOR itself mainly consists of non-magnetic materials the support structures are made of iron. It was observed that CASTOR moves several millimeters during the ramping cycles of the CMS solenoid. Once the magnetic field is stable, CASTOR also reaches a stable position. These movements have to be monitored and taken into account to reconstruct the final position during data taking conditions. To meet these circumstances, CASTOR is equipped with various sensor technologies to determine the position at all times. The different systems have different advantages and disadvantages and complement each other. A detailed description of the complete sensor system and its implementation is provided in Ref. [106]. Example pictures are shown in Fig. 4.3.

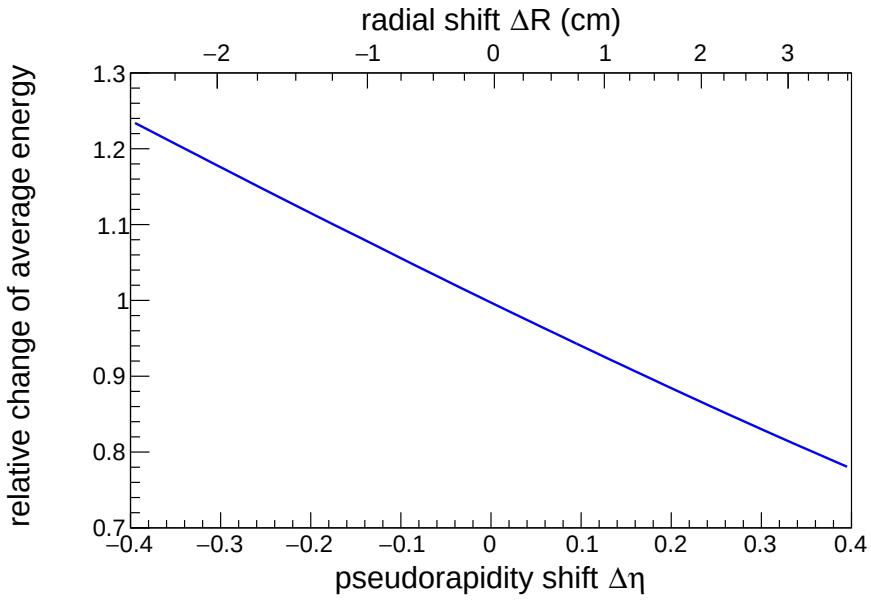



**Fig. 4.2.:** Change of the average energy deposited in CASTOR as a function of a changed acceptance, simulated for proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with PYTHIA 8 tune CUETP8M1. A displacement of CASTOR in the radial direction relative to the LHC beam pipe (upper $x$-axis) is equivalent to a changed acceptance in pseudorapidity and leads to significant changes in the expected average energy deposit.

**Reflective laser targets**

Both of the CASTOR half cylinders are equipped with three reflective targets each. After the installation of CASTOR the position of these targets can be determined at 1.0 mm precision with a special laser system. laser beams are shot onto the targets and their position can be determined from the time delay of the reflected light. The obtained position is directly given within the global reference frame of CMS. A software tool to extract the CASTOR position from this measurement was developed in Ref. [91]. The fit to the laser data provides the most accurate position measurement available. Since CASTOR is hidden behind solid iron shielding, the measurement has to be performed before the magnetic field of CMS is turned on. Possible movements during the magnet rampup can therefore not be taken into account. Although the CMS solenoid was not turned on for the low luminosity data taking campaign in 2015, the magnetic field was once tuned on and off after the installation of CASTOR. It can not be assumed that CASTOR recovers the original position after the magnet cycle.

**Contact potentiometers between CASTOR and support structures**

Overall, seven contact potentiometer sensors of type *Linear Potentiometer CLS1313* by the manufacturer *Active Sensors* are installed on CASTOR [107]. The distance between the sensor and the contact surface is determined with a metal bar being pushed into the housing and this way changing the internal resistance. Their measurement is very accurate at 0.1 mm precision [106]. Two of these potentiometers measure the opening between the

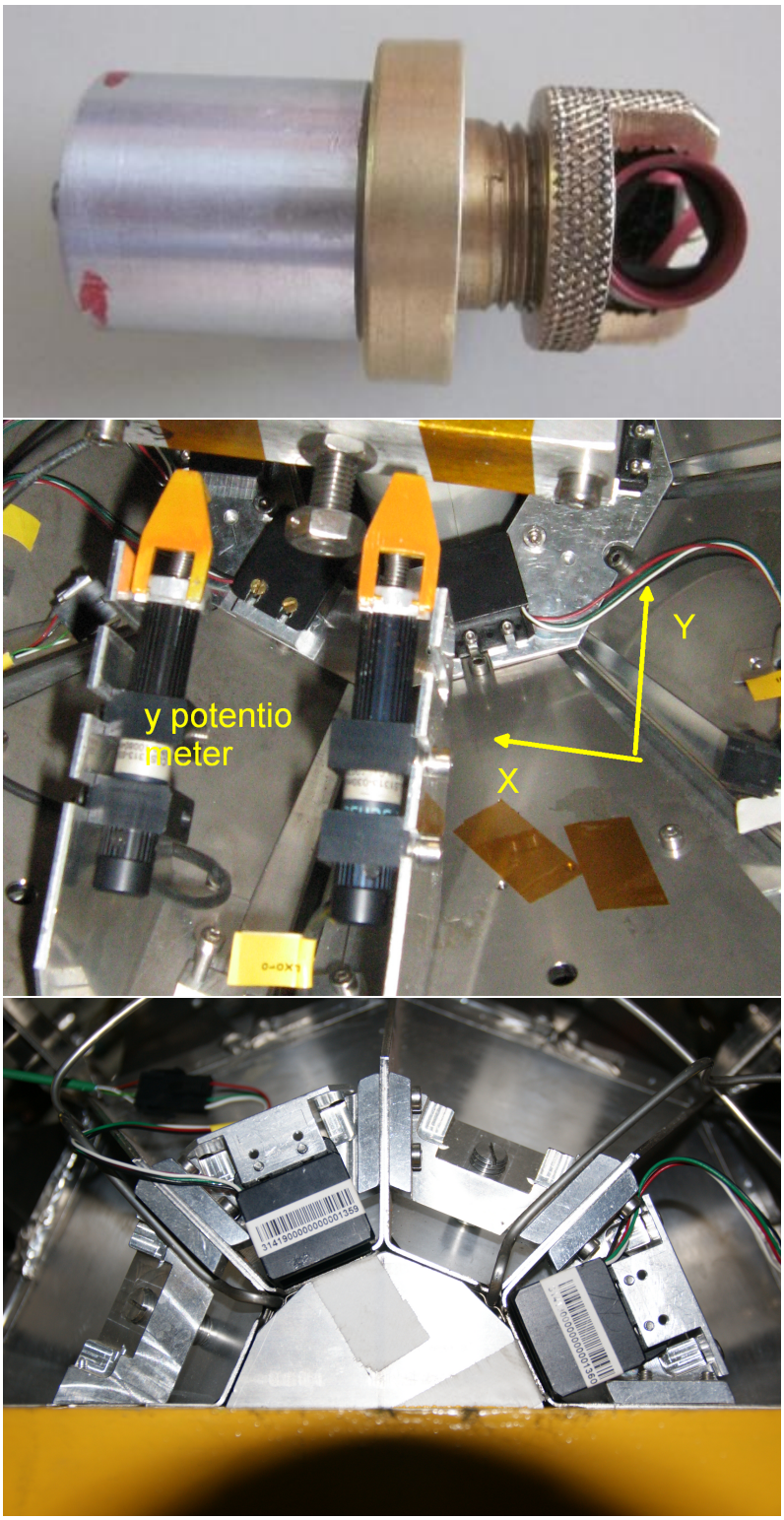



**Fig. 4.3.:** Example photographs of the devices used for the three alignment methods. Top: cylindrical adapter with a red reflective area used as target for the laser system. Center: two contact potentiometers installed on the non-IP side of CASTOR, measuring the distance to the beam pipe support structures in the $y$-direction. Bottom: two infrared distance sensor installed on the IP side of the near half of CASTOR.

two halves of CASTOR on the side pointing towards the interaction point (IP side). Five potentiometers are mounted on the side pointing away from the interaction point (non-IP side). They measure the distance to a beam pipe support structure in the $x$ and $y$ direction on both halves and in the $z$ direction on one CASTOR half only. Unfortunately, neither their exact mounting on CASTOR nor the exact position of the target surfaces are known. Therefore, they only provide relative information on movements due to magnetic forces.

**Infrared sensors between CASTOR and the LHC beam pipe**
A set of ten infrared sensors is mounted on CASTOR to monitor the distance of the inner detector boundary to the LHC beam pipe. These sensors were initially designed for safety reasons to monitor whether the detector moves too close to the fragile beam pipe in the center. Since these sensors can measure the distance to the beam pipe even when the magnetic field is turned on, their information is also used to determine the final position of CASTOR relative to the beam pipe. Their measurements are, with an accuracy of 1 mm, less precise compared to the other technologies and also the conversion of the independent sensor readings to a global position of CASTOR is less straightforward. Still, they provide valuable information. Unfortunately, the infrared sensors are not radiation hard. After few days, the stray radiation from the LHC beams significantly reduce their performance. A new set of sensors therefore has to be used for every installation.

Next to these, also alignment methods using collision data have been performed, see for example [108], and various efforts have been made to combine the information from the different methods into a global combination in order to get the most precise information possible [91]. Is was found that, with a more precise calibration, the information provided by the infrared distance sensors could improve the overall position measurement of CASTOR. The standard calibration of these sensors is not sufficient since it neglects some of the special circumstances when operated on CASTOR. Therefore, a new calibration procedure was developed as part of this thesis in order to improve the reliability of their measurements. Subsequently, the contribution of the position uncertainty to the overall energy scale uncertainty of CASTOR was significantly reduced from 16% to 7.5% for the data taking period in June 2015.

## 4.1 Two-dimensional calibration of infrared distance sensors

The infrared distance sensors mounted on CASTOR are *Long Range Retro-reflective Sensors* VTR24F1H originally by PerkinElmer [109], now produced by EXCELITAS. According to the manufacturer's specifications, the sensors are built to measure distances of up to 10 cm. They consist of a light emitting diode (LED) in the infrared range and a photodarlington transistor. A schematic view of the sensor is shown in Fig. 4.4. The sensors have four supply cables: the red and black wires serve as the power supply for the infrared LED and the green and white cable provide the bias collector-emitter voltage for the photodarlington

transistor. The electronics sit inside a polycarbonate housing which is small and can easily be mounted on CASTOR. Photographs of an individual sensor are shown in Fig. 4.5.

The sensors use photodarlington transistors to detect the intensity of the reflected infrared light emitted by the sensor's LED. Being reflected off a distant target surface, it reaches the photodarlington transistor and creates electron-hole pairs in the base–collector junction of the transistor. The so created base current is amplified by the transistor and the corresponding collector-emitted current is fed directly into a second transistor, further amplifying the current. Since both transistors share a common collector, this *darlington pair* achieves a large current gain on small space. The output current of the transistor thus changes depending on the amount of reflected light. Besides the distance between emitter and the target surface, this also depends on the properties of the target surface itself, namely its reflectivity as well as the orientation of the target surface with respect to the incident light. This is especially important for the application on CASTOR because the target surface is not an even plane but the cylindrical LHC beam pipe. While the reflectivity of the beam pipe surface is controlled with a special reflective tape, the effect of the cylindrical geometry of the target was so far neglected.

The calibration translating the output current into a measured distance was in the past determined with a plane target perpendicular to the sensor distance [106]. This corresponds to case a) in Fig. 4.6. This calibration is expected to be precise enough for a cylindrical target if the incident light hits the target perpendicularly (case b in Fig. 4.6), since the illuminated area on the target is small with respect to the curvature. If the sensor is significantly displaced from the target center (case c), the angle at which the infrared light hits the target is significantly changed. Such a displacement is expected to reduce the amount of light reflected towards the sensor. The installation of CASTOR can not be done precisely enough to ensure that the distance sensors point perpendicularly towards the beam pipe. Moreover, the movements during the magnet ramp also induce further shifts of the sensors with respect to the beam pipe. Therefore, case c) in Fig. 4.6 is a realistic scenario during the operations of CASTOR. The effect of this special geometry also depends on other parameters such as the orientation of the infrared LED with respect to the darlington transistors as well as the opening angle of the light emission towards the target. All these parameters are unknown and might differ for every individual sensor. Therefore, the exact influence of the geometry on the measured distance can only be quantified and eventually be corrected for with dedicated calibration measurements for every individual sensor. The sensor response needs to be measured precisely depending on the exact position and orientation with respect to the target. A two-dimensional calibration method was developed in the frame of this work to account for these needs. The circumstances in which the sensors are mounted on CASTOR were accurately reproduced and every sensor was individually calibrated in order to account for sensor-to-sensor variations.

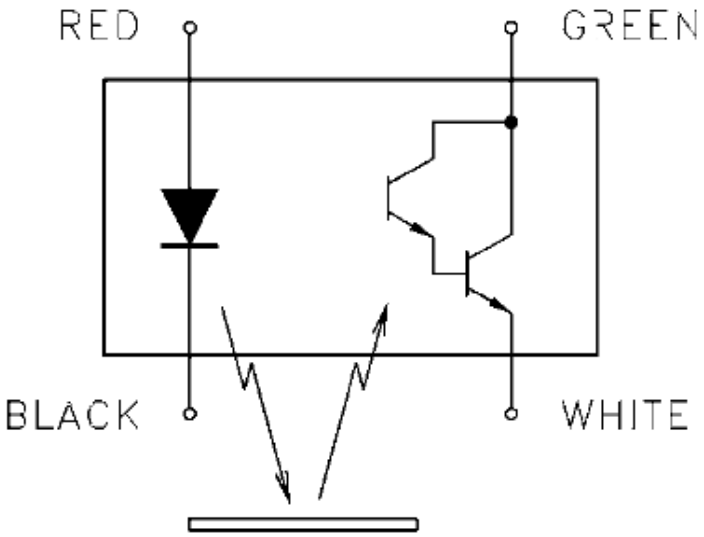



**Fig. 4.4.:** Schematic view of the working principle and internal electronics of the infrared distance sensors. The infrared LED (left) is supplied by red and black wires, the photodarlington transistors (right) via green and white cables. Taken from Ref. [109].

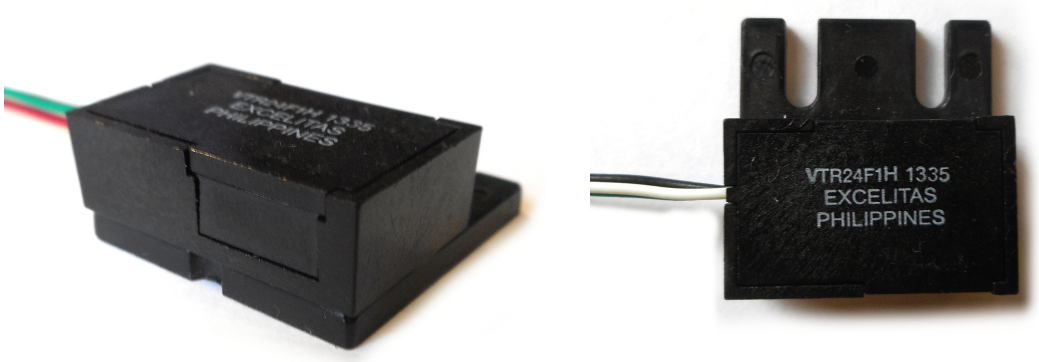



**Fig. 4.5.:** Photographs of an infrared distance sensor. The left photo shows the view from the front, the power supply and readout cables as well as the window for the infrared light can be identified. The right photo shows the sensor from the top.

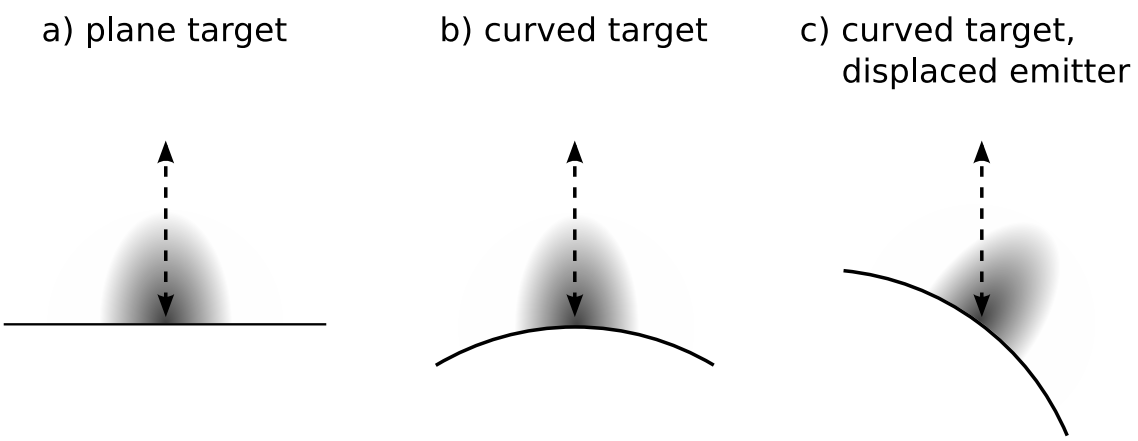



**Fig. 4.6.:** Illustration of the influence of the emitter-target geometry on the reflected light. The direction of the incident and reflected light from and towards the sensor is illustrated with dashed arrows, the target with a solid line. The reflection of the incident light is visualized with gray shaded areas assuming a diffuse but not isotropic reflection. Three geometries are shown: a) a plane target perpendicular to the direction of the incident light, b) the same for a cylindrical target, and c) for a cylindrical target and a displaced emitter.

### 4.1.1 Calibration setup

For the refined calibration of the infrared sensors, the situation in the experimental cavern of CMS is accurately rebuild in the laboratory. Special attention is paid to reproduce the geometric setup. A polycarbonate cylinder of the same diameter as the LHC beam pipe (57 mm) is mounted on a stainless steel support plate. It is further surrounded by the same reflective tape that is put around the LHC beam pipe (*PRO POWER 3130 WHITE*). An automated positioning systems with two linear stages (OWIS LT-80-150) is installed in front of the beam pipe mock-up. Each linear stage has a range of 15 cm. They are mounted perpendicular on top of each other. The top axis is additionally equipped with a mounting structure for the infrared sensors. This way, a sensor can be mounted and automatically be positioned in front of the cylindrical target at any given position with a precision of 100 $\mu$m. The positioning system defines a coordinate system by the orientation of the two linear stages: the $y$ axis is defined in the radial direction from the beam pipe center towards the sensor mounting, $x$ is the translation perpendicular to $y$. Figure 4.7 depicts a schematic view of the geometry. Figures 4.8 and 4.9 show some photos of this setup. During measurement, the whole setup is covered with black tissue in order to avoid stray background light influencing the measurement.

The electrical supply of the sensors is also of significant importance. The intensity of the infrared light emitted by the LED depends on the supply current and the sensitivity of the photodarlington transistor depends on the bias voltage. The power supply in the experimental cavern of CMS is depicted in Fig. 4.10. In order to provide constant light emission, a current regulator is added to provide a constant supply current of 3.5 mA to the

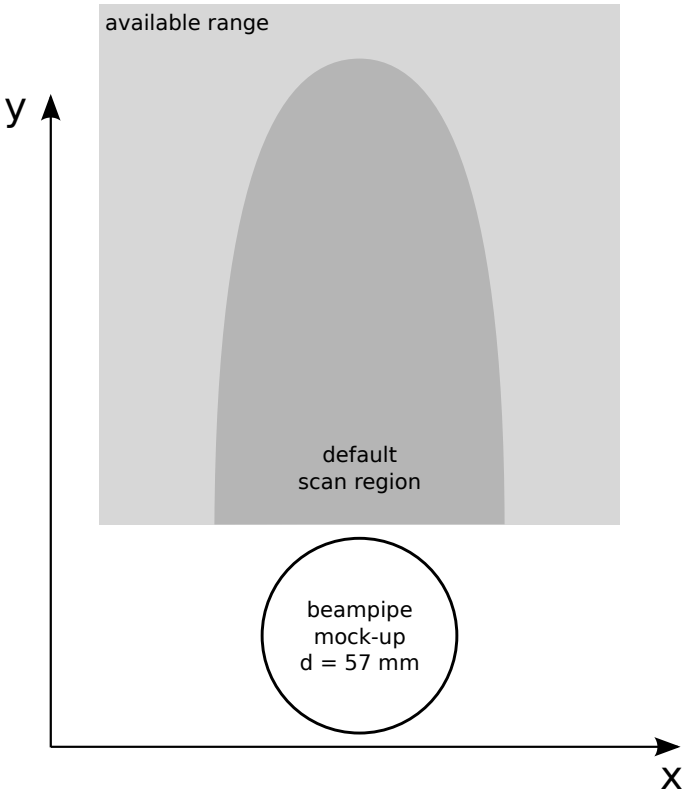



**Fig. 4.7.:** Schematic view of the realization of the two-dimensional sensor calibration. The beam pipe mock-up is shown with a black circle and the area in which the sensors are positioned with gray areas. The coordinate system used for later analysis is also indicated.

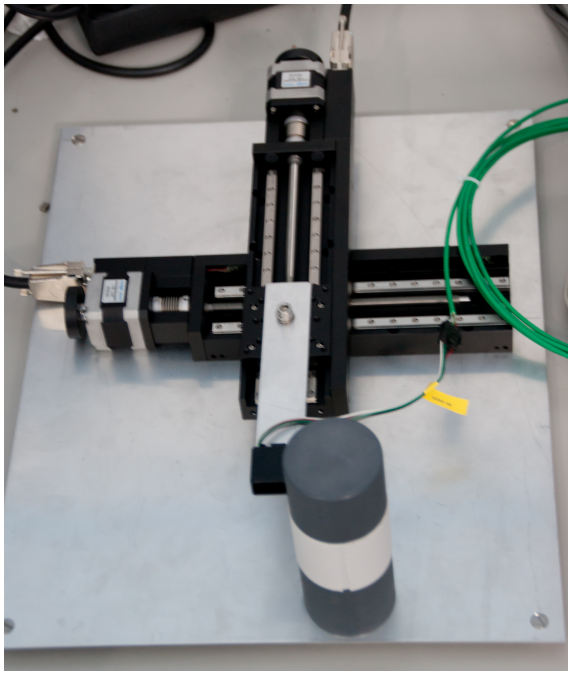


**Fig. 4.8.:** Overview of the calibration setup for the infrared distance sensors. The two linear stages are mounted perpendicularly on a stainless steel plate in order to position the infrared sensor in front of a beam pipe mock-up.

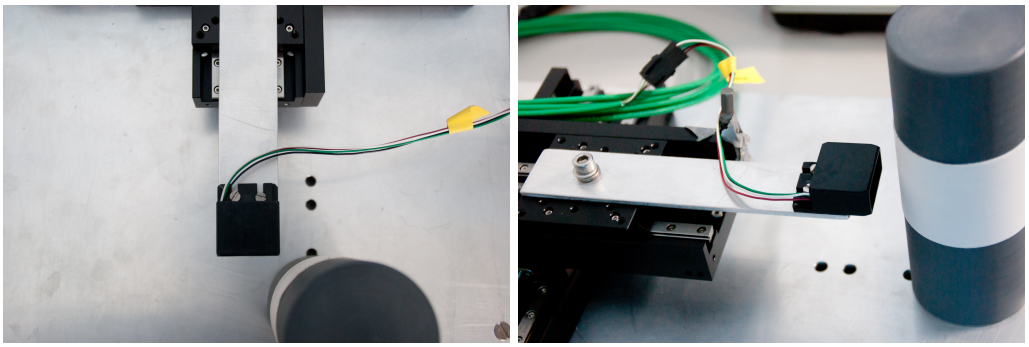


**Fig. 4.9.:** Detailed view of an infrared sensor being positioned in front of the cylindrical target. The target itself is covered with the same white tape as used in the CMS experimental cavern.

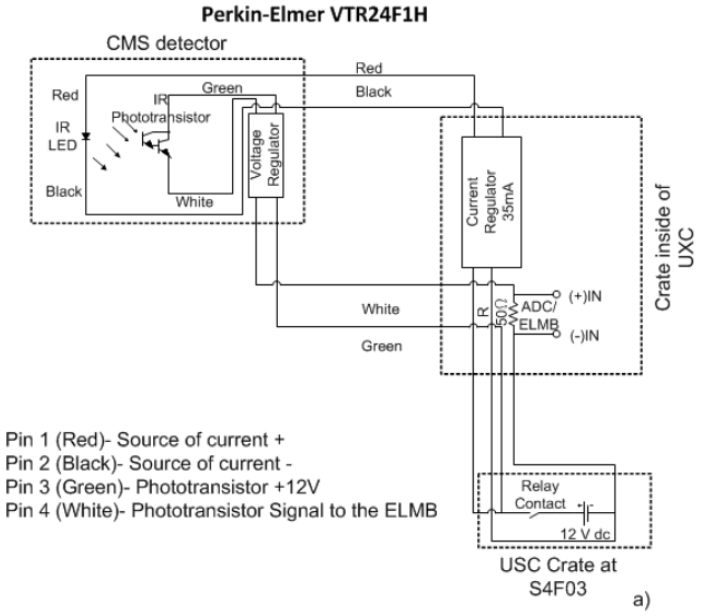



**Fig. 4.10.:** Schematic view of the supply and readout electronics for the infrared distance sensors as it is used in the experimental cavern of CMS. Taken from Ref. [106].

infrared LED. For the phototransistor, a constant operating point of 3.33 V is chosen [106]. In this setup, the collector-emitter current drawn by the darlington transistors is read out as a voltage dropping on a 50 Ω resistor. The signal is converted by an analog to digital converter (ADC) that is connected to the central CMS monitoring system. The laboratory calibration setup is realized in a similar but still more flexible way, allowing detailed systematic checks. Laboratory power supply units (*Agilent E3612A* and *TTi EX2020R*) are used to provide constant current for the infrared LED as well as a constant bias voltage to the phototransistor. With this setup, it is also possible to change the supply current and voltages in order to study systematic effects caused by possible instabilities of the power supply. Two digital multimeters are added in order to crosscheck the current and voltage supplies. A custom made connector box was built to provide an easy and straightforward access. The output current from the transistor is read out at a 100 Ω resistor. The readout is performed with an electrical multimeter *Voltcraft VC840* that is connected to a standard computer using a serial interface. Photographs of the complete setup are shown in Fig. 4.11.

A fully automated calibration procedure has been developed in order to achieve high quality in the obtained data as well as to save working time. The manufacturer of the linear stages provided a python based code package. Based on this, a software tool was written to provide a user friendly interface. It provides functionality to take user specific as well as completely automated measurements of the sensors response. The key features of this software shall briefly be explained:

**Fig. 4.11.:** Photo of the complete calibration setup (left panel) including the readout computer and the power supply units. The positioning system with the sensor and beam pipe mock-up are covered in black tissue. The right panel shows a detailed photo of the power supply units.

- At the program start, all hardware is initialized and the linear stages are automatically calibrated. For this, each axis moves to the maximum and minimum edges of their ranges. The minimum range is internally set to be the origin of the coordinate system.
- The sensor can automatically be moved to a starting position centrally in front of the beam pipe mock-up.
- A manual mode allows to adjust the starting position. Both axis can be moved in relative steps of 0.1 mm.
- The program performs an automated scan of the sensor response depending on the sensor position. The range of the measurement can be defined by the user in both $x$ and $y$ coordinates as well as the step sizes $\Delta x$ and $\Delta y$. The measurement time scales with the number of positions at which the sensor response is supposed to be measured. A default set of parameters is available through the user interface. An elliptic scan area that covers the whole sensor acceptance and a grid size of 1 mm is chosen as an ideal compromise between a precise measurement and an acceptable run time. With this parameter choice, a complete scan of one single sensor is performed in about eight hours.

During the automated sensor scan, the output current of the infrared sensor is read out multiple times in intervals of 0.2 seconds to account for fluctuations or a slow adaption of the sensor response to the new position. Once the measurement is stable on a level of five percent, five values are stored along with the sensor position in the coordinates of the positioning system as described above. The sensor is then moved to the next position. An example of such a measurement is show in Fig. 4.12.

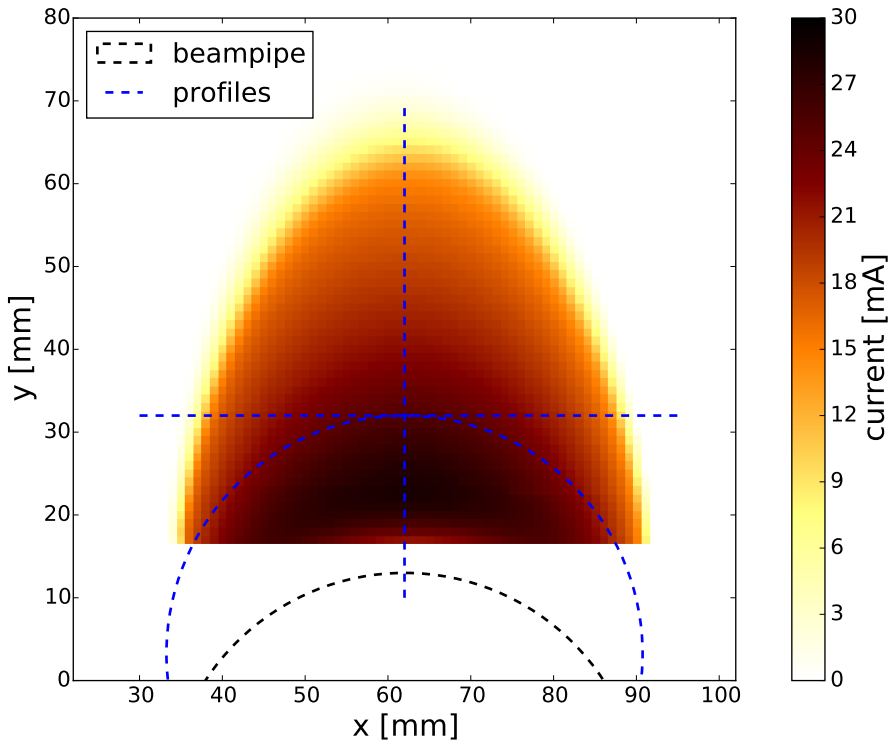



**Fig. 4.12.:** Example data of a complete calibration measurement of sensor *IR788*. The sensor response is measured in terms of the phototransistor output current as a function of the sensor position in front of the beam pipe mock-up. The position is given in the coordinate system of the positioning system. Profiles of the response along the dashed blue lines are shown in Figs. 4.13–4.15.

### 4.1.2 Results

The two-dimensional responses of 35 sensors have been recorded with the calibration setup described above. The full list is given in appendix A.1. Detailed studies of how the sensors behave at different geometries as well as under different power supply conditions are possible with this dataset. In order to study the geometrical effects of the curved target surface, profiles of the sensor response are evaluated as shown with blue dashed lines in Fig. 4.12. The studies of the sensor responses are repeated for every individual sensor leading to a precise individual calibration. In the following, data from one sensor labeled *IR788* are shown as examples.

Along a fixed value of $y$, the output current changes as the sensor passes the cylindrical target. This is shown in Fig. 4.13 for $y = 33\,\text{mm}$. As soon as the target moves into the line of sight of the sensor, the output current starts to increase, reaches a maximum when the phototransistor is at the center of the target and then decreases thereafter. It can be observed that the response is very symmetrical.

At a fixed value of $x = 63\,\text{mm}$, where the maximum along $x$ is reached, the sensor is located centrally in front of the target. Here, the curvature is minimal and the sensor response is similar to that of a plane target. The output current of the sensor as a function of $y$ shows the typical behavior also reported by the manufacturer (see Fig. 4.14). At small distances to the target, the output current slowly rises to a maximum and then falls slowly. At the end of the sensitive range, the current then drops quickly. The sensor reading therefore is ambiguous for very small distances. This has to be considered for the operation on the detector. Therefore, when mounted on CASTOR, the sensors are installed with an additional distance of about 15 mm to the detector edge.

The most relevant effect of the cylindrical beam pipe target is found when the response is investigated along a curve of constant distance to the target. This is shown in Fig. 4.15. Here the sensor response as a function of the coordinate $x$ is shown, where the considered sensor positions all have the same distance to the target along the sensor line of sight. This curve is represented by a circle with the beam pipe radius, displaced in $y$. As expected, the output current is reduced if the sensor is moved away from the central point. Due to the curvature of the target surface, less light is reflected back towards the sensors. This effect can reach a difference in output current of up to 5 mA, which corresponds to up to 10 mm in distance. In order to achieve an accurate distance measurement, this effect has to be considered and corrected for.

As mentioned earlier, another goal of these detailed measurements was to study systematic effects coming from the variation of the LED supply current and transistor bias voltage. Figures 4.16 and 4.17 show the variation of the sensor response as a function of $y$ along the central axis of the beam pipe target (similar to Fig. 4.14) when either the LED supply

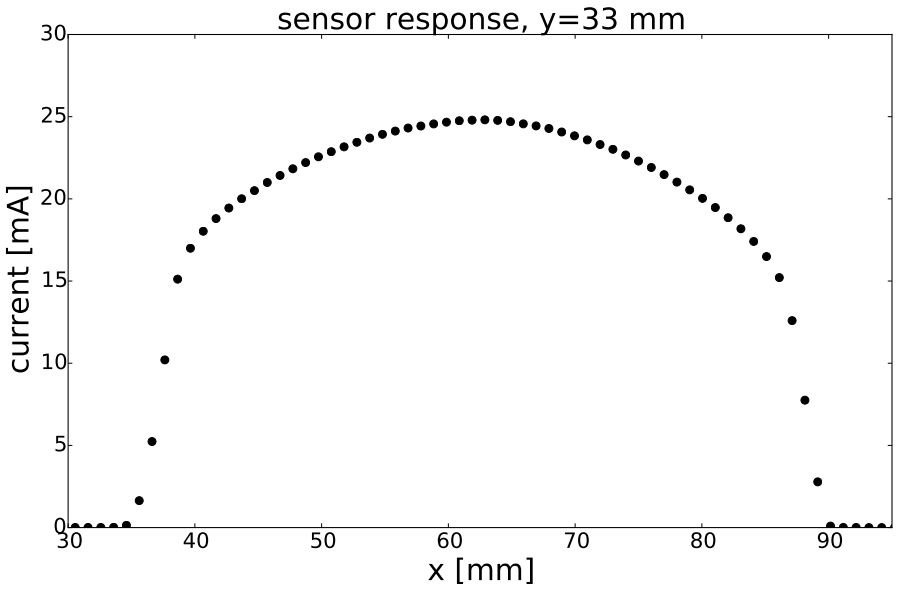



**Fig. 4.13.:** Response profile of *IR788* as a function of the position in $x$, along a line at fixed value of $y = 33\,\text{mm}$.

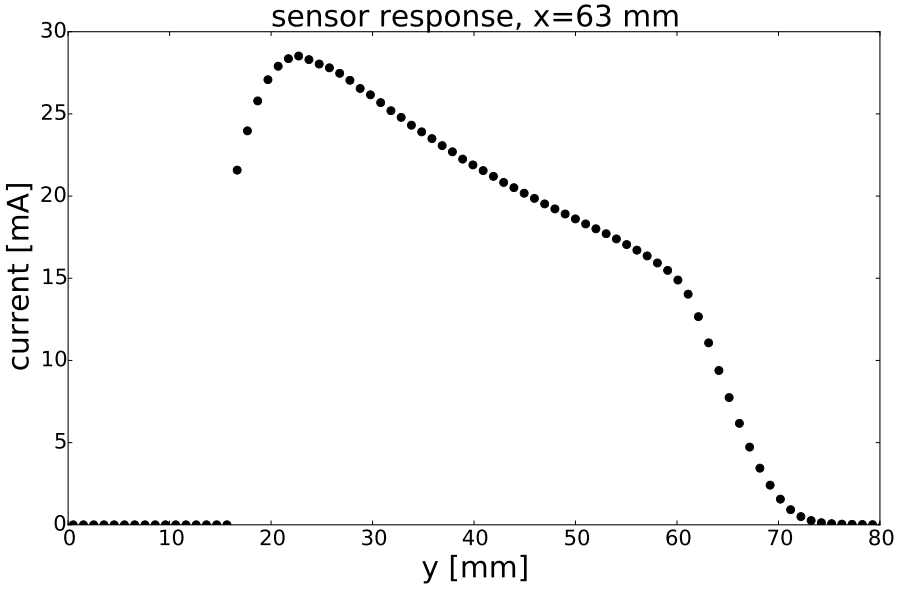



**Fig. 4.14.:** Response profile of *IR788* as a function of the position in $y$, along a line at fixed value of $x = 63\,\text{mm}$. The beam pipe surface is at $y = 14\,\text{mm}$.

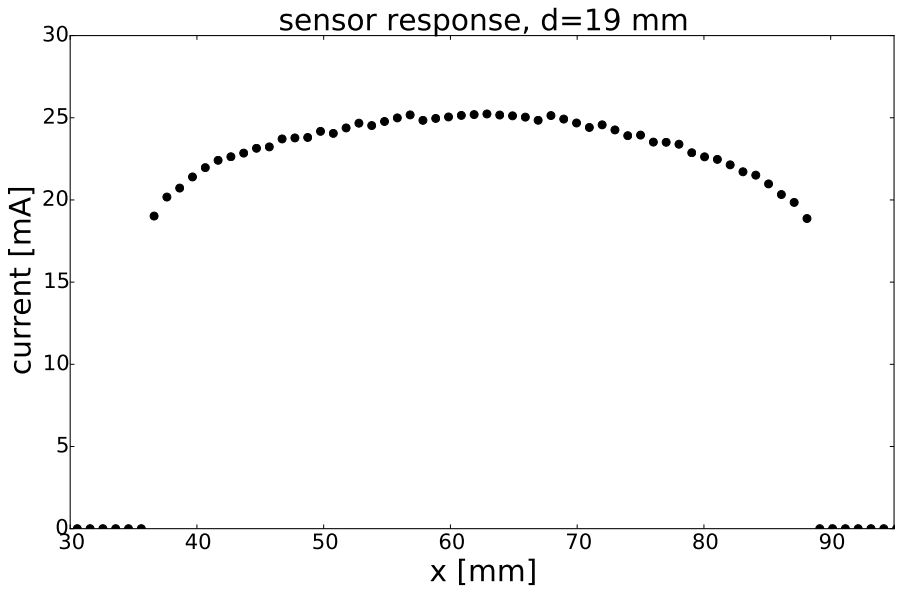



**Fig. 4.15.:** Response profile of *IR788* as a function of the position in $x$, along a line of fixed distance to the beam pipe mock-up of $d = 19\,\text{mm}$.

current or the transistor bias voltage are changed. If the LED supply current in increased (decreased), the light intensity increases (decreases) as well. This leads at a fixed distance to a larger (smaller) output current delivered by the phototransistor. Furthermore the maximal range of the sensor is increased (decreased). Changes of the transistor bias voltage change the sensitivity of the phototransistor to the reflected light. While the effect on the output current at a fixed distance is similar to the changes induced by different LED supply current, the influence on the maximal sensor range is small. The power supply in the CMS experimental cavern has been verified and found to be very stable. Correction due to a varying power supply are therefore, for the time being, neglected. Still, the available data provides this possibility in the future.

In the online monitoring system used by CMS, an immediate conversion from the sensor output current to a physical distance is performed. The main application of this system is to ensure that safety distances between CASTOR and the LHC beam pipe are kept during the installation of CASTOR and the rampup of the CMS solenoid. Within the monitoring tools, this is achieved with a ninth order polynomial conversion function:

$$d(I) = \sum_{\mathrm{i}=0}^{9} p_{\mathrm{i}} \cdot I^{\mathrm{i}} \,. \tag{4.1}$$

This particular choice of parameterization is of historic origin and a change would largely impact the CMS online monitoring system. A change is therefore beyond the scope of this work.

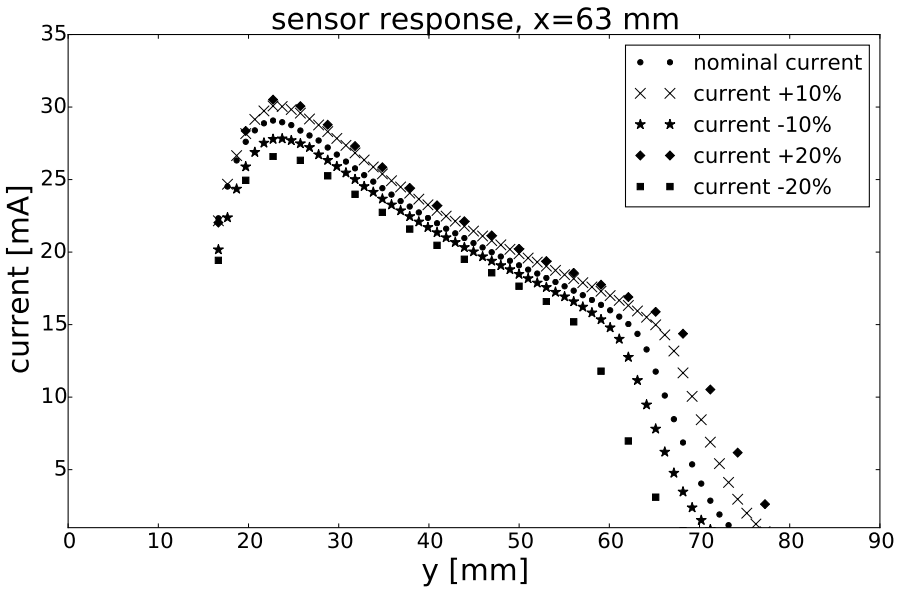



**Fig. 4.16.:** Response profile of *IR788* as a function of the position in $y$, along a line at fixed value of $x = 63\,\text{mm}$. The LED supply current is varied by 10 and 20% from the nominal value.

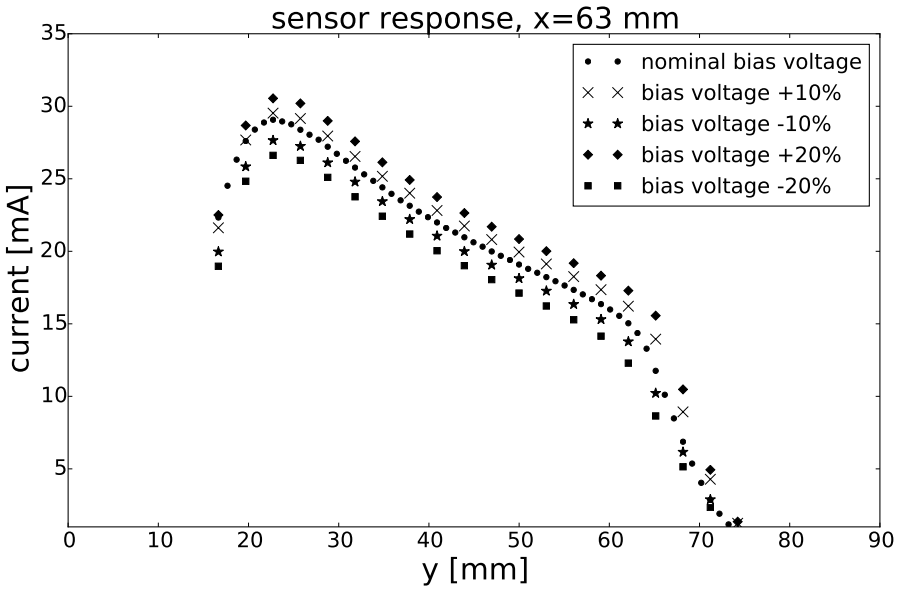



**Fig. 4.17.:** Response profile of *IR788* as a function of the position in $y$, along a line at fixed value of $x = 63\,\text{mm}$. The phototransistor bias voltage is varied by 10 and 20% from the nominal value.

Since a possible displacement of the sensors away from the central line of sight is not known a priori, an immediate correction for the geometric effects is not possible. This has to be done offline during a more detailed reconstruction of the positions (see next section for details). Nevertheless the safety requirements are still met, since the information of a large number of sensors can be combined to obtain the needed information. The $y$-dependent response curves as shown in Fig. 4.14 are used to determine these polynomial functions for every sensor individually.

Figure 4.18 shows the distance of the sensor to the beam pipe as a function of the output current. The measured position in the laboratory reference frame $y$ is first corrected to the actual distance between the sensor and the target cylinder. For this, the initial conditions that have been stored during the calibration measurement are used. Due to the sensor mounting and safety distance to the beam pipe mock-up the minimal distance between the sensor surface and the target are 4 mm. The ambiguous data points have to be removed in order to obtain a reasonable fit. Only values after the observed maximum of the response are considered. This leads to a limitation, that distances below about 10 mm can not be resolved and are avoided by installing the sensors about 15 mm inside of the physical detector edge. The resulting data is fitted with a ninth order polynomial and the parameters are stored inside the CMS online monitoring system. Figure 4.19 shows the obtained calibration curves for the four infrared sensors mounted on the front side of CASTOR facing the interaction point during the 2016 data taking period. These have in addition already been corrected for the offset of the sensors to the edge of CASTOR. While the shape of the obtained curves is very similar, significant differences in the absolute response can be observed. This emphasizes the need for individual calibrations although some of the sensors have very similar response functions.

### 4.1.3 Geometric correction during the position fit

A precise offline correction based on the detailed calibration data is applied during the fit of the actual CASTOR position to the measured sensor values. The fit algorithm was originally developed by C. Baus in Ref. [91] and extended in this work. During this procedure, both halves of CASTOR are moved independently in the $x$-$y$ plane of the CMS coordinate system and for every infrared sensor a $\chi^2$-value between the measured and calculated distances are calculated:

$$\chi^2 = \frac{d_{\mathrm{meas}} - d_{\mathrm{calc}}\,(x,y)}{\sigma_{\mathrm{meas}}} \tag{4.2}$$

with a systematic uncertainty on the sensor measurement of $\sigma_{\mathrm{meas}} = 1\,\mathrm{mm}$. Finally, the average $\chi^2$-value of all sensors $\chi^2/N_{\mathrm{sensors}}$ is calculated and minimized as a function of the ($x$,$y$)-position of the two CASTOR halves.

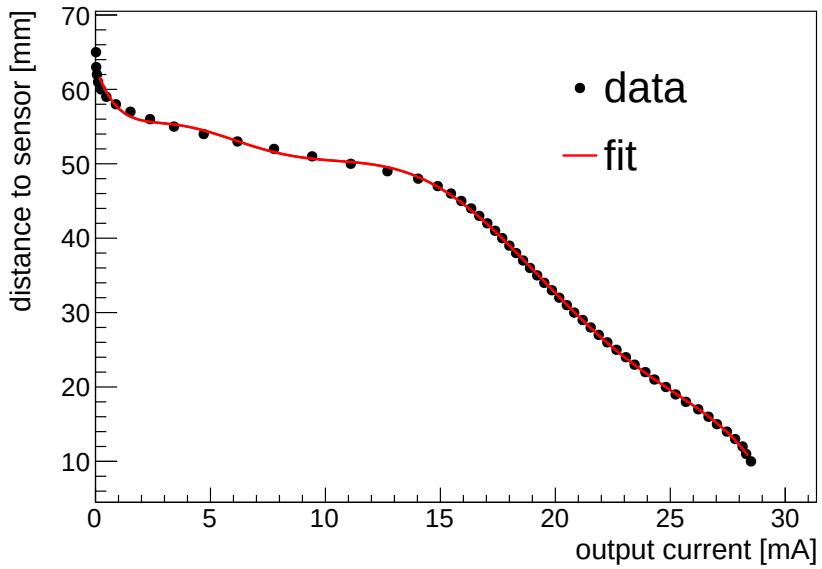



**Fig. 4.18.:** The data shown in Fig. 4.14 is used to obtain the calibration of sensor *IR788*. The distance of the sensor to the target is shown as a function of the output current. The data from the calibration measurement is fitted with the parameterization explained in the text and shown with a red line.

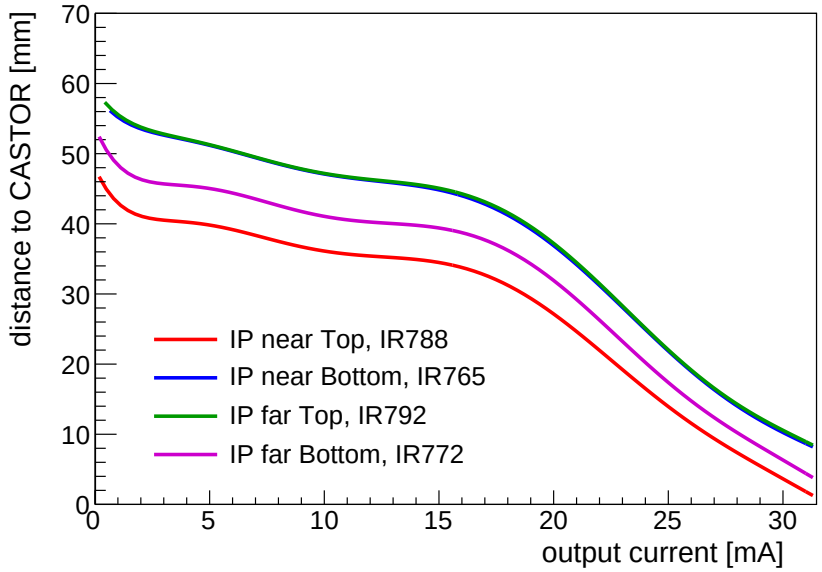



**Fig. 4.19.:** Calibration curves of four infrared sensors mounted on the front side of CASTOR during the 2016 data taking period.

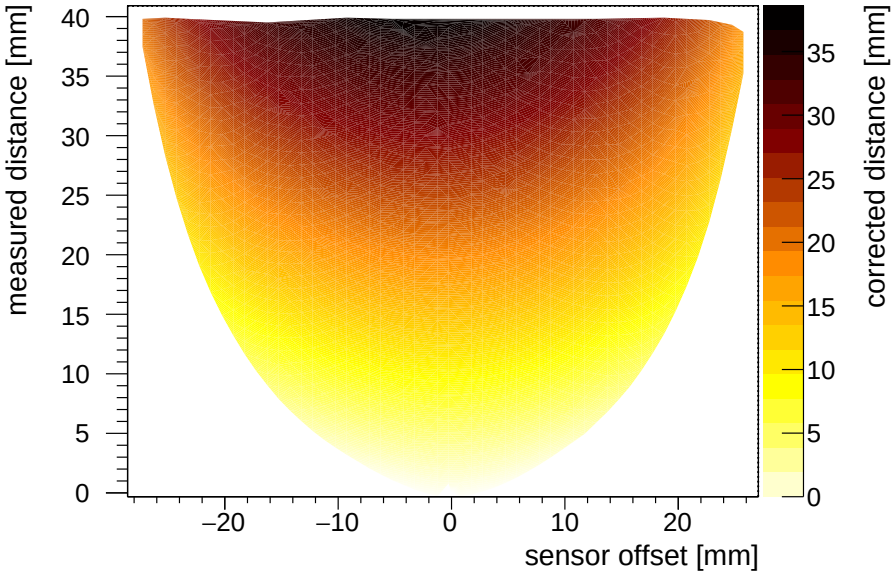



**Fig. 4.20.:** Interpolated correction function for the sensor *IR633*. The true sensor distance $d_{\mathrm{corr}}$ is shown as a function of the measured distance $d_{\mathrm{meas}}$ and the sensor offset $a$.

Based on the full calibration dataset, a two-dimensional table is calculated correcting the measured distance as a function of the sensor offset $a$ from the central beam pipe axis: $d_{\mathrm{corr}}\,(a, d_{\mathrm{meas}})$. Since the calibration measurement has been performed at finite positions, they need to be interpolated. This is done via a Delaunay interpolation of the data sample as it is implemented in the ROOT framework. As observed in Fig. 4.15, the correction is larger for larger sensor offsets. An example of this correction is shown in Fig. 4.20 for the sensor *IR633* that was used in the June 2015 data taking period. Table 4.1 summarizes the measured and corrected distances as well as the sensor offset for all ten sensors at the $\chi^2$ minimum used for the alignment in June 2015.

During the $\chi^2$ minimization the offset $a$ is calculated for every sensor independently as a function of the position of CASTOR in ($x$,$y$) and the measured distance $d_{\mathrm{meas}}$ is accordingly corrected. This is done at every minimization step as the position of CASTOR changes and the sensor positions change accordingly. This modifies Eq. (4.2) to

$$\chi^2 = \frac{d_{\mathrm{corr}}\,(x, y, d_{\mathrm{meas}}) - d_{\mathrm{calc}}\,(x, y)}{\sigma_{\mathrm{meas}}}\,. \tag{4.3}$$

**Tab. 4.1.:** Measured distance, sensor offset and corrected distance for the infrared sensors used in the June 2015 data taking period.

| sensor position | $d_{\mathrm{meas}}$ / mm | $a$ / mm | $d_{\mathrm{corr}}$ / mm |
|---|---|---|---|
| IP near top | 11.23 | 1.43 | 10.07 |
| IP near bottom | 13.18 | -2.38 | 11.11 |
| IP far top | 12.06 | 5.43 | 11.02 |
| IR far bottom | 18.99 | 0.88 | 17.43 |
| non-IP near top | 13.18 | -3.04 | 11.76 |
| non-IP near center | 17.24 | 0.41 | 17.19 |
| non-IP near bottom | 11.67 | 3.35 | 10.72 |
| non-IP far top | 14.59 | 7.46 | 12.56 |
| non-IP far center | 17.76 | 4.66 | 16.09 |
| non-IR far bottom | 21.36 | -3.89 | 20.38 |

## 4.2 Alignment of CASTOR for the 2015 proton-proton data taking

In the following the results of the CASTOR detector alignment for the 2015 proton-proton collision period in June 2015 are reported in detail. The results for the position of CASTOR during the lead-lead collisions in November 2015 as well as the proton-lead collisions in November 2016 are given in appendix A.2.

The installation of CASTOR took place on March 12, 2015. Following the installation, a position measurement was performed using the laser system. The obtained positions of each of the six targets are given in Tab. 4.2. The position of each half of CASTOR is fitted with four free parameters $x$, $y$, $\phi$, and $\rho$ to match the measured target positions. The results and uncertainties are given in Tab. 4.3 and a visualization of the obtained position in shown in Fig. 4.21. It was found, that CASTOR was installed at a position few millimeters away from the nominal position centered around (0,0) in the ($x$,$y$) plane. Furthermore, small angular rotations $\phi$ and $\rho$ are observed.

Simultaneously, a position fit based on the infrared sensors was performed. For the far half of CASTOR, the obtained position of CASTOR agrees well with the laser measurement, while for the near half the fit to the infrared data does not reproduce the shift in the $y$-direction. After the shielding of CMS was closed, the CMS magnet was turned on with a field of 3.8 T on March 19. During the magnet ramp, all sensors on CASTOR recorded significant movements (see Fig. 4.22). On April 1st, the magnetic field was ramped down and the data from the infrared sensors were evaluated again. A second position fit was performed whose result is shown in Fig. 4.23. A summary of all obtained position fits is given in Tab. 4.4. It was found that CASTOR almost recovered to the original position. The shift due to the magnet cycle was below 1 mm on the IP side of CASTOR, while it was

**Tab. 4.2.:** Results of the laser position measurement of CASTOR after the installation on March 12, 2015. The obtained positions of the laser targets are given in the CMS coordinate system.

| CASTOR half | target ID | determined position ($x$,$y$,$z$)/ m in the CMS coordinate system |
|---|---|---|
| near half | XP_ZM_Top | (0.0929 , 0.3163 , -15.8920) |
| | XP_ZP_Mid | (0.3166 , 0.0877 , -14.8825) |
| | XP_ZM_Bottom | (0.3179 , -0.0933 , -15.8929) |
| far half | XM_ZM_Top | (-0.0950 , 0.3123 , -15.8892) |
| | XM_ZP_Mid | (-0.3220 , 0.0855 , -14.8804) |
| | XM_ZM_Bottom | (-0.3208 , -0.0971 , -15.8904) |

**Tab. 4.3.:** Results of the position fit of each CASTOR half obtained with the laser position measurement on March 12, 2015. The position of the two CASTOR halves is parameterized with four parameters $x$, $y$, $\phi$, and $\rho$.

| CASTOR position fit | $x$/ mm | $y$/ mm | $\phi$ | $\rho$ |
|---|---|---|---|---|
| near half | 0.78±1.34 | -7.2±1.4 | -0.26° | 0.0064° |
| far half | -5.0±1.34 | -5.5±1.4 | -0.0016° | 0.042° |

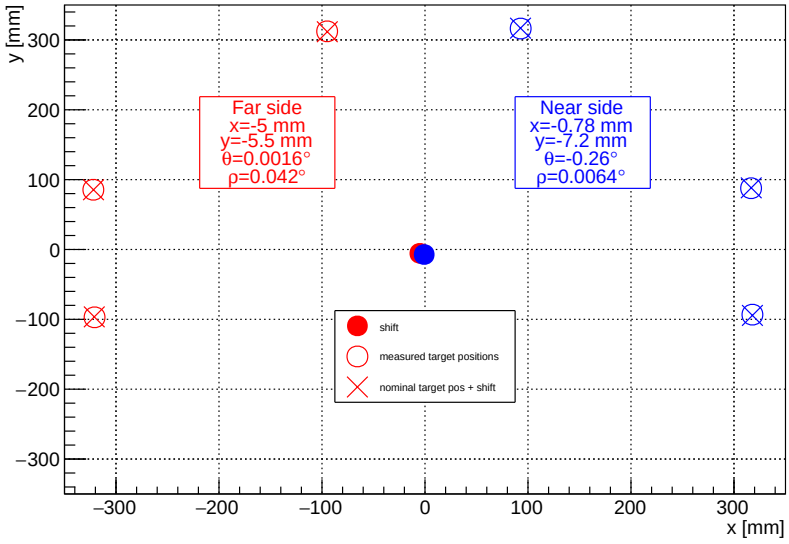



**Fig. 4.21.:** Result of the CASTOR position fit. The position of the two CASTOR halves is parameterized with four parameters $x$, $y$, $\phi$, and $\rho$. The measured positions of the laser targets are indicated with open circles and their positions resulting after the fit with crosses. The resulting shift of CASTOR in the ($x$,$y$)-plane is visualized with filled circles.

**Tab. 4.4.:** Summary of the alignment results of CASTOR. The ($x$,$y$) position is given as obtained with the laser measurement as well as the fit to the infrared sensors both before and after the magnet cycle. The shift resulting from the magnet cycle is once calculated with the infrared sensors and once with the contact potentiometer sensors.

| | near half | | far half | |
|---|---|---|---|---|
| | $x/$ mm | $y/$ mm | $x/$ mm | $y/$ mm |
| laser measurement after installation | 0.78±1.3 | -7.2±1.4 | -5.0±1.3 | -5.5±1.4 |
| IR sensors after installation | -1.9±2.2 | -1.9±1.8 | -4.3±1.9 | -2.8±2.4 |
| IR sensors after magnet cycle | -2.2±2.2 | -1.6±1.8 | -4.7±1.9 | -2.9±2.4 |
| shift after magnet cycle (IP side) | -0.33 | 0.22 | -0.40 | -0.06 |
| shift after magnet cycle (non-IP side) | -1.78 | -0.74 | -1.61 | -0.19 |
| shift of potentiometers (non-IP side) | -1.35 | -0.23 | -1.52 | -0.59 |

larger at the non-IP side. The contact potentiometers were also evaluated and recorded a very similar shift due to the magnet cycle. Since the magnet stayed off until CASTOR was removed again, a two-step procedure was chosen to identify the position of CASTOR. First, the weighted mean of the positions obtained with the laser and infrared sensors are calculated and second, the shift due to the magnet cycle as observed with the infrared sensors was added. This leads to the position of CASTOR during the June 2015 data taking period at

$$\begin{aligned} \text{Near Half } (x,y)/\,\text{mm} &= (-2.23 \pm 2.17, -1.65 \pm 1.75) \\ \text{Far Half } (x,y)/\,\text{mm} &= (-5.15 \pm 1.20, -4.86 \pm 1.46)\ . \end{aligned} \tag{4.4}$$

This alignment strategy is so far the most thorough method to determine the position of CASTOR during data taking. In future applications, new infrared sensors will need to be calibrated with the same setup. The method and tools developed in this thesis are the benchmark for the alignment of CASTOR.

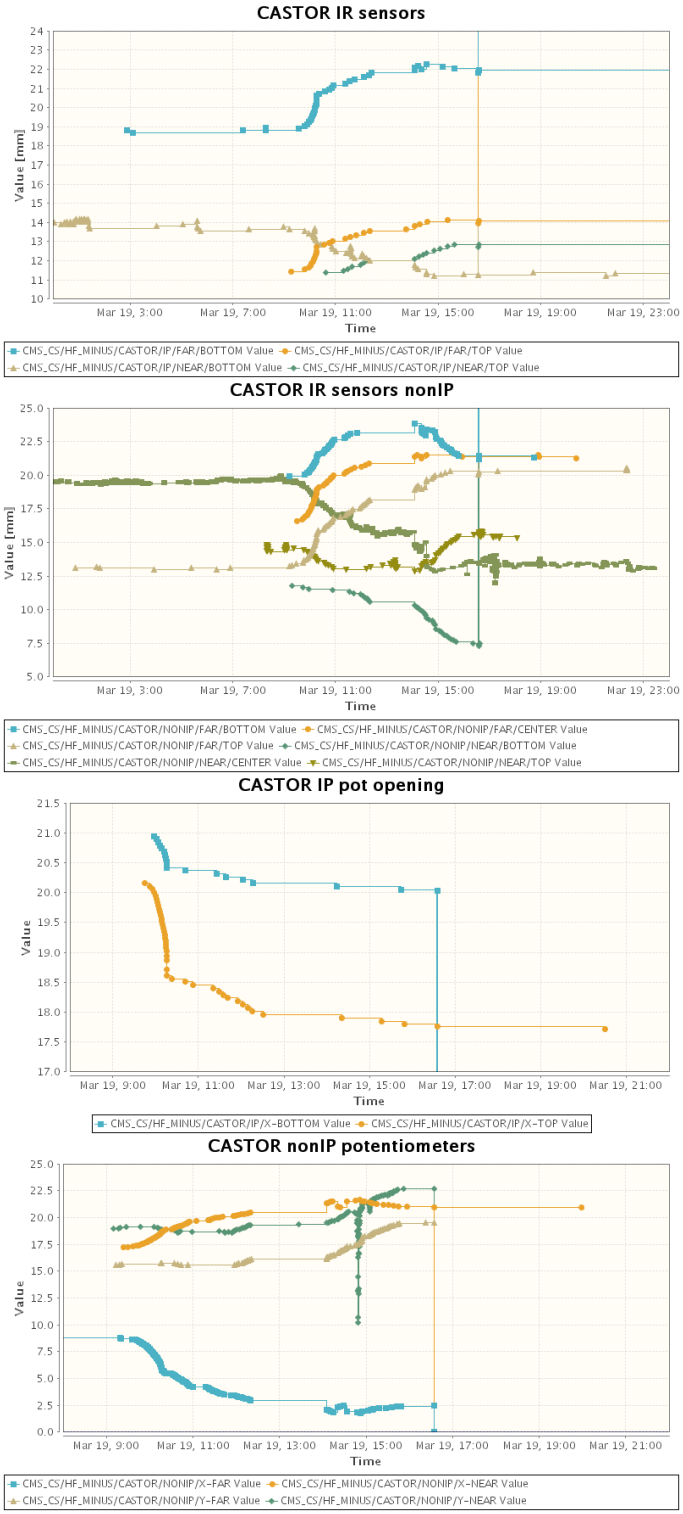



**Fig. 4.22.:** Data of all distance sensors as provided by the online CMS monitoring system during the time when the CMS solenoid was turned on. All sensors show a significant change of their readings while the magnetic field changes. This indicates a movement of the CASTOR half-cylinders. Top: infrared sensors pointing towards the LHC beam pipe on the IP side of CASTOR. Second row: infrared sensors pointing towards the LHC beam pipe on the non-IP side of CASTOR. Third row: potentiometer sensors measuring the opening between the two CASTOR halves. Bottom: potentiometer sensors measuring the movement relative the beam pipe support structure.

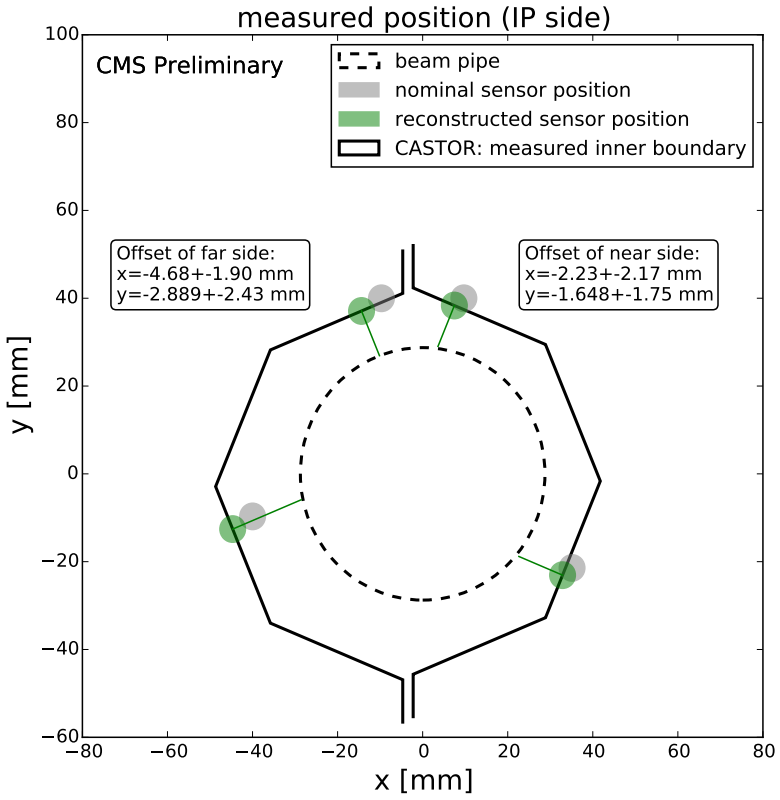



**Fig. 4.23.:** Result of the CASTOR position fit obtained with the data from the infrared distance sensors. The position of the two CASTOR halves is given by a offset from the ideal position in the $(x,y)$-plane of the CMS reference frame. The obtained positions of the infrared sensors is indicated by green circles. Gray circles indicate their respective position for a nominal CASTOR position at $(x,y)=(0,0)$. The inner boundary of the detector at the measured position is illustrated by a black line.

# 5 Simulation and data validation

Any data analysis with CASTOR requires accurate detector simulation for acceptance and efficiency corrections. Especially the detector position and other conditions need to be considered correctly in simulations as well as in the evaluation of the data quality, the reconstruction, and later the analysis of the data. After the long shutdown of the LHC, CASTOR was re-installed in CMS in June 2015 for the first weeks of data taking of the LHC Run 2. Due to problems with the cryogenic system of the CMS solenoid, there was no magnetic field inside the CMS volume. These special circumstances reinforce even more the need for a detailed check of the data quality as well as a validation of the calibration. Detailed studies were carried out in the scope of this work to ensure a good quality of the recorded data as well as a good understanding of the calorimeter properties. Most attention is payed to whether the data agrees with Monte Carlo simulations in several control observables.

## 5.1 Software and detector simulations

The reconstruction of data recorded by the CASTOR calorimeter is completely integrated in the standard CMS software *CMSSW*. This provides a straightforward implementation of all data taking conditions such as noise levels and calibrations into the central CMS databases. Furthermore, the data of CASTOR can be easily combined with data from other CMS subdetectors. Within the CMS working group on *Forward and Small-x QCD Physics* (FSQ) a new analysis framework was developed for the LHC Run 2 called *Common FSQ Framework* (CFF). This framework provides tools to produce small files with a tree-like structure based on the more complex data files produced by data reconstruction. These *trees* based on the ROOT software [92] contain only the event information that is relevant for the later data analysis and thus provide smaller file sizes as well as faster processing time than with the complex reconstruction files. Within this thesis significant effort was made to make the CASTOR data accessible within the CFF. Both low-level information such as the reconstructed energy for every individual readout channel were implemented as well as higher level objects such as calorimeter towers with applied noise cuts and jets. Table 5.1 summarizes all available CASTOR information in the CFF. With the modular design of the CFF, users are able to add or remove the CASTOR objects by using simple flags during the production of the reduced tree files. The CFF is widely used within the FSQ community of CMS and also once produced tree files have been shared and used by

**Tab. 5.1.:** List of CASTOR reconstruction objects that are available in the Common FSQ Framework.

| Reconstruction Step | Information | Comment |
|---|---|---|
| RecHits | Energy | Calibrated energy in units of GeV |
| | Sector | Azimuthal sector of the readout channel |
| | Module | Longitudinal module of the readout channel |
| | Quality | Boolean flag whether or not the readout channel is operational |
| | Saturation | Boolean flags whether or not a readout channel was saturated and successfully corrected |
| Towers | p4 | 4-vector containing position and energy of the tower |
| | emEnergy | Energy contained in the electromagnetic part of the tower |
| | hadEnergy | Energy contained in the hadronic part of the tower |
| | Size | Number of operational channels within the tower |
| Jets | p4 | 4-vector containing position and energy of each reconstructed jet |
| | fem | Fraction of the jet-energy contained in the electromagnetic part of the calorimeter |
| | Depth | Depth of the jet maximum in the longitudinal direction |
| | Size | Number of calorimeter towers clustered to one jet |

various people and groups. The CASTOR data therefore is easily accessible and the unique data is used in published and ongoing analyses. All data analysis presented in this work was performed with the CFF.

The use of the CFF is not only restricted to reconstructed data but can in the same way be used for Monte Carlo simulations as well. Also here, CASTOR is completely integrated into the CMS detector simulation. The detector simulation is based on the GEANT 4 toolkit [110]. A precise geometrical description of the CASTOR calorimeter is implemented in this framework, see Ref. [91] for details. The quality of the detector simulation of CASTOR is of great importance for precise physics analyses. Special care is taken to ensure that the simulations correctly reproduce the detector response as measured in the test beam [111]. All possible crosschecks are made to validate the quality of the simulations.

## 5.2 Data and Monte Carlo samples

Five Monte Carlo samples have been produced and are used to validate the data, correct it to particle level and to estimate systematic uncertainties. The samples are generated with the CMS software framework version *7_1_20_patch3*. An overview of the samples is given in table 5.2. The generated Monte Carlo samples are provided in the form of ROOT trees compatible with the Common FSQ Framework. This allows to easily process Monte Carlo predictions along with the data, and to distribute the samples.

Special attention is payed to reproduce the detector conditions during data taking. Thus, no magnetic field is simulated as well as a realistic position of the proton-proton luminous region within the CMS reference frame. Moreover, 11 of the 224 readout channels are found to be unsuited for data analysis because they have too high noise, can't be calibrated with the muon intercalibration method or have power supply problems (see Tab. 5.3). These channels are excluded in the simulations as well as in the validation procedure and data analysis. They can be identified as white areas in Fig. 5.1. Furthermore, the position of CASTOR as it is measured after the installation is taken into account (numbers given in Eq. (4.4)). Two samples are generated with an additionally shifted position to study the influences of the alignment uncertainty on the data analyses. Therefore, CASTOR is once shifted in the positive directions in both $x$ and $y$ by the respective uncertainty and once in the negative direction. Three different Monte Carlo event generators are chosen in order to cover a broad range of different descriptions of hadronic interactions. Moreover, the considered models have proven to provide good descriptions of proton-proton collisions at more central pseudorapidities. These are PYTHIA 8.2 with tune CUETP8M1 and tune 4C combined with the MBR model for diffraction as well as EPOS LHC.

The data used for validation were recorded in CMS run 247324 using an unbiased trigger. For noise studies, dedicated data sets with only single or no beam presence combined with a random trigger are used. For all further validation and analysis steps, a special event selection is chosen to select collision events. At least one reconstructed calorimeter tower above a noise threshold is required in the HF calorimeters on either the positive or negative hemisphere of CMS. The thresholds are chosen in a way to minimize the selection of events due to electronic noise or beam backgrounds. Figure 5.2 shows the energy distributions of the HF towers close to the noise level for unbiased data as well as single- and no-beam-samples, and Monte Carlo simulations. A threshold of 5 GeV (indicated by the red dashed line) separates clearly the noise peak from the signal tail and is therefore chosen for the further event selection.

**Tab. 5.2.:** Monte Carlo samples used for data validation and data analysis.

| Event Generator | Events | CASTOR position |
|---|---|---|
| PYTHIA 8 Tune CUETP8M1 | 4917500 | as measured |
| PYTHIA 8 Tune 4C+MBR | 4862000 | as measured |
| EPOS LHC | 4978400 | as measured |
| PYTHIA 8 Tune CUETP8M1 | 960000 | as measured plus systematic shift |
| PYTHIA 8 Tune CUETP8M1 | 980000 | as measured minus systematic shift |

**Tab. 5.3.:** Reasons for CASTOR readout channels being excluded from the analysis.

| Sector | Module | Reason for exclusion |
|---|---|---|
| 2 | 10 | very faint muon signal |
| 3 | 8 | high noise |
| 3 | 11 | high noise |
| 5 | 4 | high noise |
| 7 | 5 | high noise |
| 8 | 8 | unexpected muon signal |
| 9 | 4 | high noise |
| 12 | 12 | high noise |
| 14 | 8 | unexpected muon signal |
| 15 | 4 | HV problems |
| 16 | 4 | HV problems |

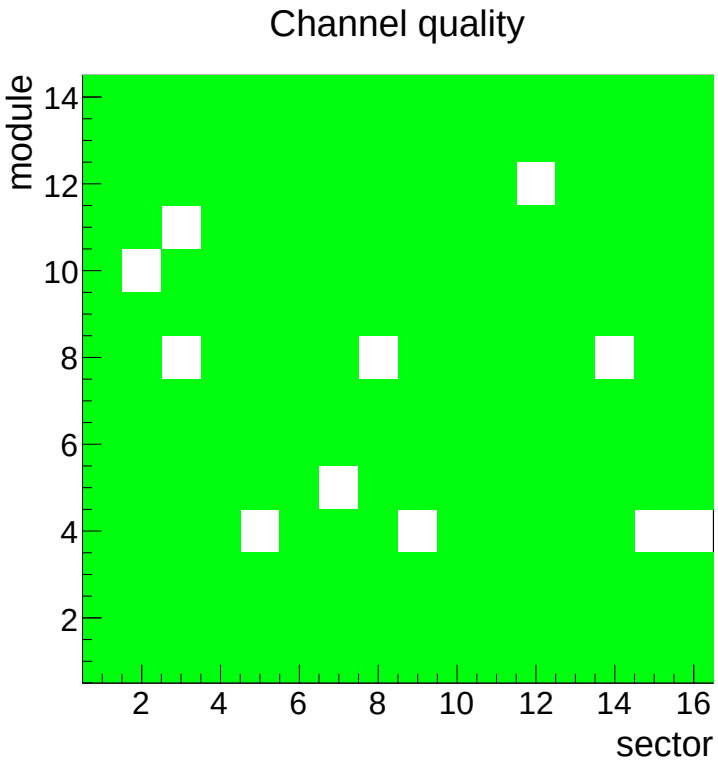



**Fig. 5.1.:** Quality of the CASTOR readout channels, in the sector-module representation. Fully operational channels are marked green, non-operational channels are white.

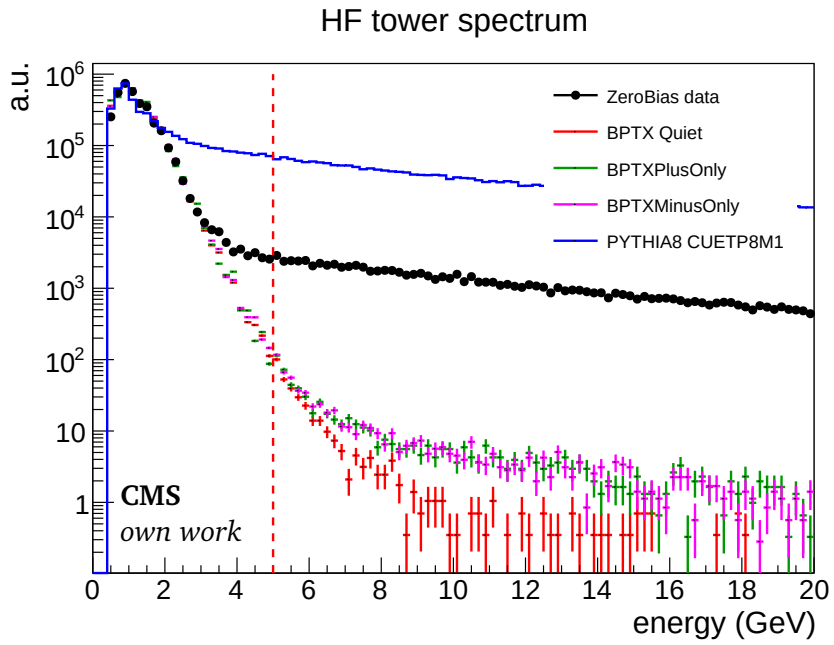



**Fig. 5.2.:** Energy distribution of HF-towers for ZeroBias data (black markers), noise samples (colored markers) and PYTHIA 8 Monte Carlo simulations (blue line). The chosen selection cutoff threshold of 5 GeV is indicated by a red dashed line.

## 5.3 Data quality and validation

Due to the very forward phase space, small shifts of the CASTOR calorimeter will result in significant changes of the expected observed energy and thus directly affects the calibration of the calorimeter. The uncertainty on the energy scale due to the position uncertainty of CASTOR is investigated with the PYTHIA 8 CUETP8M1 samples at the various detector positions. In order to study the effect of possible misalignment, the ratio of reconstructed energy with respect to the true generator level energy in the acceptance $-6.6 < \eta < -5.2$ is calculated for all three positions. Figure 5.3 shows this distribution. The peak around $0.6$ is due to the non-compensating properties of CASTOR. The yellow error band indicates the variation due to the position uncertainty of CASTOR. As a result, the uncertainty on the energy scale of the detector due to the alignment uncertainty is found to be 7.5%. Until now, this is the most precise determination of the CASTOR alignment being implemented in Monte Carlo simulations and used for data analysis. It represents a significant improvement with respect to previous data taking periods, reducing the total uncertainty on the CASTOR energy scale from 22% to 17%.

In order to check the performance of the calibration as well as of the alignment, profiles of the average energy deposited per collision in CASTOR are investigated in the azimuth plane as well as in the longitudinal direction. For an ideal geometry, when the interactions take place on the central axis of the calorimeter, a flat distribution of the average energy is expected in the azimuthal plane. Due to the fact that the interaction point is shifted with respect to the CMS coordinate system and furthermore CASTOR is not perfectly aligned with the central axis of CMS, a wavelike structure is obtained as a function of the azimuth $\phi$. This is reproduced by Monte Carlo simulations that account for these shifts. Figure 5.4 compares the $\phi$-profile observed in data to those obtained from simulations. The standard muon intercalibration (open markers labeled *Data original*) is not in very good agreement with what is obtained from simulations. Although there are hints for a general agreement, some sectors show significant deviations from the predictions. Therefore a new constraint is added to the intercalibration method. The response of the channels to the beam halo muons is in addition required to be symmetric in $\phi$ while conserving the overall energy scale. This improves the description of the azimuthal profile significantly. The data points (solid markers) are mostly enclosed by the curves representing the alignment uncertainties (red and green curves in Fig. 5.4) and the data agrees within the energy scale uncertainties with the prediction for the measured geometry.

The longitudinal profile is also of major importance since, for example, the front two modules are used to separate the electromagnetic from the hadronic component. A change in the relative intercalibration of the front to the back channels would result in a misreconstruction of the ratio of the electromagnetic component to the hadronic energy. Here, data and Monte Carlo simulations show good agreement. The increase of the response in the data in the most backward modules is understood to be a feature of

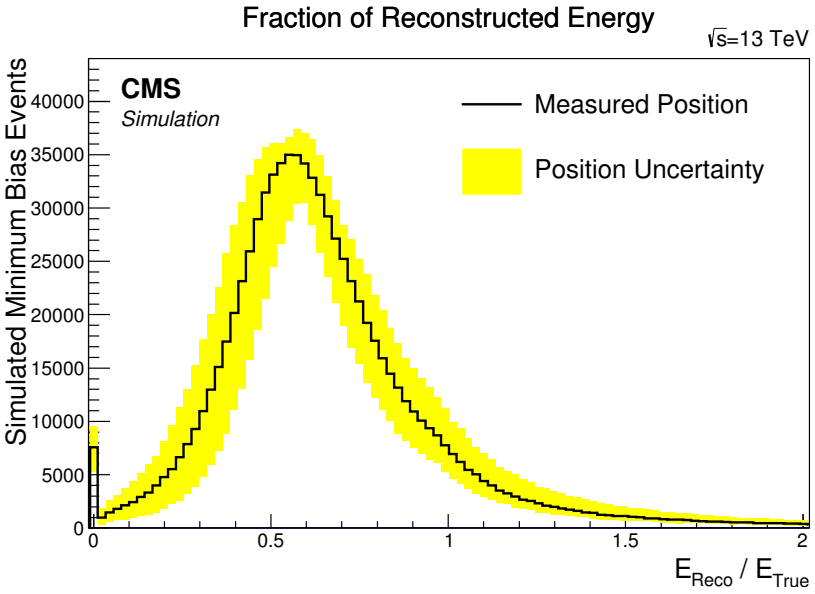



**Fig. 5.3.:** Distribution of the ratio between reconstructed energy and true energy in the acceptance $-6.6 < \eta < -5.2$ based on PYTHIA 8 tune CUETP8M1. The distribution at the measured position is shown as a black line. The yellow error band indicates the variation of this distribution due to the position uncertainty of CASTOR as given in Eq. (4.4). Published in Ref. [112].

the muon intercalibration. A signal in this region is required in the selection procedure for muon-like events. Since especially the low energetic muons may already to be stopped inside the calorimeter, this leads to a slight artificial increase in the response correction. In order to avoid a bias due to this effect the last two modules (13 and 14) are excluded from any data analysis. Since they do not contribute significant energy deposits, this has negligible impact on the reconstruction of the overall energy. In general, it is one result of this thesis to demonstrate how difficult it is to consistently describe all longitudinal and azimuthal channels of CASTOR. Fortunately, this is not absolutely required by most data analyses. Also here, the average over $\phi$ is used and the longitudinal separation is only used to split into the electromagnetic and hadronic sections of single towers. Thus, it is a particularly important benchmark test to verify the longitudinal intercalibration is the fraction of electromagnetic energy within one tower. This is compared in Fig. 5.6 to simulations and is found to be in excellent agreement. This is the most relevant property for the analysis of this thesis.

The electronic noise of the readout channels is investigated on the level of reconstructed calorimeter towers. The energies of all working channels in one azimuthal sector are summed up to form one calorimeter tower. Only towers with an energy of at least $\sqrt{N} \times 0.65$ GeV are to be considered in data analysis, where $N$ is the number of functioning and calibrated channels in this sector. Figure 5.7 shows the distribution of the tower energy divided by $\sqrt{N}$ for four different data samples. For data with either no or only one bunch coming from the minus direction (*BPTXquiet* and *BPTXMinusOnly* respectively),

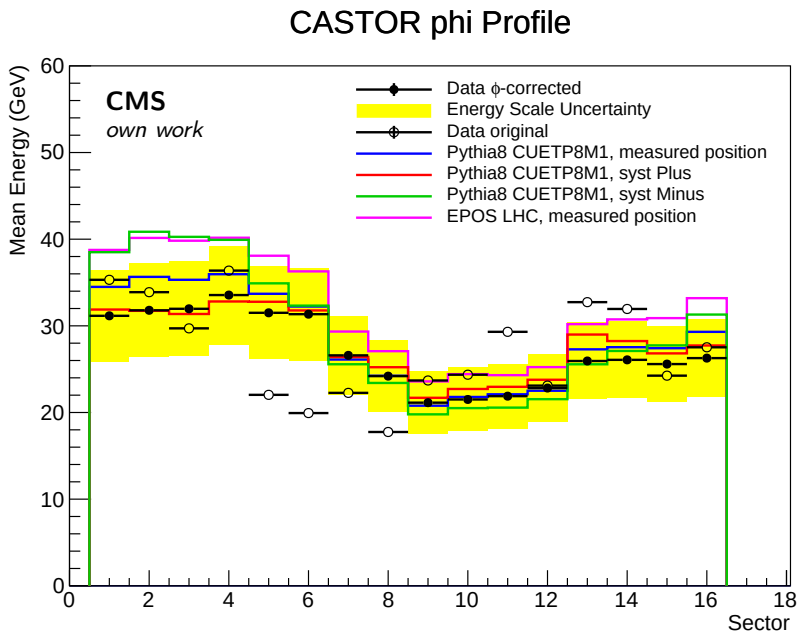



**Fig. 5.4.:** Profile of the average energy reconstructed in CASTOR as a function of the azimuthal sector. Data is shown with black markers and the energy scale uncertainty as yellow error band. This is compared to PYTHIA 8 tune CUETP8M1 and EPOS LHC with the measured CASTOR position as well as systematic shifts in plus and minus direction.

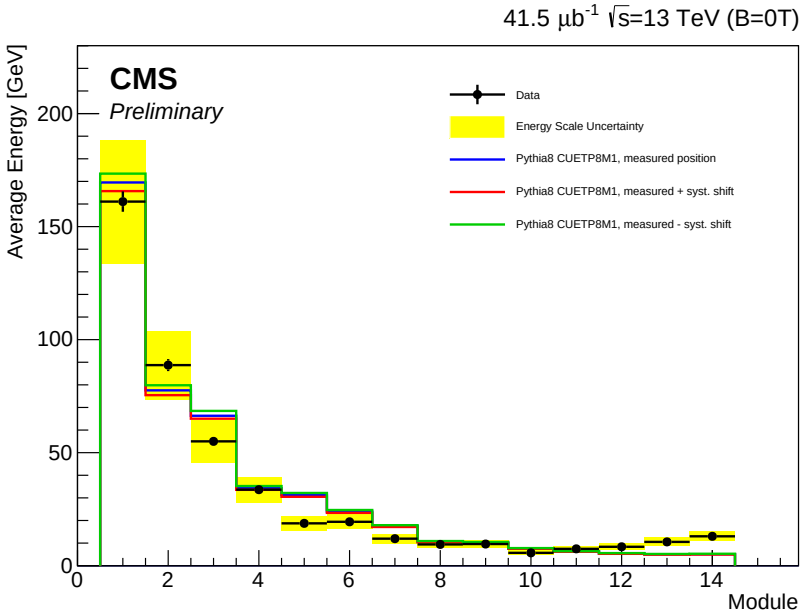



**Fig. 5.5.:** Profile of the average energy reconstructed in CASTOR as a function of the longitudinal module. Data is shown with black markers and the energy scale uncertainty as yellow error band. This is compared to PYTHIA 8 tune CUETP8M1 and EPOS LHC with the measured CASTOR position as well as systematic shifts in plus and minus direction. Published in Ref. [112].

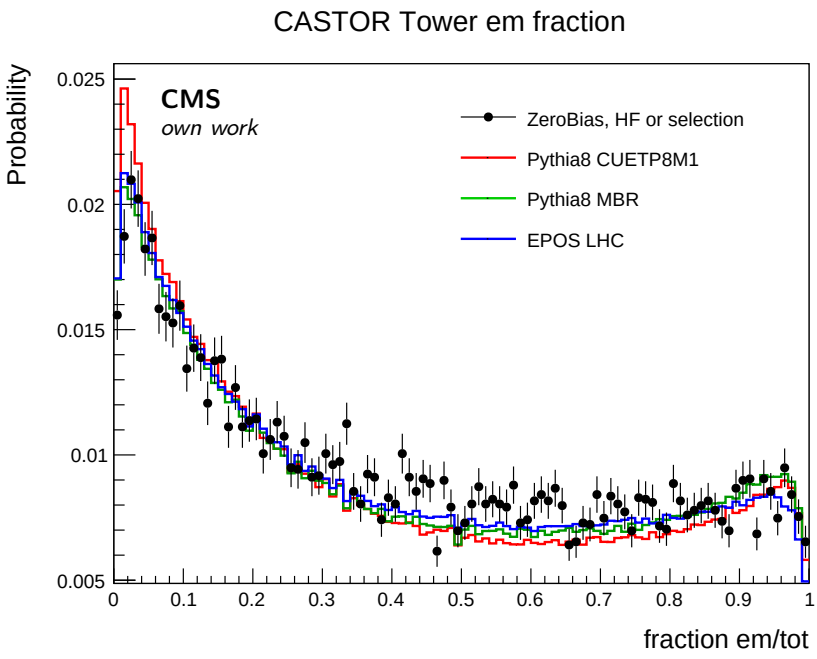



**Fig. 5.6.:** Event-by-event fraction of reconstructed electromagnetic energy within one CASTOR tower. Data is compared to Monte Carlo simulations.

the electronic noise is mainly well below the threshold energy of 0.65 GeV. A tail towards higher energies is observed in events with only one bunch coming from the plus direction and thus going towards CASTOR. This is due to beam induced backgrounds. However, the probability to see these kind of energies is about two orders of magnitude lower than for data with colliding bunches (ZeroBias). The average energy of beam backgrounds is therefore about a factor of 200 lower. The distribution of the largest reconstructed tower energy is shown in Fig. 5.8. This demonstrates, that the electronic noise is well cut away and behaves similar in data and in simulations. The fact that the fraction of events with no tower above threshold is higher in data is due to the low interaction probability of only 5 % in the ZeroBias data. Thus, almost 95 % of all events do not have any LHC collision.

As a final step, the calibration of the absolute energy scale is verified. The calibration performed with proton-proton collision data at a center-of-mass energy of 7 TeV (see chapter 3.2 for details) is transferred to the running conditions at the higher center-of-mass energy using the muon intercalibration method. This transformation is validated by reconstructing the data with two different sets of calibration constants. As a reference, the calibration used during the 2013 data taking period is used. These are corrected for gain differences due to different high voltage settings and the fact of no CMS magnetic field.[1] This is compared to the updated calibrations for the 2015 data taking period. One obtains an energy scale which is consistent within the systematic uncertainties.

[1] A LED pulser introduces light flashes directly into the PMTs in order to compare the signal strengths measured by the PMTs at various conditions. The calibration constants can then be corrected for the different running conditions.

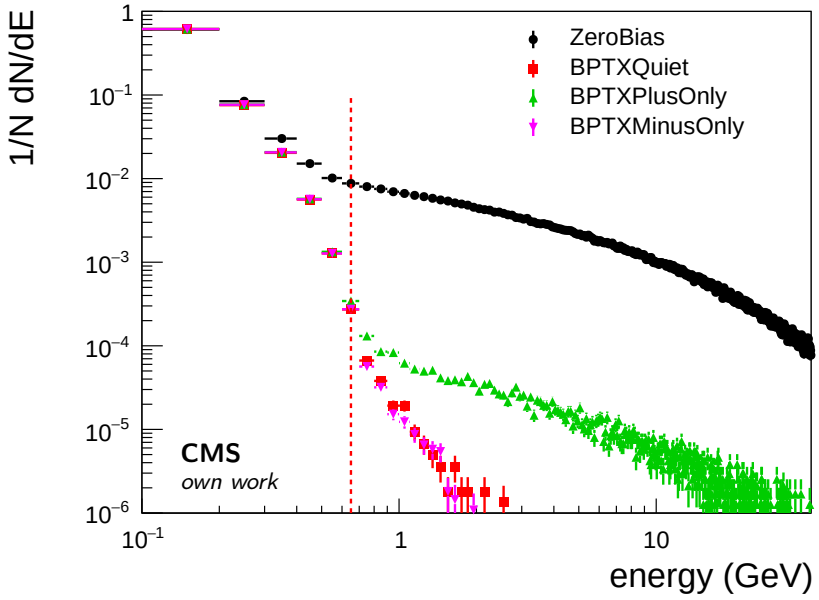



**Fig. 5.7.:** The energy in a calorimeter tower divided by the square root of the number of good channels in the respective tower is shown for various event selections. The cutoff of 0.65 GeV (indicated by a dashed red line) discriminates well between signal and electronic noise (BPTXQuiet and BPTXMinusOnly). Beam induced backgrounds (seen in BPTXPlusOnly) are reduced by about two orders of magnitude.

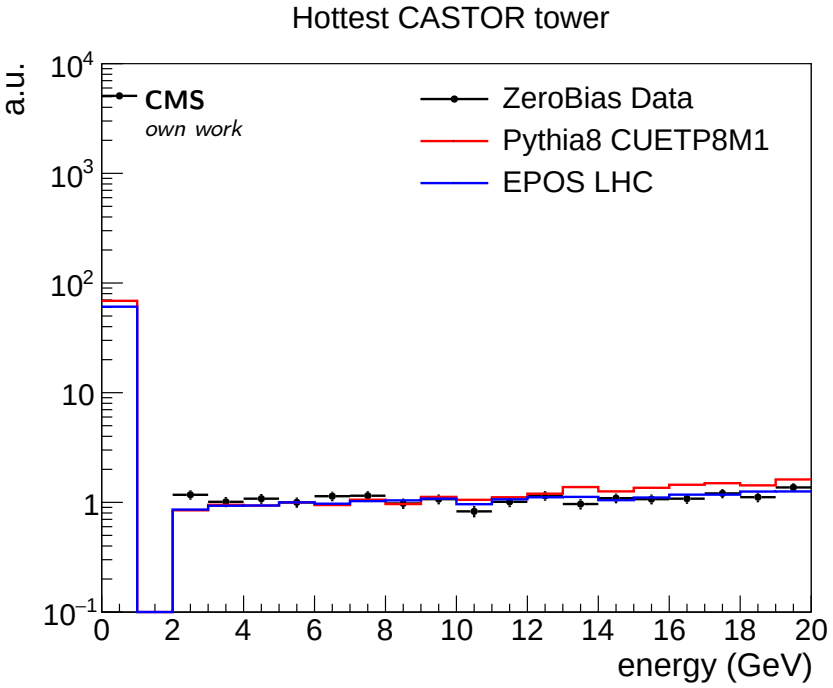



**Fig. 5.8.:** Energy distribution of the hottest CASTOR tower for ZeroBias data and Monte Carlo simulations.

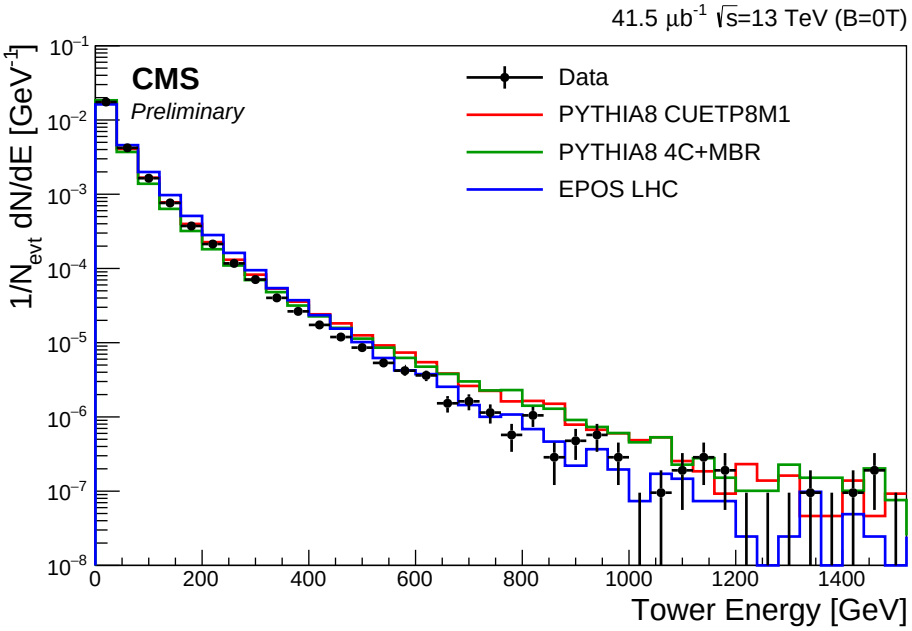



**Fig. 5.9.:** Energy spectrum of single reconstructed CASTOR towers, normalized by the number of events that pass the offline event selection. Very good agreement between data and Monte Carlo simulations is found. Published in Ref. [112].

As a result, the energy distribution of reconstructed calorimeter towers is shown in Fig. 5.9 for both data and Monte Carlo. Excellent agreement between data and simulations is found. Overall, the performance of the CASTOR calorimeter and the data quality is very good. Key features such as the alignment and the calibration are well understood and provide consistent results.

## 5.4 Improved online data quality monitoring during the 2016 proton-lead data taking

With the experience gained during the data validation of the 2015 proton-proton data taking, the validation procedure was changed to an on-the-fly system during the 2016 proton-lead data taking period. The CMS software provides an online Data Quality Monitoring (DQM) tool, that is designed to study small collision data sets within few hours after they have been recorded. A set of distributions was therefore selected and added to the DQM software so that a group of trained experts was able to immediately investigate the data quality. As an example a screenshot of the *RecHit Occupancy Map* is shown in Fig. 5.10. This method was used for the first time by the CASTOR group. As a consequence, the CASTOR detector was also included in the central CMS data certification program. Run periods are defined in which the detector was fully operational and simplifies data analysis in the future. The fact that CASTOR is included in the central validation procedure is a big achievement for the community and the recorded data can be analyzed without the need

**Tab. 5.4.:** List of observables present in the CASTOR online DQM, together with a detailed description of the desired behavior. All requirements have to be met in order to flag a run as good.

| Observable | Requirement |
|---|---|
| Unpacker Report | The number of CMS readout units (FEDs) must be equal to four for all events. |
| Dead Channel Map | The dead channel map must be uniform and equal to zero. |
| Report Summary Map | The fraction of events with a reported error must be below one permille. |
| CASTOR DigiSize | The size of the Digi collection should always be equal to six. |
| Readout Timing | The main signal should be visible in time slice 4 for all channels. The signal in time slice 5 should not exceed 15% of the signal in time slice 4. |
| RecHit Occupancy Map | The characteristic structure of CASTOR should be seen with an increased signal in the front modules, an increased signal in module 5 and almost uniform along the sectors, see Fig.5.10. |
| Noise | No significant noise level is allowed to be observed. |
| Trigger | The trigger thresholds should match the running conditions. |

of many additional technical crosschecks. A detailed list of the quality criteria is given in Tab. 5.4 and a complete list of all certified runs is given in appendix B. The data that in the end was labeled to have good quality corresponds, for the different data taking periods, to an integrated luminosity of :

- 241.892 $\mu\mathrm{b}^{-}1$ (5 TeV proton-lead)
- 64.378 $\mathrm{nb}^{-}1$ (8 TeV proton-lead)
- 115.277 $\mathrm{nb}^{-}1$ (8 TeV lead-proton)

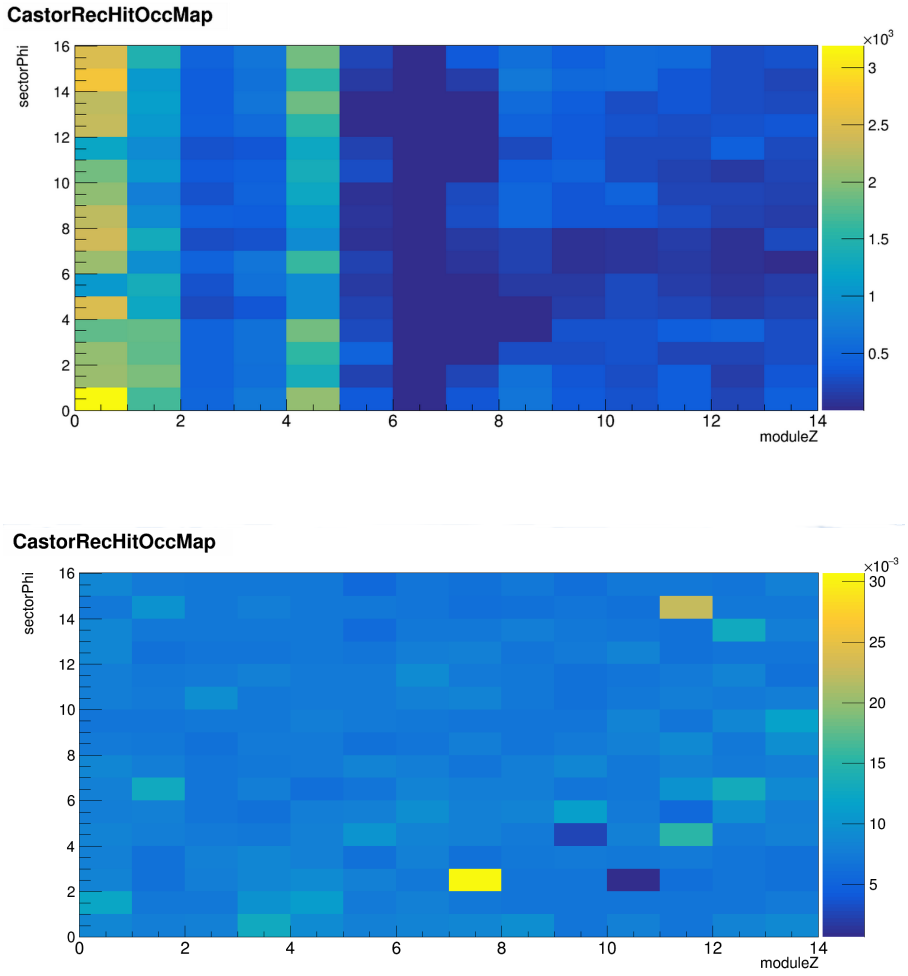



**Fig. 5.10.:** Examples of the *RecHit Occupancy Map* in a good run (top) where the characteristic structure of the CASTOR response is seen. On the bottom, the same distribution is shown for a data sample without collisions. Here no increase noise level is seen except for two known noisy channels.

# 6 Measurement of the very forward energy with CASTOR

> *If you torture the data long enough, it will confess.*
>
> **— Ronald Coarse**
> (Economist, Nobel Laureate)

The description of the underlying event in proton-proton collisions remains a big challenge. Since the collision activity in the very forward phase space with pseudorapidities $|\eta| > 5$ is significantly influenced by the underlying event mechanisms, measurements in this region will contribute to a better understanding of the related physics. Moreover, the particle production in the very forward direction also provides crucial input for extensive air shower calculations since the bulk of the available collision energy is carried by the particles produced into this region of phase space (see also Fig. 4.1). Very forward particle production has already been studied at the LHC at center-of-mass energies of 0.9, 2.76, 7 and 8 TeV for example by the CMS [57, 113], TOTEM [113–115], and LHCf [116, 117] experiments. In the present work these measurements are extended with data taken by the CASTOR calorimeter at a center-of-mass energy of 13 TeV.

The most fundamental measurement are the inclusive very forward energy spectra. The spectra are determined in terms of the production cross section $\mathrm{d}\sigma/\mathrm{d}E$ as a function of the energy that is carried by the particles in the acceptance of CASTOR. This already lead to the first publication of a comparable measurement in the very forward acceptance at 13 TeV [118] and is one of the main results of this thesis. The design of CASTOR is exploited to also separate between energies deposited in the electromagnetic and hadronic sections of the calorimeter, which gives additional information on the description of the energies carried by different particle species in the very forward phase space. The results are fully corrected to the particle level with an unfolding technique. The average energy is included in the measurement of the energy density as a function of pseudorapidity, $\mathrm{d}E/\mathrm{d}\eta$, in the forward region using also the HF calorimeter [119].

In a further step the energy deposit in CASTOR is correlated to the collision activity in the central region. The average energy is evaluated as a function of the charged particle

multiplicity at central rapidities ($|\eta| < 2$). Also here the separation into the electromagnetic and hadronic contributions ultimately enhances the physics reach of the measurement.

The presented measurements are unique among the LHC experiments and provide important insights to understand global properties of hadronic collisions.

## 6.1 Inclusive very forward energy spectra

The data used for this analysis was recorded in CMS run 247324. After lumisection 83, the LHC beams were separated at the CMS interaction point thus providing data with very low luminosity. The CMS data taking conditions (e.g. trigger prescales) were adapted to this change at lumisection 97. Therefore, only data recorded after lumisection 97 are analyzed. Furthermore, the analysis is restricted to a single proton bunch in CMS with bunchcrossing number 208, which had a higher intensity, leading to a lower noise contribution in the event selection. The proton-proton interaction probability per bunch crossing of this bunch corresponds to about $6\%$. This is the same dataset and event selection as used for the first LHC paper at 13 TeV, the measurement of the charged hadron multiplicity distribution dN/d$\eta$ by CMS [120]. The data corresponds to a integrated luminosity of $0.34\,\mu\mathrm{b}^{-1}$ [121].

### 6.1.1 Analysis strategy

Collision events are selected online using the ZeroBias trigger and offline by requiring at least one calorimeter tower in the HF calorimeters above a noise threshold of 5 GeV. The total energy measured in CASTOR is obtained by summing up all calorimeter towers. The fraction of energy deposited in the first two longitudinal modules of the calorimeter can be interpreted as the electromagnetic component of the energy. Thus, the total energy deposited in CASTOR can be divided into the energy deposited by electromagnetic and hadronic particles. The reconstructed spectra of the total, electromagnetic and hadronic energy on detector level are normalized to the number of events that pass the offline event selection. The results on detector level are shown in Fig. 6.1 for data and different Monte Carlo event generators, the systematic uncertainties are illustrated with a yellow band and will be discussed in detail in section 6.1.3. Significant differences between the models and the data can be seen.

These three spectra are correlated in a non-trivial way and thus considering the electromagnetic and hadronic spectra individually provides complementary information on particle production in the very forward region. On detector level the ratio of the electromagnetic energy to the total energy $f_{\mathrm{em}}$ is also determined event by event. The corresponding distribution is shown in Fig. 6.2. Although the data suggest a slight shift towards a higher fraction of electromagnetic energies, the detector simulation does agree with the data

**Tab. 6.1.:** Relative amount of fake and missed events with the event selection described in the text.

| Model | Fake | Missed |
|---|---|---|
| PYTHIA 8 CUETP8M1 | 0.2% | 1.7% |
| PYTHIA 8 4C+MBR | 0.37% | 4.7% |
| EPOS LHC | 0.4% | 3.9% |

within the uncertainties. The detector level spectra are corrected for experimental effects to the energy on particle level with an unfolding technique and afterwards normalized to the differential cross section $\mathrm{d}\sigma/\mathrm{d}E$. This allows to most generally compare the data to a large set of model predictions. Moreover, also in the future these data can be easily used to benchmark or tune upcoming models.

The corresponding event selection on stable particle level that the data are corrected to, is defined in order to provide high selection efficiency relative to the detector level selection based on the HF calorimeters. A selection based on the Lorentz-invariant fractional proton-momentum-loss $\xi$ was chosen. This variable is commonly used to describe diffractive processes but is well defined for any type of final state, independent of the underlying collision mechanisms. It provides a very simple and straightforward way to define the acceptance of the measurement at particle level. All stable (lifetime: $c\tau > 1\,\mathrm{cm}$) final state particles are divided into two systems, $X$ and $Y$, based on their position with respect to the largest rapidity gap in the event. All particles on the negative side of the largest gap are assigned to the system $X$, while the particles on the positive side are assigned to the system $Y$. The invariant masses, $M_\mathrm{X}$ and $M_\mathrm{Y}$, of each system are calculated using the four-momenta of the individual particles. Their squared ratios to the center-of-mass energy, $\xi_\mathrm{X}$ and $\xi_\mathrm{Y}$, are given by

$$\xi_\mathrm{X} = \frac{M_\mathrm{X}^2}{s} \quad \xi_\mathrm{Y} = \frac{M_\mathrm{Y}^2}{s} \,, \tag{6.1}$$

and $\xi$ can then be defined as

$$\xi = \max(\xi_\mathrm{X}, \xi_\mathrm{Y}) \,. \tag{6.2}$$

Figure 6.3 shows the efficiency of the detector level event selection as a function of the $\xi$-value of the event. Aiming for an optimal compromise between high visible phase space and optimal selection efficiency, a cut on $\xi > 10^{-6}$ is chosen. This corresponds to a model dependent selection efficiency which is on average $97.3 \pm 1.6\,\%$. Events that are selected on the generator level but not on detector level are called *missed* events, events that are seen on detector level but do not fulfill the required generator selection are called *fake* events. The amount of these events is model dependent and is summarized in Tab. 6.1.

The total energy in the CASTOR acceptance on particle level is calculated by summing up the energy of all particles, except muons and neutrinos, within the acceptance of CASTOR

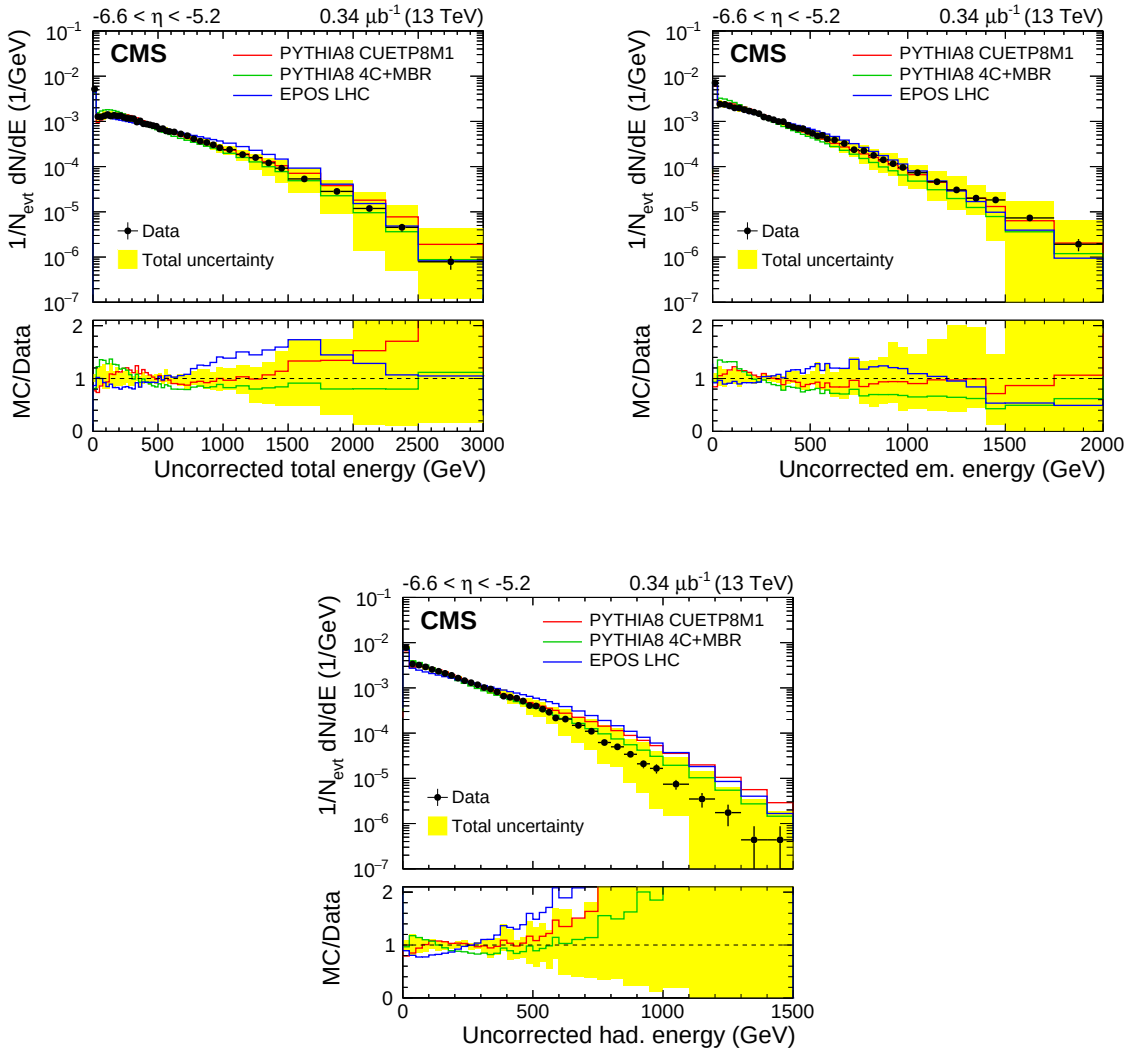



**Fig. 6.1.:** Spectra of the energy reconstructed in CASTOR, normalized to the number of events that pass the offline event selection, compared to the detector level predictions of various event generators. The total energy spectrum is shown in the top left panel, the electromagnetic in the top right, and the hadronic in the bottom panel. Statistical (systematic) uncertainties are shown with error bars (yellow band). Published in Ref. [118].

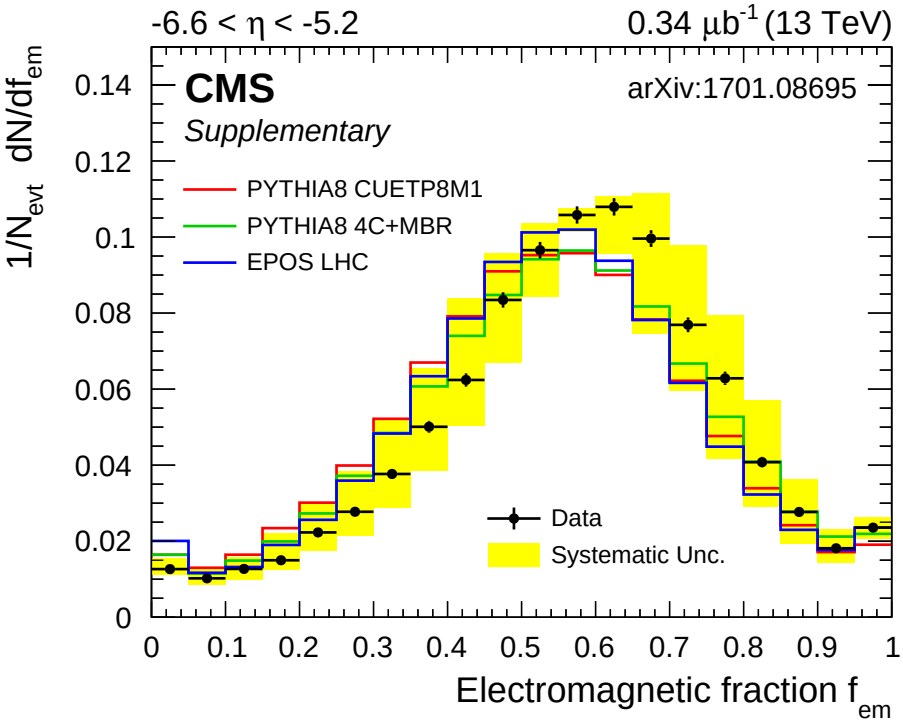



**Fig. 6.2.:** The distribution of the ratio of electromagnetic to total reconstructed energy in CASTOR. The error bars denote statistical uncertainties, while the yellow band indicates the systematic uncertainty. Published as supplement material for Ref. [118].

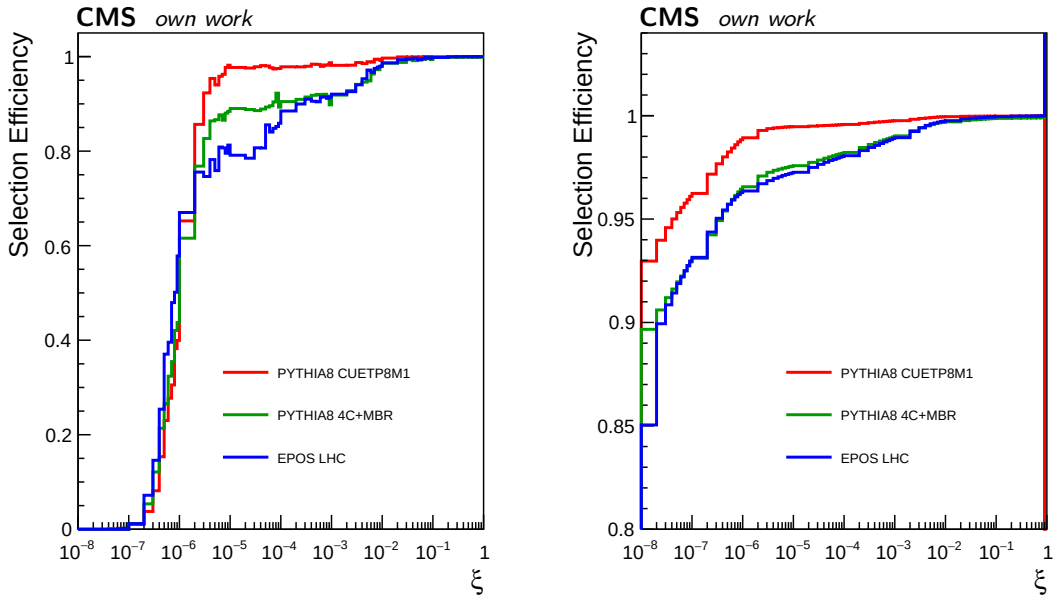



**Fig. 6.3.:** Efficiency of the offline event selection for events based on their value of $\xi$, on the left in differential and on the right panel in integrated representation.

at pseudorapidities $-6.6 < \eta < -5.2$. Muons and neutrinos are excluded since they do not deposit visible energies in the calorimeter. For the electromagnetic spectrum, only energies of electrons and photons are summed up, whereas they are excluded for the hadronic energy. Neutral pions are not considered stable in this analysis, thus their decay photons contribute to the electromagnetic spectrum.

Distributions of the reconstructed energy in CASTOR as a function of the true energy on particle level are constructed for the PYTHIA 8 Tunes CUETP8M1 and 4C+MBR as well as for EPOS LHC. The three compositions of total, electromagnetic, and hadronic energies are shown independently in Fig. 6.4. These distributions will serve as reference matrices during the unfolding procedure. Some important detector properties can be identified in these distributions:

- On average about 60% of the total particle level energy is reconstructed in CASTOR. This is due to the non-compensating properties was was also identified in Fig. 5.3.
- The detector response to electrons is per definition equal to unity both in data as well as in simulations. This is visible in the distribution of the electromagnetic energy, which is located on the diagonal.
- The model differences in the detector descriptions are small.

### 6.1.2 Correction for experimental effects

In a mathematical representation, the spectra can be written as vectors of order $n$ where each element represents the number of events in a certain range in energy. We write $\vec{E'}$ for the energy spectrum on detector level and $\vec{E}$ for the spectrum on particle level, $\vec{m}$ and $\vec{f}$ for the missed and fake events. The general problem to correct the measured distribution to the particle level, is to invert and solve the equation

$$\vec{E'} = \mathbf{S} \times \left(\vec{E} - \vec{f}\right) + \vec{m} \; . \tag{6.3}$$

Here, the matrix $\mathbf{S}$ represents the smearing of the measurement due to detector effects. This information is provided by Monte Carlo simulations and is illustrated by Fig. 6.4. In the presented case, the inversion of this problem is not possible, since the matrices $\mathbf{S}$ can not be numerically inverted in a precise way. Therefore, a regularized unfolding procedure is developed using the so-called Bayesian iterative method with early stopping as it is implemented in the ROOUNFOLD package [122]. It exploits an algorithm proposed by d'Agostini [123].

The cells of the smearing matrices $\mathbf{S}$ can be interpreted as probabilities $P\left(E_i|E'_j\right)$ to obtain a certain energy $E'$ on detector level caused by a certain energy $E$ on particle level. For

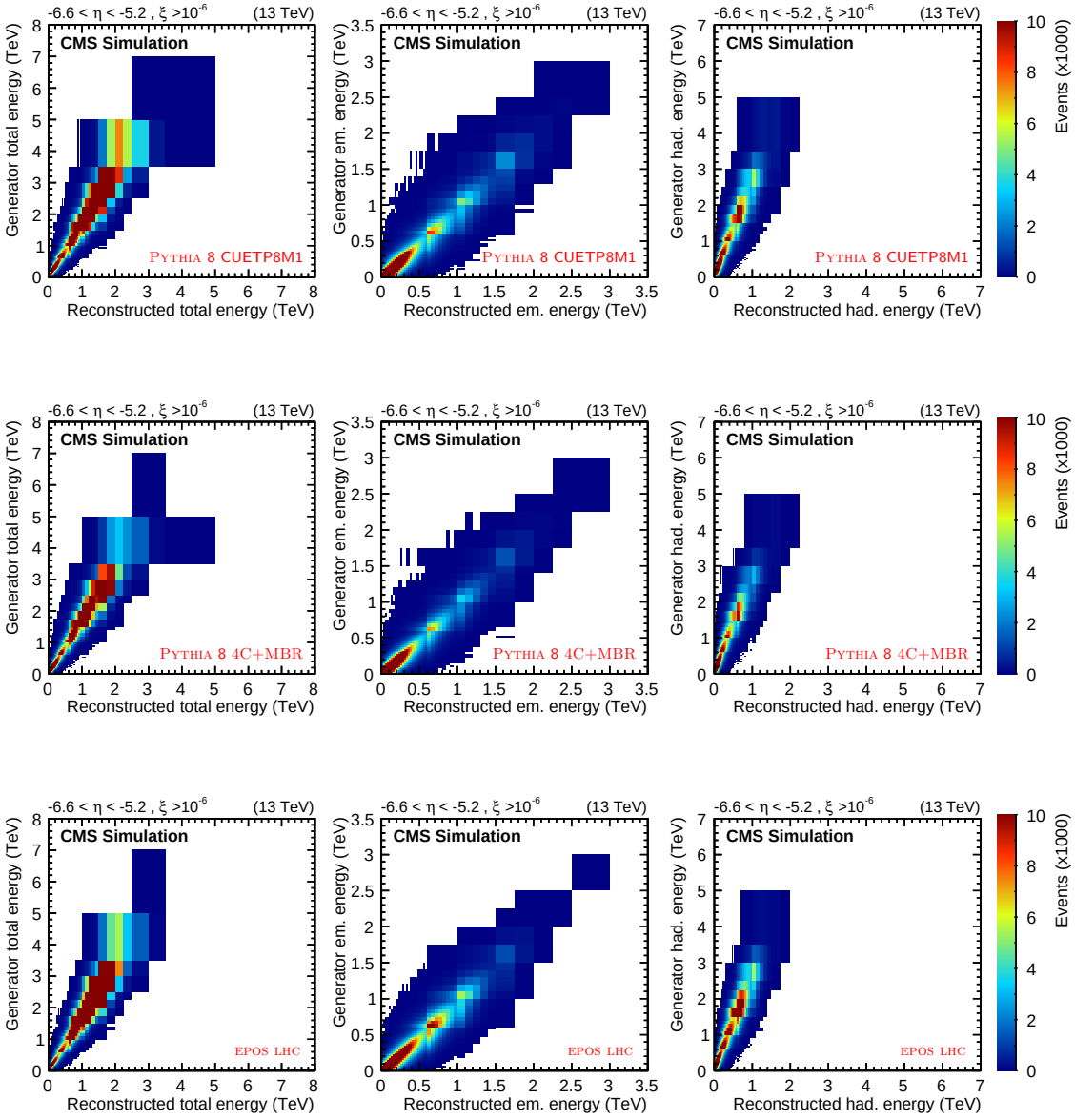



**Fig. 6.4.:** Distributions of reconstructed energy with respect to the particle level energy for PYTHIA 8 CUETP8M1 (top), PYTHIA 8 4C+MBR (middle), and EPOS LHC (bottom row). The distributions are shown for the total (left), electromagnetic (middle), and hadronic energy in CASTOR (right column). The color indicates the number of events. The event selection is explained in the text.

this, the distributions shown in Fig. 6.4 are normalized to 1 in bins of the detector level energy, leading to:

$$\sum_{i}^{n} P\left(E_i|E'_j\right) = 1 \,. \tag{6.4}$$

Assuming a known initial probability distribution of energies at particle level $P_0\left(\vec{E}\right)$ and applying Bayes' theorem to this case, one can rewrite $P\left(E_i|E'_j\right)$ as

$$P\left(E_i|E'_j\right) = \frac{P\left(E'_j|E_i\right) P_0\left(E_i\right)}{\sum_{l=1}^{n} P\left(E_j|E'_l\right) P_0\left(E_l\right)} \tag{6.5}$$

The distribution of events on particle level can now be calculated with

$$E_i = \sum_{j_1}^{n} E'_j P\left(E_i|E'_j\right) \tag{6.6}$$

and including the correction for fake and missed events

$$E_i = \left(\sum_{j_1}^{n} \left(E'_j - f_j\right) P\left(E_i|E'_j\right)\right) + m_i \,. \tag{6.7}$$

After the first iteration the obtained result depends on the initial assumption on the true probability distribution $\vec{P_0}$. If it is not consistent with the input data, there will be a disagreement between the unfolded result and the initial guess. Therefore, in a second iteration, $P_0$ is replaced by the normalized distribution $\vec{P_1} = 1/N\vec{E}$. From Eq. (6.7) it is also obvious that in the first iteration, this methods is equivalent to simple correction factors. This way, the expected true distribution on particle level will be approached iteratively. It is a feature of iterative d'Agostini unfolding that for very large number of iterations the method approaches the maximum likelihood estimator of matrix inversion. This is characterized by a small bias but large fluctuations.

The optimal number of iterations has to be determined for every application. In the present case this is done using evolution curves of the reduced $\chi^2$ value between:

- an unfolded independent Monte Carlo sub sample and the corresponding generator truth and
- between the untreated data and the unfolded data that is afterwards smeared back to the detector level.

These curves are shown in Fig. 6.5 for PYTHIA 8 CUETP8M1. The chosen number of iterations is marked with a red dashed line and is 4 (6, 5) for the total (electromagnetic, hadronic) spectrum . The final choice is based on the second mentioned comparison

between the unfolded and smeared data and the original data to avoid Monte Carlo biases. As the final selection criteria, the point is chosen, where either the reduced $\chi^2$ drops below 1 or the change $\Delta\chi^2$ to the previous iteration is below 0.1. The reduced $\chi^2$ is calculated with the diagonal elements of the covariance matrix that is calculated by ROOUNFOLD.

Furthermore, corrections for selection inefficiencies need to be applied. This is done during the unfolding procedure. From Monte Carlo simulations, spectra of the missed and fake events are obtained. The spectra of missed and fake events are shown relative to the total number of events in Figs. 6.6 and 6.7. The spectrum of fake events (scaled by the actual size of the dataset) is subtracted from the data before the unfolding procedure. Correspondingly, the scaled spectrum of missed events is added after the unfolding. This correction is certainly model dependent, but within a few percent and therefore small compared to other uncertainties.

Standard *Closure tests* are performed to validate the procedure. First, an independent detector level distribution from Monte Carlo simulations is unfolded and compared to the corresponding generator level truth. Additionally, the unfolded and afterwards back-smeared data is compared to the untreated data. Results are shown as ratios in Figs. 6.8. All tests show a good agreement within the statistical uncertainties. In the first case, this shows that the unfolding procedure correctly corrects for detector effects in simulations and reproduces the underlying generator level truth. The second study shows that the unfolding does not change the information content of the data in any way.

As a further crosscheck of the unfolding procedure, a bottom-line test is performed. The unfolding is required not to increase the significance of a measurement. The described mathematical operations during the unfolding can not add any information to the data, their discriminative power can only be reduced. As a consequence, the reduced $\chi^2$ between the data and model predictions is expected to be larger in the smeared than in the unfolded space. The reduced $\chi^2$ values of the data compared to various model descriptions are calculated once in the unfolded and once in the smeared space. The results are shown in table 6.2. The majority of the studied models show the expected result.

The unfolding procedure induces correlations between neighboring bins. Since the detector smearing is rather large for CASTOR, bin migrations are large, leading to larger correlations which can be both positive and negative. Therefore, uncertainties need to be calculated with the full covariance matrices of the unfolded results. They are shown in Fig. 6.9.

### 6.1.3 Systematic uncertainties

Various sources of systematic uncertainties are studied and their effect on the final result is calculated:

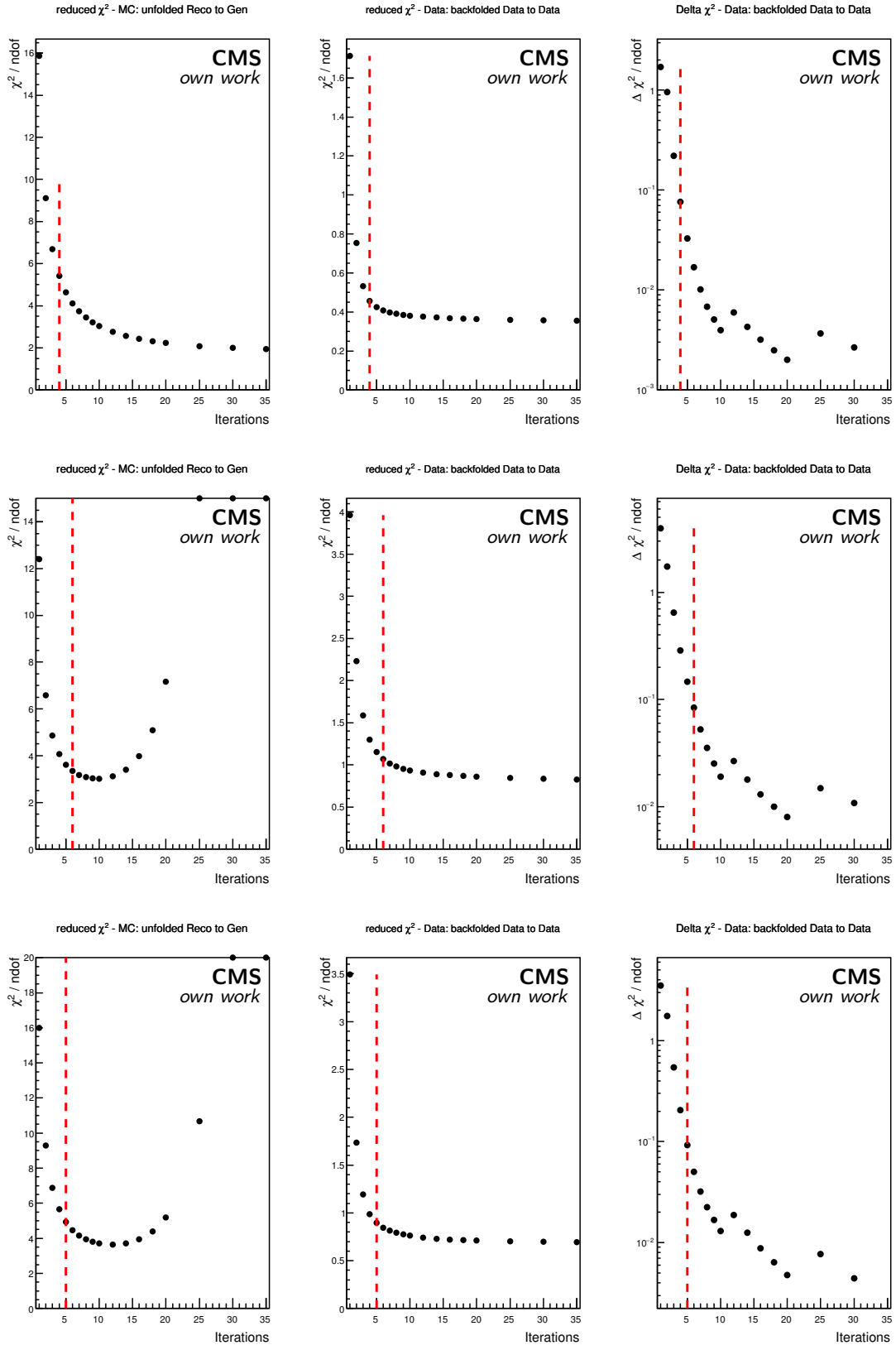



**Fig. 6.5.:** Evolution of $\chi^2$ depending on the number of iterations for the total (top), electromagnetic (middle), and hadronic (bottom) energy spectrum. 4 (6, 5) Iterations are chosen for the final unfolding (red dashed line).

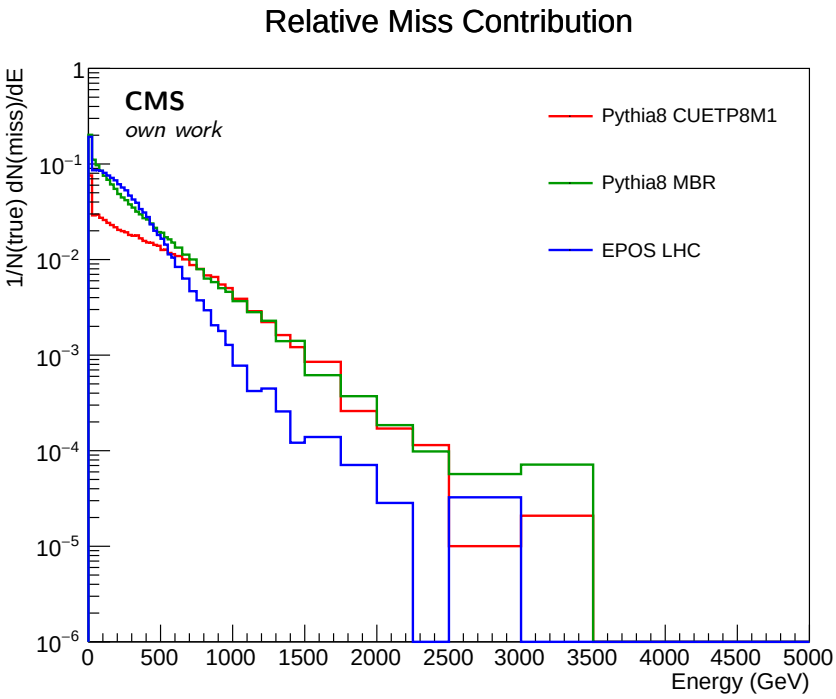



**Fig. 6.6.:** Contribution of missed events to the unfolding procedure. The distribution is scaled to the total number of events selected on particle level.

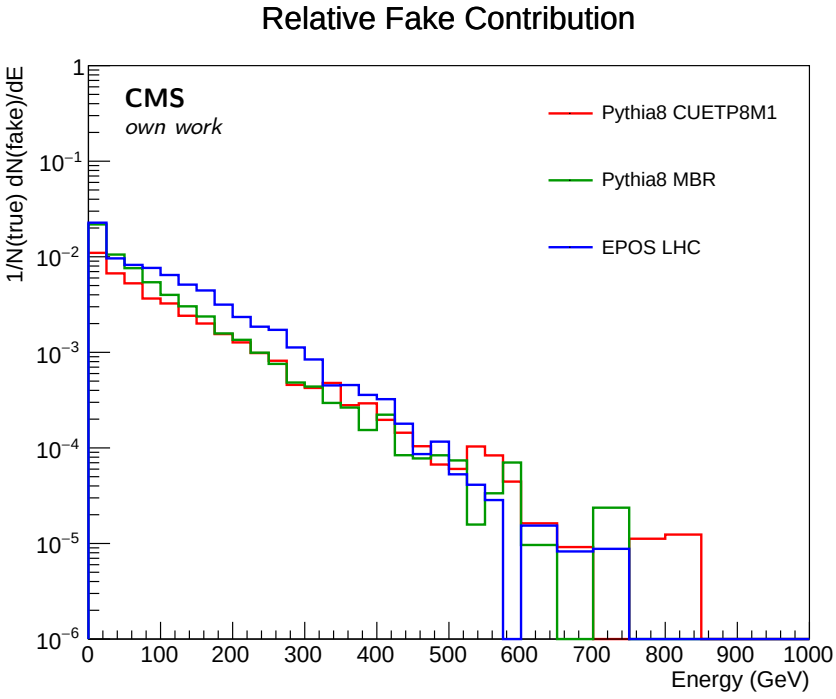



**Fig. 6.7.:** Contribution of fake events to the unfolding procedure. The distribution is scaled to the total number of events selected on detector level.

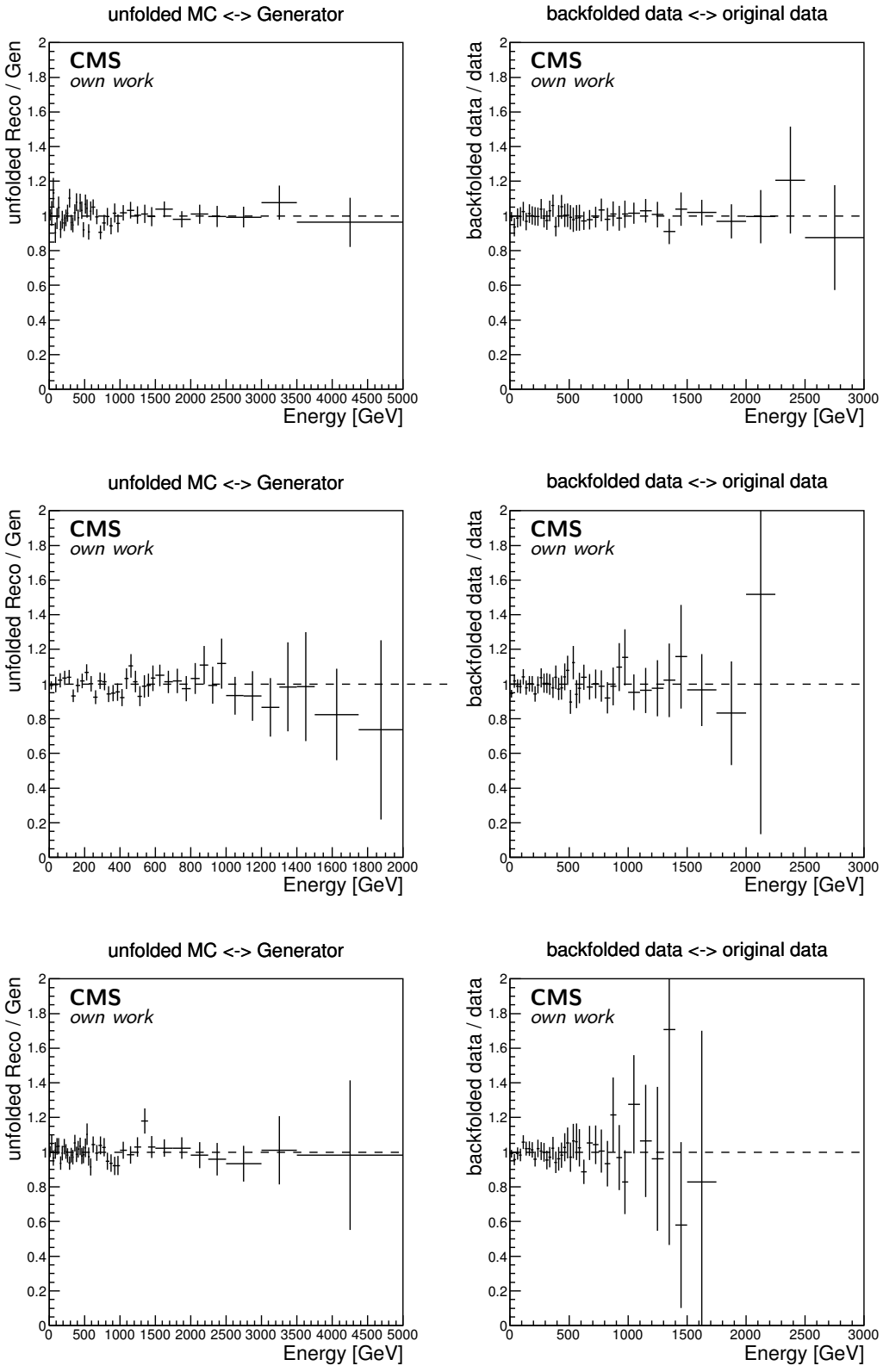



**Fig. 6.8.:** Closure test of the unfolding procedure for the total (top), electromagnetic (middle), and hadronic (bottom) energy spectrum.

**Tab. 6.2.:** $\chi^2/ndof$ calculated between the data and the model predictions in the unfolded and smeared space.

| $\chi^2$/ndof | unfolded space | smeared space |
|---|---|---|
| EPOS LHC | 7.897 | 64.4343 |
| EPOS 1.99 | 7.45962 | 22.9681 |
| QGSJETII.03 | 38.711 | 140.859 |
| QGSJETII.04 | 13.1669 | 36.3026 |
| SIBYLL 2.1 | 40.435 | 35.5441 |
| SIBYLL 2.3 | 23.4733 | 55.3507 |
| PYTHIA 8 CUETP8M1 | 13.1692 | 13.9723 |
| PYTHIA 8 4C+MBR | 25.6856 | 38.9246 |

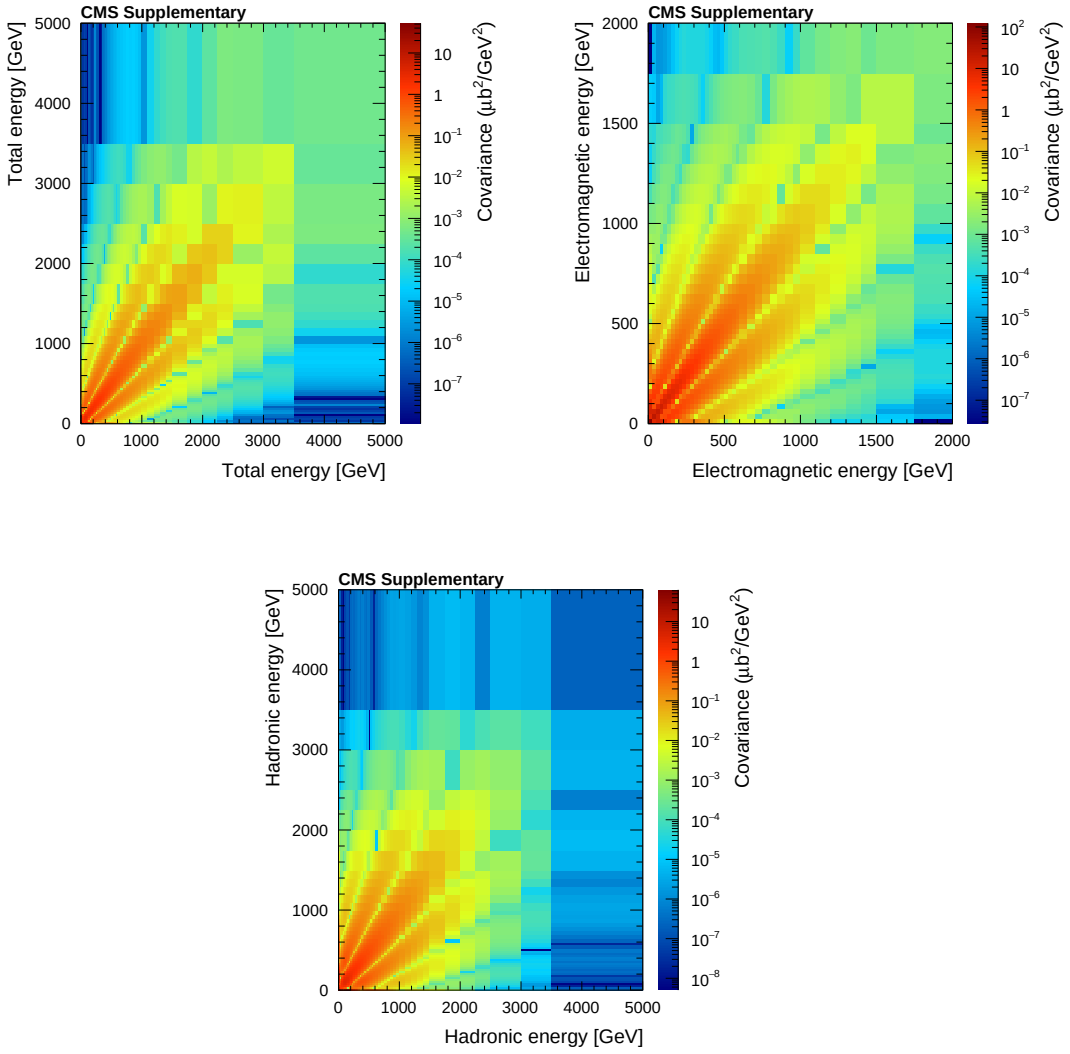



**Fig. 6.9.:** Covariance matrices of the unfolded spectra. Top left for the total spectrum, top right for the electromagnetic and bottom center for the hadronic spectrum. Published as supplement material for Ref. [118].

**CASTOR energy scale**

The uncertainty of the energy scale is the dominant systematic uncertainty. The different contributions are listed in table 6.3. Together, they lead to an overall uncertainty of 17%. Thus, the measured energy spectrum on detector level is scaled both up and down by 17% and the resulting spectra are unfolded using the same unfolding procedure as for the central spectrum. The differences between the resulting spectra are taken as the systematic uncertainty of the result. The effect of the energy scale uncertainty is different for different bins in the spectrum. Furthermore, as a result of the normalization, the uncertainty becomes largest at very low and very high energies, and very small at intermediate energies.

**Model dependence**

Since the unfolding relies on Monte Carlo simulations, three models are used to unfold the detector level spectra, namely PYTHIA 8 CUETP8M1, PYTHIA 8 4C+MBR, and EPOS LHC. The average of the three resulting spectra is used as the result and the largest difference as a systematic uncertainty. Over the whole total energy spectrum, the average uncertainty is $5.1\%$, varying from $0.6$ to $17\%$. For the electromagnetic and hadronic spectrum, the model dependence is slightly higher with $6.7\%$ and $8.7\%$ respectively. This effect is partially also related to limited MC statistics.

**Event selection**

The effect of a variation of the HF tower threshold in the event selection is also studied. It is observed, that with lower (or higher) threshold, the event selection is more (less) sensitive to events with low energy in the forward region, resulting in a larger (smaller) fraction of events with no or few energy in CASTOR. Therefore, this effect will be accounted for in the systematic uncertainties. Since the energy scale uncertainty of the HF calorimeter is 10%, a variation to 4.5 and 5.5 GeV is chosen. The residual sensitivity to noise or beam induced backgrounds in the event selection are well below 1%.

**Pileup**

The average pileup of the analyzed data is of the order of 6%. This means that more than one interaction occurred in only 3% of all selected events. Additional checks are performed to investigate the effect of pileup events. A rejection method based on vertex reconstruction is applied and it is found that the data do not show any noticeable deviation from Monte Carlo simulations without pileup. As an additional crosscheck, data with higher interaction probability are investigated in a similar way and corrected for pileup. In these studies no indication is found that the spectra are affected in a statistically meaningful way. Therefore, the remaining contribution of pileup events to the measured spectra is considered negligible, especially compared to other uncertainties.

All contributions to the systematic uncertainty are treated in the same way: the detector level spectra, varied by each systematic uncertainty, are unfolded and for every bin, the

**Tab. 6.3.:** Components of the systematic uncertainty on the CASTOR energy scale

| | |
|---|---|
| HF energy scale | 10% |
| Extrapolation + model dependence | 10% |
| CASTOR non-compensation | 5% |
| CASTOR position | 7.5 % |
| Total | 17 % |

**Tab. 6.4.:** Uncertainties on the differential cross sections at a few selected values of the total, electromagnetic, and hadronic energies.

| | Total | | Electromagnetic | | Hadronic | |
|---|---|---|---|---|---|---|
| | 300 GeV | 3000 GeV | 300 GeV | 1200 GeV | 300 GeV | 2000 GeV |
| Energy Scale | $^{+17\,\%}_{-14\,\%}$ | $^{+94\,\%}_{-77\,\%}$ | $^{+5.9\,\%}_{-2.1\,\%}$ | $^{+93\,\%}_{-65\,\%}$ | $^{+11\,\%}_{-10\,\%}$ | $^{+169\,\%}_{-80\,\%}$ |
| Unfolding | $\pm 5.8\%$ | $\pm 6.4\%$ | $\pm 5.2\%$ | $\pm 4.1\%$ | $\pm 6.9\%$ | $\pm 17\%$ |
| Event selection | $\pm 0.5\%$ | $<0.01\%$ | $\pm 0.14\%$ | $<0.01\%$ | $\pm 0.06\%$ | $<0.01\%$ |
| Luminosity | $\pm 2.6\%$ | | | | | |
| Statistical | $\pm 1.2\%$ | $\pm 4.3\%$ | $\pm 1.5\%$ | $\pm 5.9\%$ | $\pm 1.0\%$ | $\pm 4.2\%$ |

spread of these unfolded spectra is used to define the systematic uncertainty on the result due to this contribution. The different contributions are added in quadrature for every bin independently. Example values at some selected energies are given in Table 6.4.

### 6.1.4 Results

The total, electromagnetic, and hadronic energy spectra are measured in the region $-6.6 < \eta < -5.2$ and corrected to the particle level for events with a fractional proton momentum loss of $\xi > 10^{-6}$. The results are shown in Figs. 6.10–6.12 as they are published in Ref. [118] and compared to the predictions of EPOS, QGSJETII and SIBYLL (left plots) and various PYTHIA 8 tunes (right plots). The same data is shown in Fig. C.1 in appendix C.1 with a linear scale and zoomed into the low energy region. All spectra feature a sharp peak at zero. This peak is due to events with large gaps in the acceptance of CASTOR. This peak is therefore sensitive to the presence and amount of diffractive or otherwise highly elastic events. The total and hadronic energy spectra exhibit further peaks at about 300 and 100 GeV respectively, followed by a long tail towards higher energies. Such a peak is not observed in the electromagnetic spectrum. Therefore, the structure seen in the total spectrum can be attributed to the hadronic component.

In Fig. 6.10, the differential cross section of the total energy is shown. The models considered perform differently in reproducing the observed spectrum, but none of the models reproduce all features of the data. The bump at about 300 GeV is visible in all of them. The spectrum is best described by EPOS LHC and QGSJETII.04. The tunes based on the PYTHIA 8 model tend to overestimate the contribution of the soft part of the spectrum.

The same trend is observed for SIBYLL 2.3. The high energy tail on the other hand is very well described by PYTHIA 8 and SIBYLL, whereas EPOS LHC and QGSJETII.04 overestimate the region between 1 and 2.5 TeV. It can also be observed that the spectra are very sensitive to the scaling parameter $p_{\mathrm{T},0}^{\mathrm{ref}}$ of PYTHIA 8, which is an important parameter to describe multiparton interactions.

The electromagnetic spectrum is shown in Fig. 6.11. The data is relatively well described by most of the models within uncertainties. PYTHIA 8 4C+MBR and SIBYLL 2.3 fail to correctly model the shape of the soft part of the spectrum up to about 500 GeV. The electromagnetic energy distribution is found to be also very sensitive to the underlying modeling of MPI in PYTHIA 8.

Figure 6.12 shows the hadronic energy distribution. While EPOS LHC and QGSJETII perform well at lower energies, they predict too large a cross section in the range of 600 to 1800 GeV. This feature is also observed in the total energy spectrum, suggesting that the excess originates from the production of charged hadrons. SIBYLL 2.3 on the other hand reproduces the slope of the spectrum over a larger energy range, but significantly overestimates the cross section at very low energy, while SIBYLL 2.1 shows a large excess at around 500 GeV, similar to that observed in the total energy spectrum.

As mentioned earlier, the mechanisms of the underlying event are expected to be a dominant contribution at forward rapidities in proton-proton collisions. The present measurements confirm that the modeling of multiparton interactions has a significant impact. PYTHIA 8 CUETP8M1 without MPI is ruled out by the data, which exhibit much harder spectra than predicted by the model. The shape of the spectra are moreover significantly influenced by the MPI-related settings in PYTHIA 8. The present results can therefore contribute to improvements in future Monte Carlo parameter tunes.

The measurement of the CASTOR energy spectra provides detailed information on the energy distribution in the very forward direction. The average of the total energy is included in a different measurement by CMS using the same data and particle level definition. The acceptance of the HF calorimeters is extended by CASTOR and the average energy density $\mathrm{d}E/\mathrm{d}\eta$ is studied as a function of pseudorapidity in the acceptance $3.15 < |\eta| < 6.6$ [119]. The results are shown in Fig. 6.13. They provide complementary information on the energy production in the forward direction on a more general level.

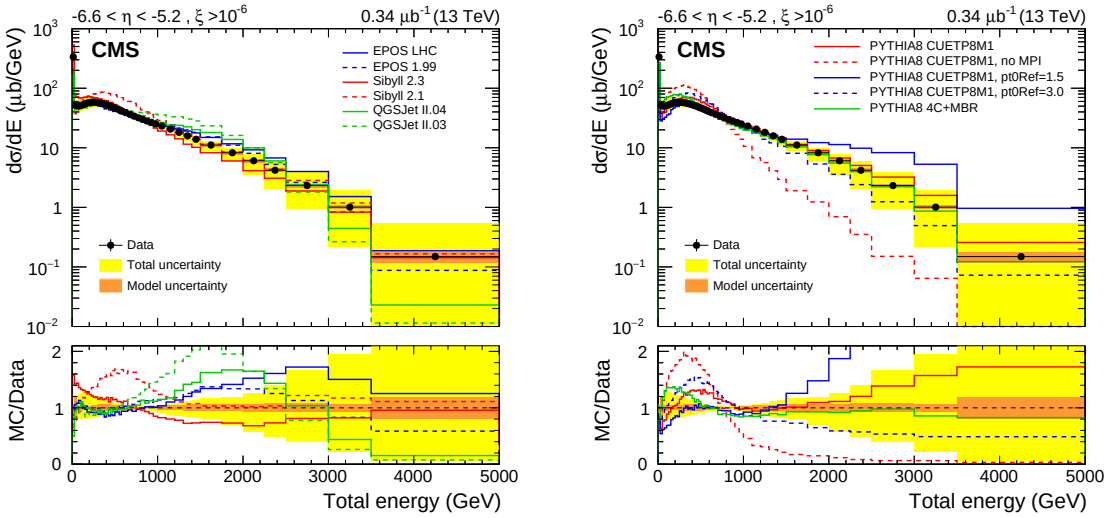



**Fig. 6.10.:** Differential cross section as a function of the total energy in the region $-6.6 < \eta < -5.2$ for events with $\xi > 10^{-6}$. The left panel shows the data compared to MC event generators mostly developed for cosmic ray induced air showers, and the right panel to different PYTHIA 8 tunes. Published in Ref. [118].

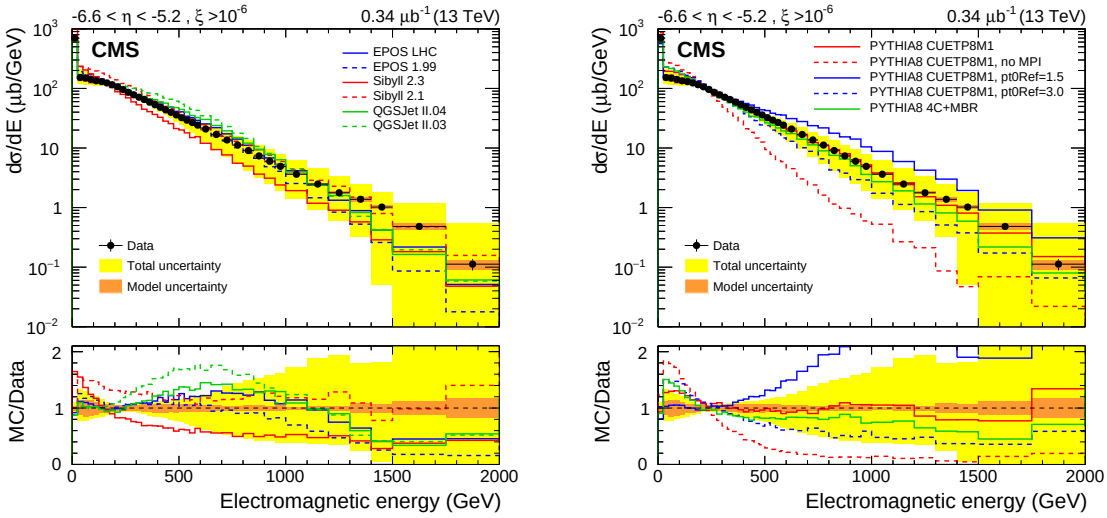



**Fig. 6.11.:** Differential cross section as a function of the electromagnetic energy in the region $-6.6 < \eta < -5.2$ for events with $\xi > 10^{-6}$. The left panel shows the data compared to MC event generators mostly developed for cosmic ray induced air showers, and the right panel to different PYTHIA 8 tunes. Published in Ref. [118].

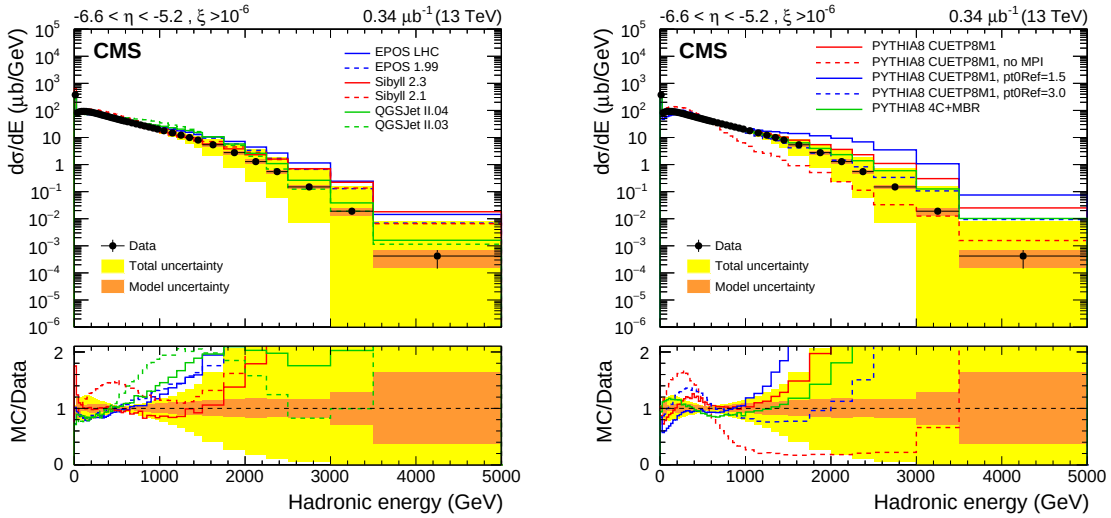



**Fig. 6.12.:** Differential cross section as a function of the hadronic energy in the region $-6.6 < \eta < -5.2$ for events with $\xi > 10^{-6}$. The left panel shows the data compared to MC event generators mostly developed for cosmic ray induced air showers, and the right panel to different PYTHIA 8 tunes. Published in Ref. [118].

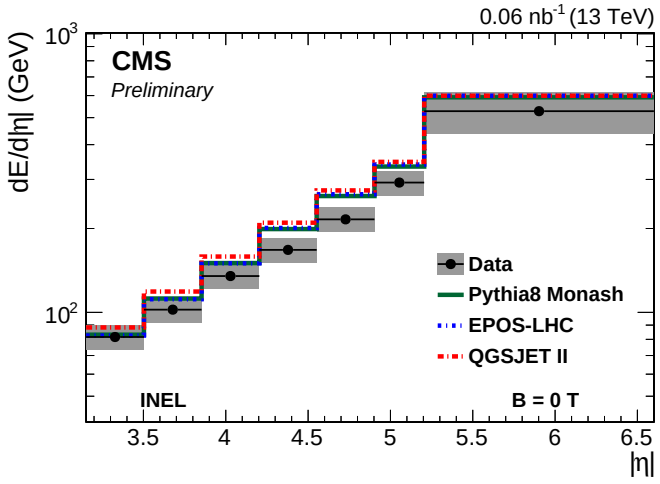



**Fig. 6.13.:** Energy density $\mathrm{d}E/\mathrm{d}\eta$ in the forward direction measured with HF and CASTOR at $\sqrt{s}$=13 TeV. Published in Ref. [119].

## 6.2 Multiplicity dependent energy

A natural extension to the presented inclusive measurements is to study the forward particle production in relation to the collision properties at central rapidities. This is especially interesting in the context of the underlying event mechanisms since they both affect the forward and central particle production at the same time. This was already shown at lower center-of-mass energies [57]. There, the average energy deposit per proton-proton collision in the CASTOR acceptance was studied as a function of the transverse momentum of the hardest jet within $|\eta| < 2$. A comparable study was now performed in the frame of this work with the data recorded with CASTOR at 13 TeV center-of-mass energy. Due to the lack of a magnetic field, no information on the transverse momenta of particles or jets in the central region is available. Therefore the multiplicity of charged particles $N_{\mathrm{ch}}$ with $|\eta| < 2$ was chosen in order to quantify the collision activity in the central region. This is often referred to as a measure of the centrality or the hardness of a proton-proton collision. In the following, the average energy deposit in CASTOR is studied as a function of $N_{\mathrm{ch}}$.

### 6.2.1 Tracking without magnetic field

Due to the lack of a magnetic field, particle tracks in the central tracker of CMS are not bend but are simple straight lines. Therefore it is not only impossible to measure the particle's transverse momenta but also the standard track reconstruction fails. The reconstruction software assumes and requires curved tracks in order to identify them properly. This arises the need for a special tracking algorithm that was developed for the measurement of the charged hadron multiplicity [120] and was kindly provided for this work.

The CMS pixel tracker provides information on where charged particles traverse the detector layers. These three dimensional points are used to reconstruct straight lines. Since the combinatorial background is large, strict quality criteria are applied to clean the ensemble of possible tracks. The first and most strict is that at least three pixel hits are required to lie with a cone of radius $R = \sqrt{\Delta\phi^2 + \Delta\eta^2} < 0.02$ to the reconstructed line. The efficiency to find more than two hits in the pixel detector drops quickly for $|\eta| > 2$. Therefore the tracking range is restricted to $|\eta| < 2$. Furthermore only those tracks are kept that cross others in a region close to the expected interaction point. This means that the tracks are required to form at least one interaction vertex in the center of the detector volume. Two examples of single reconstructed events are shown in Fig. 6.14. Note that 3D projection effects significantly contribute to the visual impression of the tracks. This method has been crosschecked intensively and was validated against Monte Carlo simulations based on PYTHIA 8 tune CUETP8M1. Details can be found in appendix C.2.

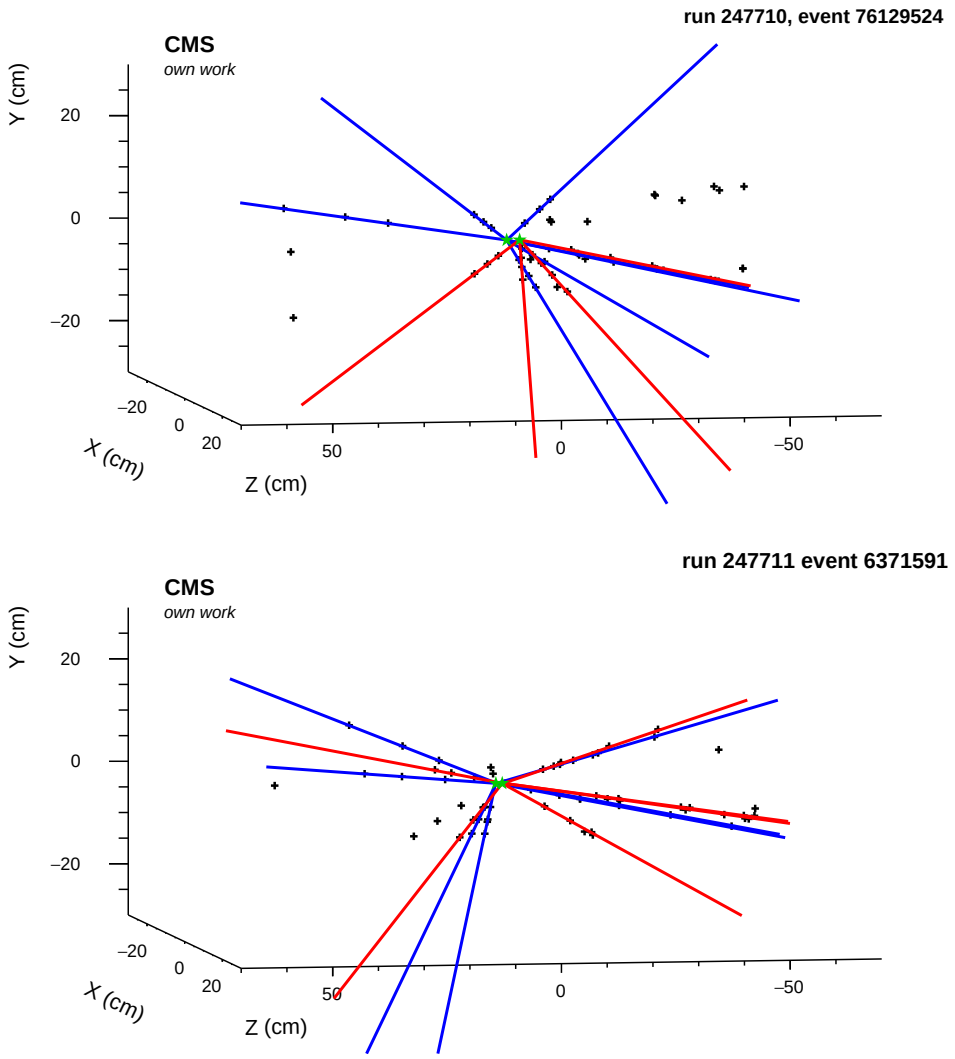



**Fig. 6.14.:** Visualizations of two example events with two reconstructed vertices and 11 (top) and 13 (bottom) reconstructed tracks. The positions of the pixel hits are illustrated with black crosses and the tracks with red and blue lines. The color indicates to which vertex the tracks are assigned. The green stars illustrate the location of the reconstructed vertices at a distance of 2.8 cm (top) and 1.4 cm (bottom).

In addition to the validation of general tracking performance the track reconstruction efficiency is determined with Monte Carlo simulations. The distance between the reconstructed tracks and the charged particles on the generator level is evaluated in terms of $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$. This is shown in Fig.6.15 for all possible combinations of tracks and particles. The distance between reconstructed tracks and the particles they are originating from is small and is visible in the form of a peak at low $\Delta R$. For larger distances the combinatorial background is dominant. The transition between the actual tracking and the background takes place at about $\Delta R \approx 0.06$.

The tracking efficiency can therefore be calculated as the probability to find a reconstructed track within a distance of $\Delta R < 0.06$ to a charged stable particle. The efficiency is evaluated as a function of the transverse momentum of the generator level particle and is shown in Fig. 6.16. It is observed that the tracking efficiency rises sharply with $p_{\mathrm{T}}$ and reaches a plateau of about 80% for values of $p_{\mathrm{T}}$ larger than 200 MeV. Complementary, the tracking fake rate can be estimated by calculating the probability to find a reconstructed track with no stable particle being present within the chosen radius. This is evaluated in Fig. 6.17 as a function of $|\eta|$. Here values between 5 and 12% are obtained. The calculated efficiency and fake rate is comparable to those obtained with standard tracking algorithms [124], especially bearing in mind that only the pixel tracker is used here.[1] As a conclusion, the tracking is considered efficient for particles with a transverse momentum $p_{\mathrm{T}}$ larger 200 MeV and absolute pseudorapidity smaller 2. Therefore, all reconstructed tracks with $|\eta| < 2$ are considered for the detector level observables, while on particle level, only charged particles with $p_{\mathrm{T}} > 200\,\mathrm{MeV}$ and $|\eta| < 2$ are used to calculate the particle multiplicities.

[1] While typical tracking efficiencies reach 100% for isolated muons, the average efficiencies for all particles are significantly reduced, especially at $p_{\mathrm{T}}$ <1 GeV. Moreover the lack of a magnetic field increases the number of particles with very small transverse momenta which suffer from multiple scattering thus reducing the overall efficiency and increase the fake rate due to additional hits in the pixel detector that are incompatible with straight lines.

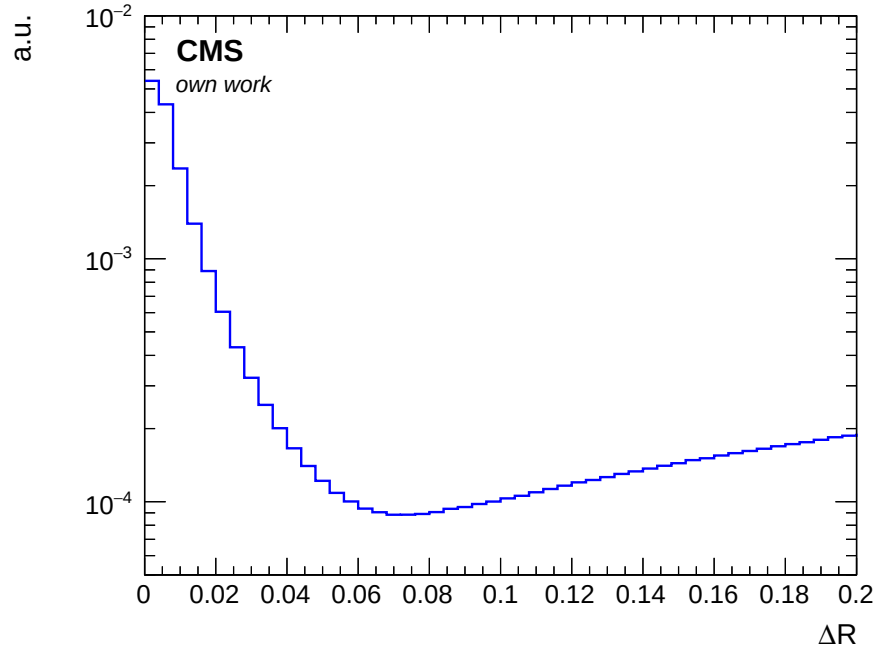



**Fig. 6.15.:** Distance of reconstructed tracks and generated charged particles in units of $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$. The peak at small values of $\Delta R$ consists of correctly matched particles within the reconstruction uncertainty while the combinatorial background is dominant for larger values of $\Delta R$.

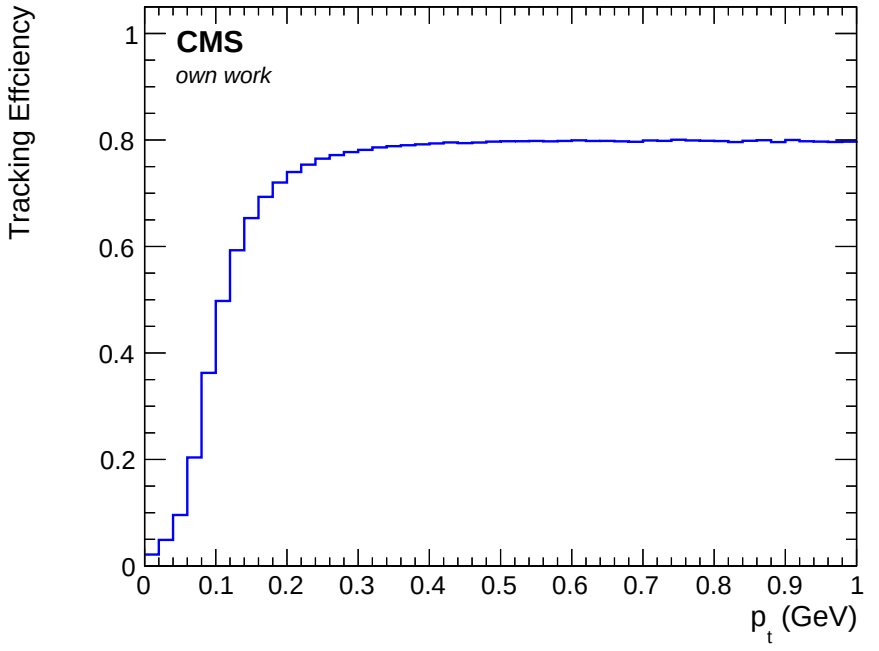



**Fig. 6.16.:** Efficiency of the tracking algorithm as a function of the particle transverse momentum $p_{\mathrm{T}}$. The reconstruction becomes efficient to about 80% for particles with $p_{\mathrm{T}} > 200\,\mathrm{MeV}$.

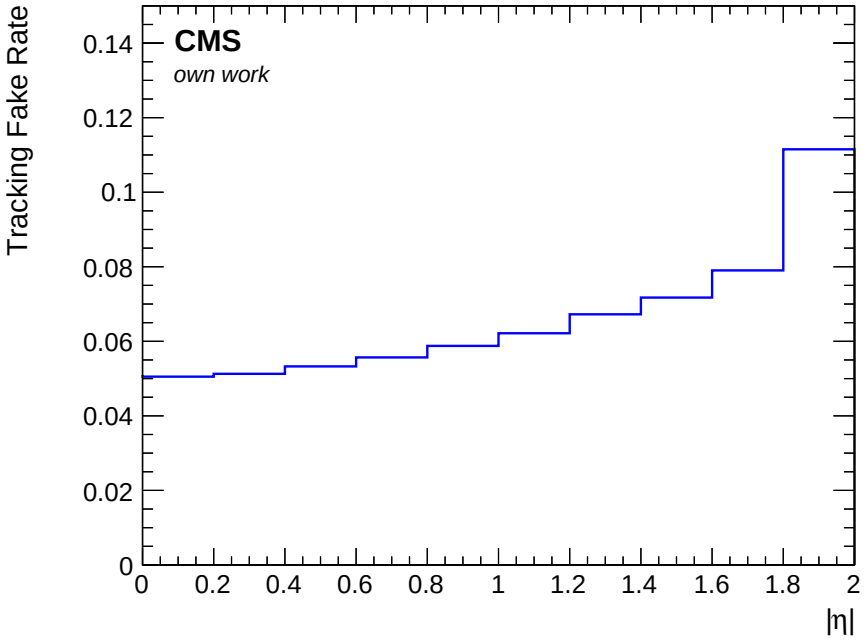



**Fig. 6.17.:** Fake rate (fraction of reconstructed tracks that are not associated with a nearby charged particle) of the tracking algorithm as a function of $|\eta|$.

### 6.2.2 Trigger and Event Selection

While the CASTOR energy spectra are measured at the lowest possible beam intensity in order to minimize pileup related effects, this also means that the available statistics is very small. A larger dataset is needed in order to extend the measurement in a statistically meaningful way to events with higher multiplicities. For this purpose data from CMS runs 247920 and 247934 are used also requiring a ZeroBias trigger. The total integrated luminosity of the data is about 56 $\mu\mathrm{b}^{-1}$. This is about 160 times the statistics as available for the energy spectra. The event selection is adapted from the measurement of the CASTOR energy spectra. On detector level, collision events are selected by requiring activity in the HF calorimeters with a threshold of 5 GeV per tower. On top of this, at least one reconstructed particle track is required. These criteria are applied for data as well as simulations. The intensity of the proton beams was higher than for the run used for the energy spectra and the average interaction probability was about 30%. This also increases the probability of recording two simultaneous collisions. In order to reject these pileup events, an additional cut on the reconstructed tracks and vertices is developed.

It is observed in Monte Carlo simulations that even if there is only one proton-proton interaction, sometimes two vertices are reconstructed. These so-called *split-vertices* occur especially in events with high multiplicity. These two vertices are located closely together in the z-direction. This is illustrated in Fig. 6.18, where the distance of two vertices in

the z-direction is shown. A very similar distribution is found for data with low interaction probability, while data with higher interaction probability also shows a significant tail with larger distances. It can therefore be assumed that events with two reconstructed vertices that are close together mostly originate from one single collision which gets split up by the vertexing algorithm. Since this happens mostly for high multiplicity events, these events should not be rejected. Therefore only events with more than two reconstructed vertices are rejected, while events with exactly two vertices are kept if those are separated by less that 0.5 cm. The cut value of 0.5 cm is derived from Fig. 6.18 and is verified with data. The track multiplicity distribution is shown in Fig. 6.19 for simulations as well as data with and without the explained vertex cut. It can be observed that the tail of high multiplicity is suppressed by the vertex cut towards a level that agrees well with simulations. Since there are significant model differences here, a perfect agreement is not required but still the obtained level of agreement justifies the use of this kind of vertex selection. A second crosscheck is done by studying the spectrum of the total energy in CASTOR as shown in Fig. 6.20 for data taken in different runs. The data from run 247324 was used for the previously presented measurement of the CASTOR energy spectrum and serves here as a reference. It can be seen that by applying the vertex selection the tail of high energetic events in the high pileup data also gets suppressed towards the same level as found in the the low pileup data. It can therefore be concluded that the selection criteria on the number and distribution of reconstructed vertices significantly reduces the effect of pileup events in the observables relevant for this study, reaching a level where it can be neglected.

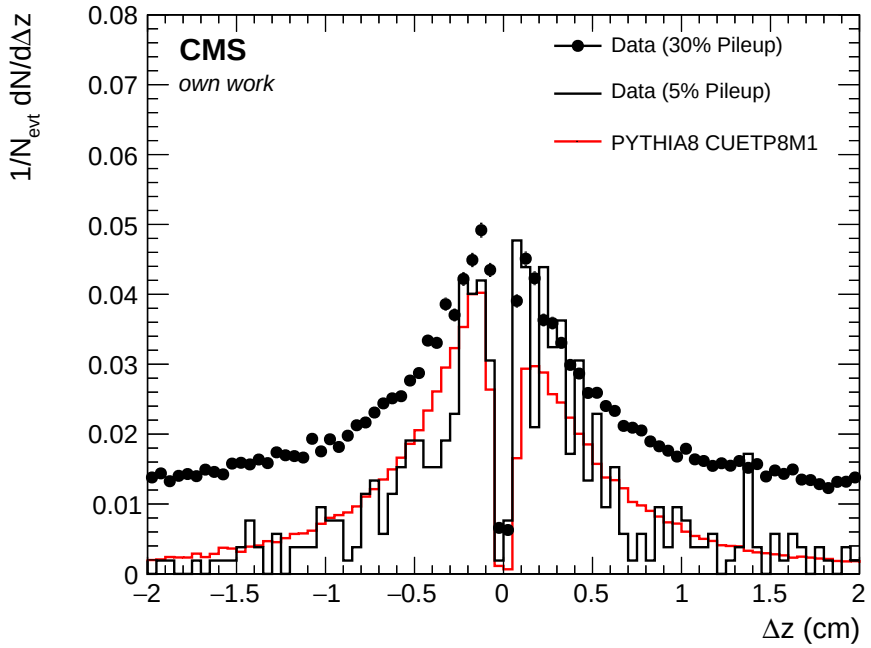



**Fig. 6.18.:** Longitudinal distance of two reconstructed vertices for Monte Carlo simulations with only one proton-proton interaction (red line) as well as for data samples with small (black line) and larger (black markers) pileup.

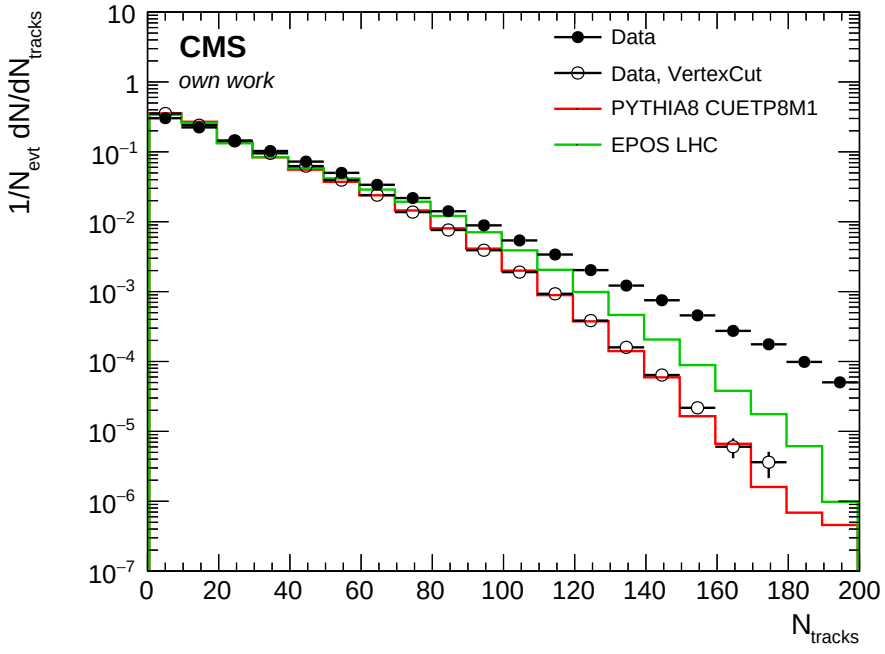



**Fig. 6.19.:** Distribution of track multiplicities normalized to the number of events for Monte Carlo simulations with only one proton-proton interaction (red and green lines) as well as for high pileup data samples without (filled black markers) and with (open black markers) the vertex cut described in the text.

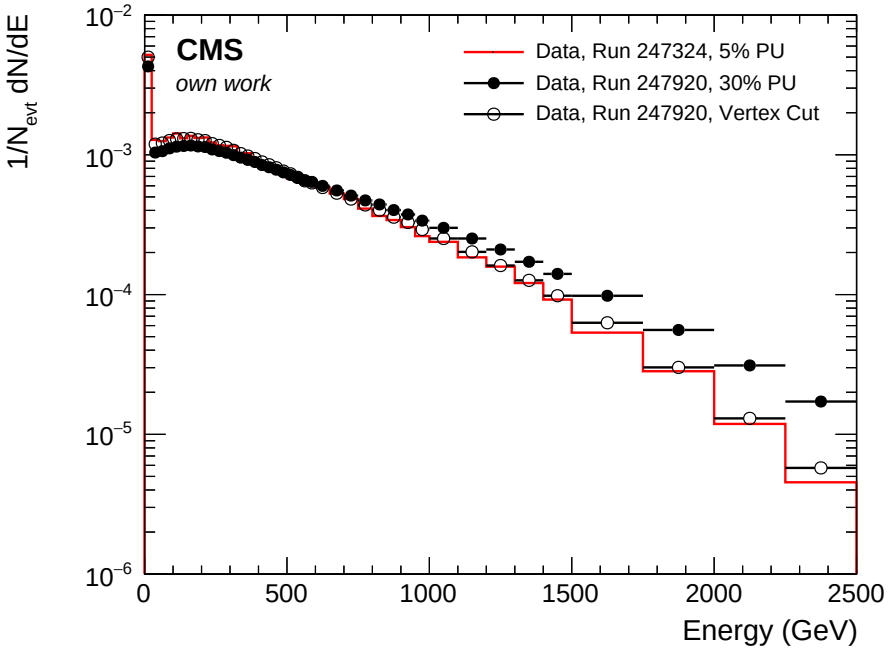



**Fig. 6.20.:** CASTOR energy spectrum for data at different pileup conditions. The data from run 247324 used for the analysis of the energy spectrum is shown as a reference with a red line. The data from run 247920 used for this study is shown without (filled black markers) and with (open black markers) the vertex cut. The tail of high energies due to two overlapping collisions is significantly reduced and the agreement with the low pileup data is obtained.

### 6.2.3 Detector level results

The total energy deposit in CASTOR is calculated in the same way as for the energy spectrum. On detector level all calorimeter towers above noise threshold are summed up. Furthermore, the energy is also separated into the electromagnetic and hadronic contribution. The events are classified according to the number of reconstructed tracks and the average energy is calculated. For each bin in track multiplicity, with $N_{\mathrm{i}}$ events and their energies $E_{\mathrm{j}}$, this is given by

$$\langle E_{\mathrm{i}} \rangle = \frac{1}{N_{\mathrm{i}}} \sum_{\mathrm{j}}^{N_{\mathrm{i}}} E_{\mathrm{j}} \; . \tag{6.8}$$

The used multiplicity bins are given in table 6.5 along with the average number of reconstructed tracks within this bin. The available statistics are sufficient to calculate the average energies up to track multiplicities of 150. At higher multiplicities, the statistical uncertainties in the data as well as the available Monte Carlo simulations become dominant because the number of recorded events decreases dramatically. This is also shown in Fig. 6.19.

The average total reconstructed energy in CASTOR is shown in Fig. 6.21 as a function of the track multiplicity. Data is compared to model predictions of PYTHIA 8 tune CUETP8M1 and tune 4C+MBR as well as EPOS LHC. Systematic uncertainties are shown with a yellow band and are described later in detail. It can be observed that the average energy in CASTOR increases with the track multiplicity. This is reproduced by all models considered. The PYTHIA 8 tunes have very similar shape, although the energy is significantly reduced for tune 4C+MBR over the whole range. The predicted increase at low track multiplicities is steeper than the data, while it is too flat at higher multiplicities. This behavior is even stronger for EPOS LHC. A flat saturation level as predicted by EPOS LHC at multiplicities of 80 and higher is not confirmed by the data. Overall EPOS LHC and PYTHIA 8 CUETP8M1 describe the data well within the uncertainties, while the energies predicted by PYTHIA 8 4C+MBR are significantly too low for $N_{\mathrm{tracks}} < 30$.

The same distribution is shown for in Fig. 6.22 for the electromagnetic and hadronic component. While the shape of the distributions remain unchanged, the absolute scale of the measured energies changes significantly. In the electromagnetic component, PYTHIA 8 CUETP8M1 and EPOS LHC agree well with the data withing the uncertainty whereas PYTHIA 8 4C+MBR is even lower compared to the total energy. An opposite effect is seen for the hadronic component, where PYTHIA 8 4C+MBR performs well within the uncertainties. PYTHIA 8 CUETP8M1 and EPOS LHC on the other hand overestimate the hadronic energies.

**Tab. 6.5.:** Multiplicity classes used in the analysis together with the measured average number of reconstructed tracks in the respective class.

| $N_{\text{tracks}}$ | $\langle N_{\text{tracks}} \rangle$ |
|---|---|
| $0 < N \leq 10$ | 5.05 |
| $10 < N \leq 20$ | 13.97 |
| $20 < N \leq 30$ | 24.15 |
| $30 < N \leq 40$ | 34.17 |
| $40 < N \leq 50$ | 44.12 |
| $50 < N \leq 60$ | 54.11 |
| $60 < N \leq 70$ | 64.08 |
| $70 < N \leq 80$ | 74.04 |
| $80 < N \leq 90$ | 84.01 |
| $90 < N \leq 100$ | 93.93 |
| $100 < N \leq 110$ | 103.9 |
| $110 < N \leq 120$ | 113.8 |
| $120 < N \leq 130$ | 123.8 |
| $130 < N \leq 140$ | 134.0 |
| $140 < N \leq 150$ | 143.7 |

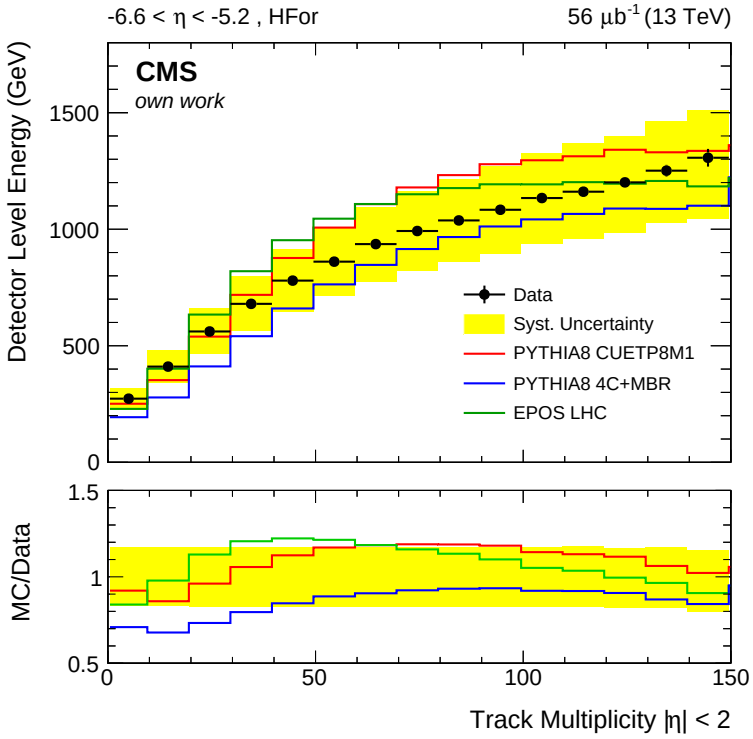



**Fig. 6.21.:** Average total in the CASTOR calorimeter as a function of the number of reconstructed tracks with $|\eta| < 2$ after the offline event selection. Data is shown with black markers along with predictions from PYTHIA 8 tune CUETP8M1 and 4C+MBR. Systematic uncertainties are shown with a yellow band.

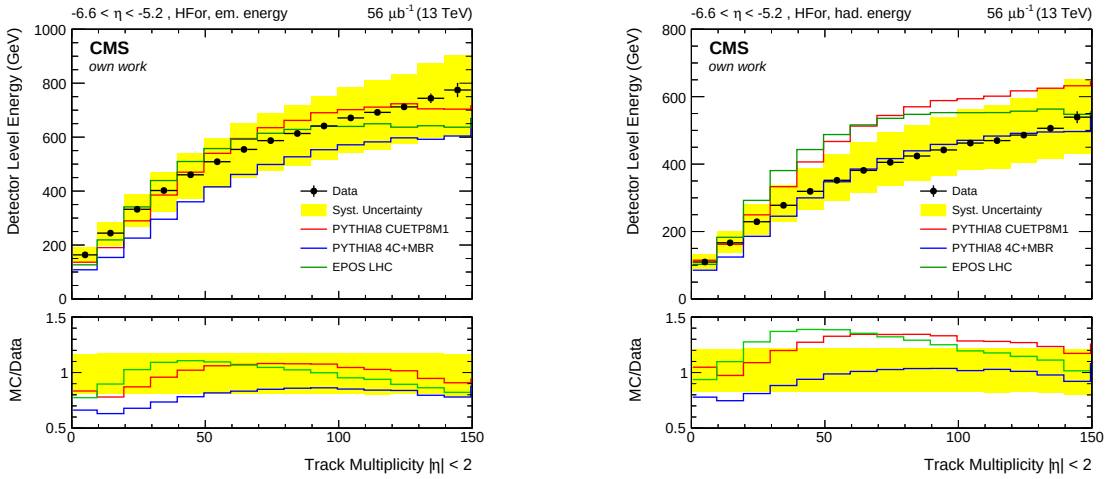



**Fig. 6.22.:** Average electromagnetic (left panel) and hadronic (right panel) in the CASTOR calorimeter as a function of the number of reconstructed tracks with $|\eta| < 2$ after the offline event selection. Data is shown with black markers along with predictions from PYTHIA 8 tune CUETP8M1 and 4C+MBR and EPOS LHC. Systematic uncertainties are shown with a yellow band.

### 6.2.4 Correction to particle level and uncertainties

A bin-by-bin correction method is chosen to correct the obtained results from the detector to the stable particle level, similar to what is used in Refs. [57] and [119]. Two correction factors are calculated for every bin in track multiplicity: one to correct for tracking inefficiencies and the second to correct the detector level to the particle level energy. Both corrections are combined to obtain a global correction as a function of the track multiplicity $C\left(N_{\text{tracks}}\right)$. It is given by:

$$C\left(N_{\text{tracks}}\right) = \frac{\langle E_{\text{true}}\rangle\left(N_{\text{ch}}\right)}{\langle E_{\text{reco}}\rangle\left(N_{\text{ch}}\right)} \times \frac{\langle E_{\text{reco}}\rangle\left(N_{\text{ch}}\right)}{\langle E_{\text{reco}}\rangle\left(N_{\text{tracks}}\right)} \,. \tag{6.9}$$

These corrections are calculated three times for the total, electromagnetic and hadronic energies independently. The stable particle level is defined in a way to match the detector level event selection as well as the stable particle definition of the very forward energy spectra. Therefore events with $\xi > 10^{-6}$ (as defined in Eq. (6.2)) and at least one charged particle with $p_{\text{T}}$ >200 MeV and $|\eta| < 2$ are chosen. Motivated by the above described study of the tracking efficiency the charged particle multiplicity of $N_{\text{ch}} = N_{\text{ch}}\left(p_{\text{T}} > 0.2\,\text{GeV}, |\eta| < 2\right)$ is chosen. The correction for tracking inefficiency (the second term in Eq. (6.9)) is close to one and is independent of the choice of the energy composition. The main contribution to the correction $C$ arises therefore from the non-compensating properties of CASTOR. On average, about 60% of the total particle level energy is reconstructed. The resulting correction factor for the total energy is therefore on average about 1.67. Since the fraction of energy that is not reconstructed is increasing with energy the correction also

increase slightly with higher average energy and consequently also with higher particle multiplicities. The correction of the electromagnetic energy is flat with particle multiplicities since the electromagnetic showers do not suffer from non-compensation. The only correction is due to a small fraction of hadronic energy also being deposited in the electromagnetic section of the calorimeter. The energy correction for the electromagnetic energy is therefore about 0.9. The hadronic energy on the other hand suffers significantly from non-compensation and the correction factor reaches values between 2.5 and 3. The correction factors are shown in detail in Figs. 6.23–6.25. The two contributions to the correction are shown in the top panels separately as a function of track multiplicity or charged particle multiplicity respectively. The overall corrections $C\left(N_{\mathrm{tracks}}\right)$ following Eq. (6.9) are shown in the bottom panels as a function of the track multiplicity. The differences between the three models considered are small and originate from the different fraction of electromagnetic and hadronic energies. The average of the three models is taken to correct the data to the particle level. The spread of the three correction curves is used as a systematic uncertainty of the correction due to model differences.

Various sources of systematic uncertainties are investigated and their effect on the final result is studied. In general, every uncertainty is applied on the detector level distribution and is corrected to the particle level with the same procedure as the central value. The deviations from the central value are calculated for every source of uncertainty and every bin individually. All uncertainties are considered uncorrelated and therefore added in quadrature to obtain the final systematic uncertainties.

**CASTOR energy scale**

As for the energy spectra measured with CASTOR, the uncertainty on the CASTOR energy scale is dominant. The reconstructed energy is scaled up and down by 17%. Due to the fact, that average energies are calculated, this leads to a constant uncertainty of 17% over the whole range of multiplicity.

**Model dependence**

As already described, the bin-by-bin correction introduces a slight model dependence. A systematic uncertainty is estimated by varying the correction factors within the range given by the models (see for example Fig. 6.23, bottom panel). The effect on the final result is below 4%.

**Pileup rejection**

The derived cut on the vertex multiplicity is varied in a conservative way in order to estimate the effect on the ensemble of selected events and the average energies accordingly. Instead of the standard cut on events with maximal 2 vertices that are less than 0.5 cm apart from each other, once only events with maximally one reconstructed vertex are selected. The cut is also softened by allowing two-vertex events with a distance of 0.7 cm.

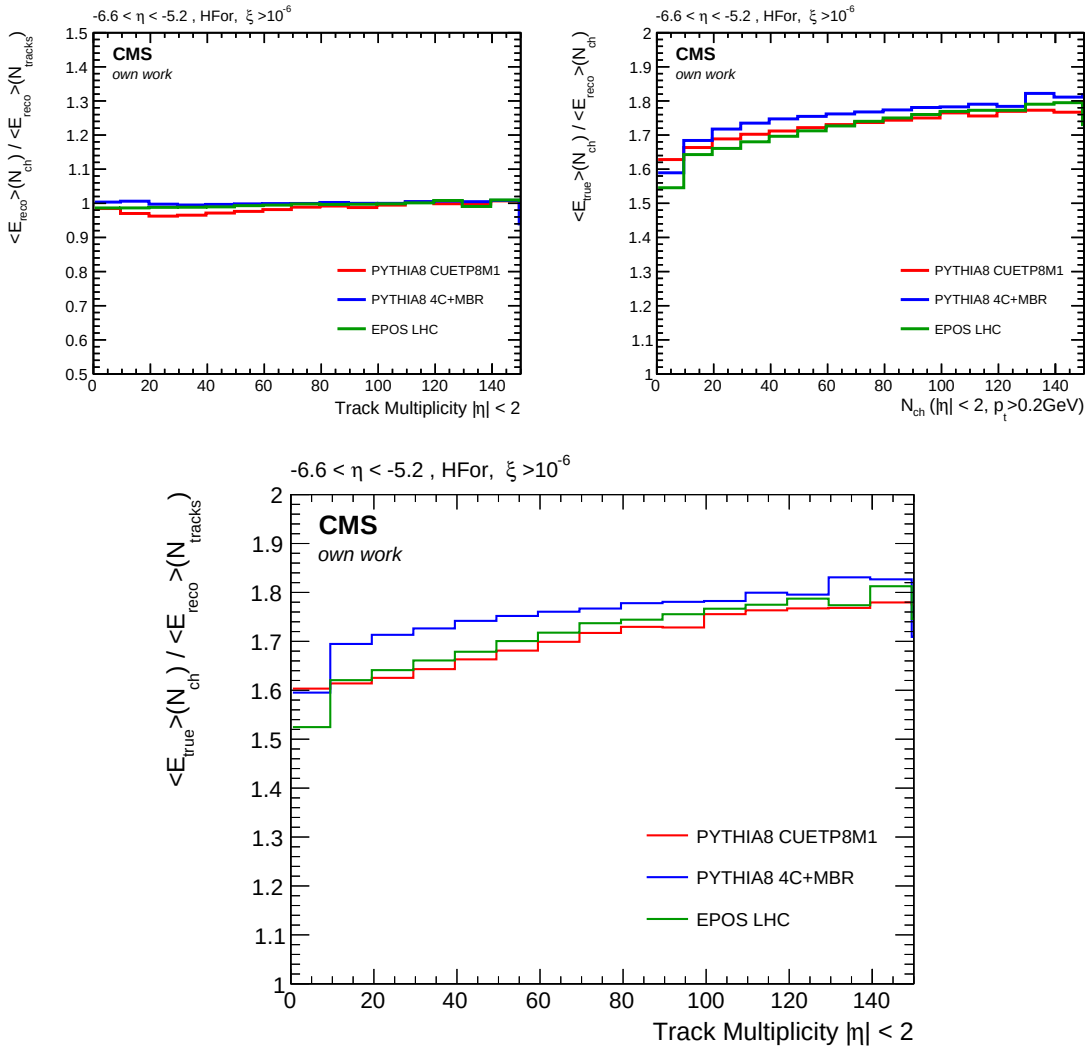



**Fig. 6.23.:** Components of the bin-by-bin correction factors as defined in Eq. (6.9) used to correct for tracking inefficiencies (top left panel) and energy reconstruction of the total energy (top right panel). The combination of both is shown in the bottom panel. The factors are shown for PYTHIA 8 tune CUETP8M1 and 4C+MBR and EPOS LHC for which detailed detector simulation is available.

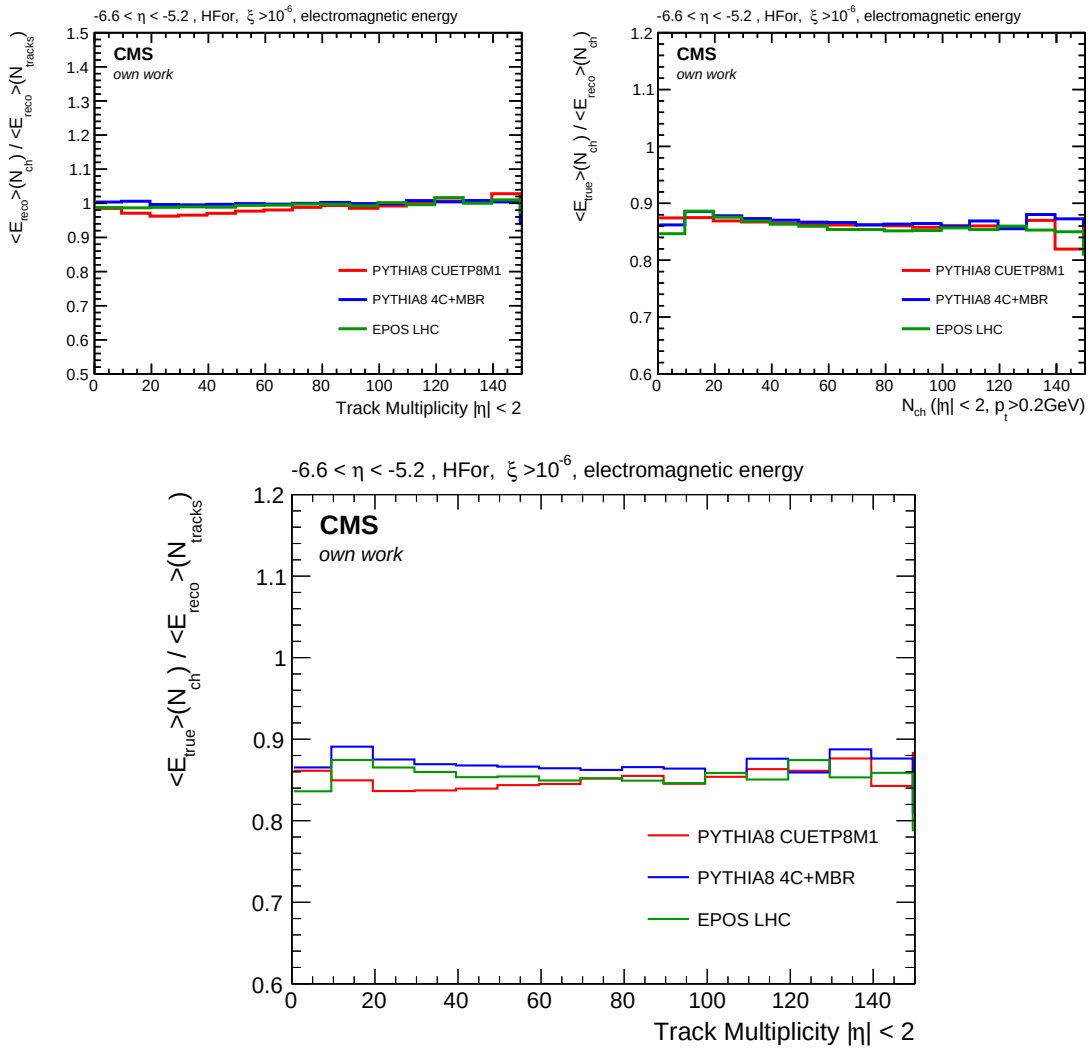



**Fig. 6.24.:** Components of the bin-by-bin correction factors as defined in Eq. (6.9) used to correct for tracking inefficiencies (top left panel) and energy reconstruction of the electromagnetic energy (top right panel). The combination of both is shown in the bottom panel. The factors are shown for PYTHIA 8 tune CUETP8M1 and 4C+MBR and EPOS LHC for which detailed detector simulation is available.

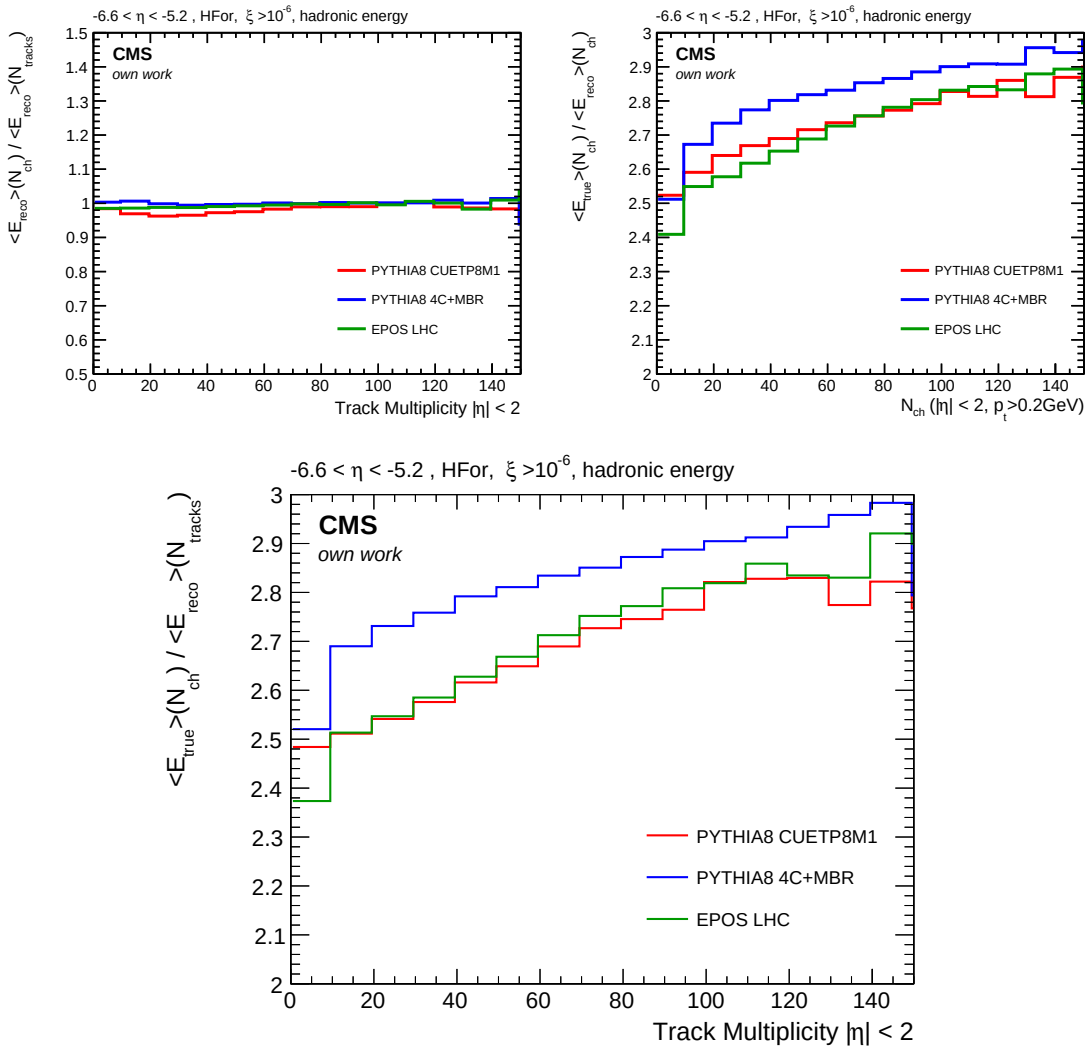



**Fig. 6.25.:** Components of the bin-by-bin correction factors as defined in Eq. (6.9) used to correct for tracking inefficiencies (top left panel) and energy reconstruction of the hadronic energy (top right panel). The combination of both is shown in the bottom panel. The factors are shown for PYTHIA 8 tune CUETP8M1 and 4C+MBR and EPOS LHC for which detailed detector simulation is available.

These changes in the event selection mainly impact higher multiplicities and lead to a systematic uncertainty of 6% in the highest bin.

**HF energy scale**
The uncertainties on the reconstructed energies in the HF calorimeters is 10%. The effect of varying the noise threshold for the event selection from 5.0 GeV to 4.5 and 5.5 GeV respectively was studied and found to be below 1% over the whole multiplicity range.

**Tracking**
The correction to the stable particle level relies on the assumption that the properties of the CMS pixel tracker are exactly reproduced in simulations. Moreover, the tracking efficiency is assumed to be the same in data than in simulations. These kind of systematic uncertainties have been studied in detail in Ref. [120]. It is found that the uncertainties on tracking and vertexing efficiency are 1.8 and 2-3% respectively. In this analysis, these are combined in a conservative way and 5% uncertainty on the track reconstruction is assumed. The data has been reprocessed once with 5% more and once with 5% less reconstructed tracks per event. This leads to a change in the average energies per bin of particle multiplicities due to migrations from one bin to another. The total effect on the result is below 6% over the whole range.

**Intercalibration**
The separation of electromagnetic and hadronic energy suffers in addition from an intercalibration uncertainty. This amounts to -8% for the electromagnetic and +12% for the hadronic energy.

Adding up all these contributions in quadrature, one obtains an almost flat uncertainty of about 18%, dominated by the CASTOR energy scale uncertainty. The single components to the systematic uncertainties are shown in the appendix in Figs. C.8–C.10 for the total, electromagnetic and hadronic energies respectively.

### 6.2.5 Results

The average energy per proton-proton collision that is produced in the acceptance of the CASTOR calorimeter is measured as a function of the charged particle multiplicity in the central region and fully corrected to the stable particle level. Various sources of systematic uncertainties are studied and the final results are compared to various model predictions in Figs. 6.26–6.28. The results are given for the total energy as well as the electromagnetic and hadronic energies.

All models agree with the data in the fact that the average energy in the very forward direction rises with the general collision activity, which here is characterized by the charged particle multiplicity. Only PYTHIA 8 fails to describe this rise when multiparton

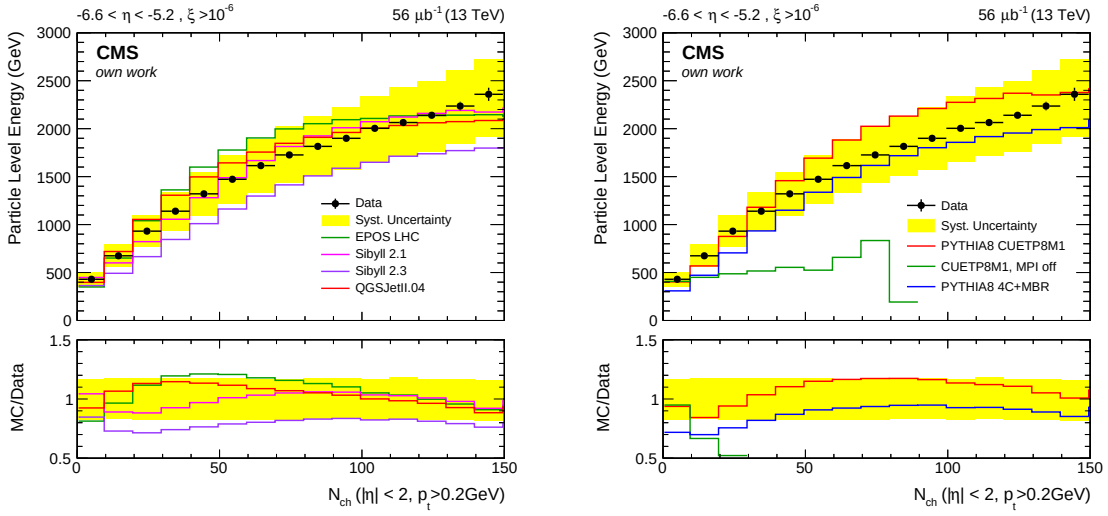



**Fig. 6.26.:** Average total energy in the acceptance of CASTOR at $-6.6 < \eta < -5.2$ for events with $\xi > 10^{-6}$ as a function of the charged particle multiplicity $N_{\mathrm{ch}}$ ($p_{\mathrm{T}} > 0.2\,\mathrm{GeV}, |\eta| < 2$). The data is fully corrected for detector effects and compared to predictions by hadronic event generators commonly used in high energy cosmic ray physics (left panel) and variations of PYTHIA 8 (right panel).

interactions (MPI) are switched off. This emphasizes once more the fact that the particle production at very forward pseudorapidity is fundamentally sensitive to the modeling of MPI. Most of the models considered also reproduce the total amount of energy within the experimental uncertainties (see Fig. 6.26). PYTHIA 8 4C+MBR predicts too little energy at low multiplicities. SIBYLL 2.3 underestimates the energy production over the whole range. While especially EPOS LHC and QGSJETII.04 predict a flat plateau region at multiplicities larger than about 100, the data show a steady increase with energy. Even at the highest studied multiplicities, this increase is stronger than predicted by all the models considered here.

The separation of the energy into the contribution from electromagnetic and hadronic particles reveals more detailed characteristics of the models. While most of the models agree very well with the electromagnetic energies (see Fig. 6.27) SIBYLL 2.3 and PYTHIA 8 4C+MBR undershoot significantly and in the case of SIBYLL 2.3 dramatically. This is consistent with the observation of a too steep electromagnetic spectrum compared to the data shown in Fig. 6.11. In the hadronic energy (see Fig. 6.28) the opposite effect is seen. Here EPOS LHC and PYTHIA 8 CUETP8M1 predict too large energies at intermediate multiplicities, SIBYLL 2.3 and PYTHIA 8 4C+MBR agree with the data within uncertainties.

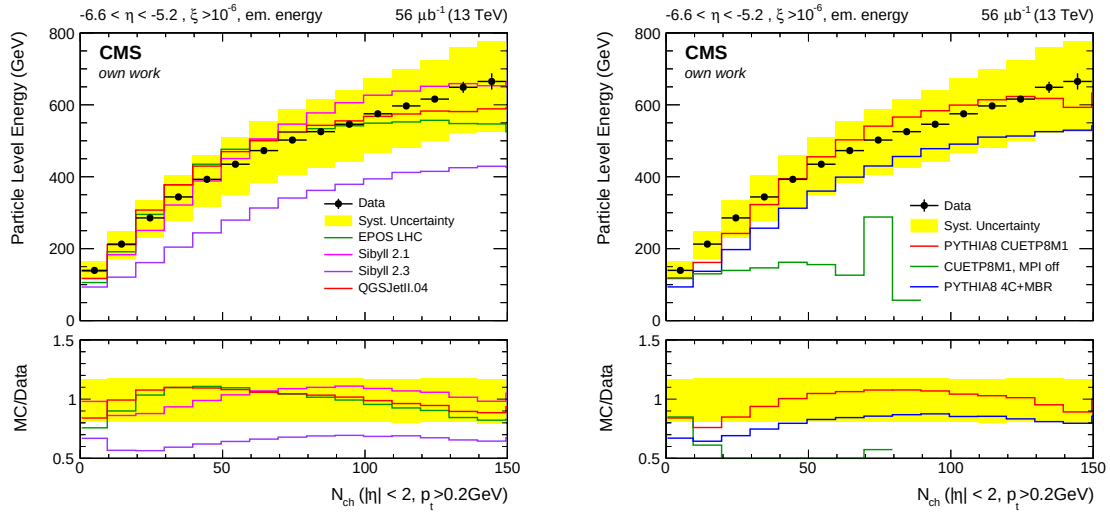



**Fig. 6.27.:** Average electromagnetic energy in the acceptance of CASTOR at $-6.6 < \eta < -5.2$ for events with $\xi > 10^{-6}$ as a function of the charged particle multiplicity $N_{\mathrm{ch}}\,(p_{\mathrm{T}} > 0.2\,\mathrm{GeV}, |\eta| < 2)$. The data is fully corrected for detector effects and compared to predictions by hadronic event generators commonly used in high energy cosmic ray physics (left panel) and variations of PYTHIA 8 (right panel).

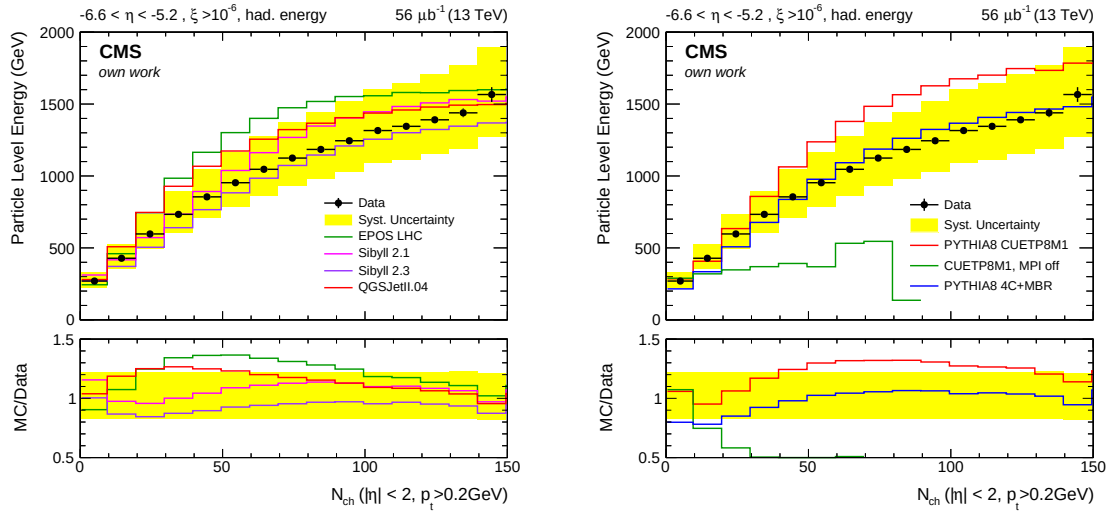



**Fig. 6.28.:** Average hadronic energy in the acceptance of CASTOR at $-6.6 < \eta < -5.2$ for events with $\xi > 10^{-6}$ as a function of the charged particle multiplicity $N_{\mathrm{ch}}\,(p_{\mathrm{T}} > 0.2\,\mathrm{GeV}, |\eta| < 2)$. The data is fully corrected for detector effects and compared to predictions by hadronic event generators commonly used in high energy cosmic ray physics (left panel) and variations of PYTHIA 8 (right panel).

# 7 Implications for hadronic interaction models

The presented measurements in the very forward phase space at 13 TeV are unique at the LHC and bear an enormous potential to improve the understanding of hadronic collisions. The detailed interpretation of the data in order to find precise implications on the underlying physics is an extensive topic. Some of the most prominent and apparent results are described in this chapter.

## 7.1 Tuning of hadronic event generators

It is another highlight of this thesis that a new parameter set for the PYTHIA 8 event generator is determined. Such an improved parameter set is typically called *tune*. The tune derived as part of this thesis is expected to become one of the main references for CMS data analyses at 13 TeV in the future.

Multipurpose event generators face the challenge to describe the full final state of hadronic collisions. While some aspects are well described theoretically, for example within the framework of perturbative QCD, and can be calculated very precisely, others remain to be approximated by the models. This is especially true for non-perturbative physics whose description is deeply phenomenological. Event generators therefore rely on parameterizations with a large set of parameters that can be freely chosen within a limited range set by the physics framework. These parameters include for example the hadronization processes as well as soft interactions such as multiparton interactions and the treatment of the beam remnant. The event generators are continuously tested and improved as more collider data are analyzed. Especially the extrapolation to higher center-of-mass energies needs to be verified as the LHC energies increase. Special attention is payed to optimizing the free parameters in a way that the models describe the measured collision properties as accurately as possible while the key features of the model remain unchanged. This process is often referred to as *generator tuning*.

Since many years, tuning efforts are carried out in different ways and to different extents. Major model tunes usually include a large set of modified parameters combined with a large set of input data from both electron-positron as well as hadron colliders. Examples are the Perugia Tunes of PYTHIA 6 [125], the tunes 4C [126] and Monash [127] of PYTHIA 8, or the updated version of EPOS 1.99, EPOS LHC.

Next to these extensive ones, smaller tunes are obtained to improve the models descriptions for only some specific aspects of hadron collisions. Consequently only a small number of parameters are tuned using a smaller set of measurements. These tunes usually base upon extensive tunes and improve only parts of the generator in order to reach optimal performance for physics analyses. Examples are the CMS underlying event tune CUETP8M1 [58] or the ATLAS tunes A2/AU2 [128] of PYTHIA 8.

### 7.1.1 RIVET and PROFESSOR

A systematic tuning of model parameters represents a big computational and scientific challenge since the interplay between different parameters is not easy to understand and the change of the generator predictions to a given parameter change is not straightforward. A manual tuning *by eye* requires deep understanding of the impact of every single parameter and can only be done by few experts. Furthermore it requires a lot of time to process and evaluate the generator response to a parameter change and validate that a given change would not decrease the performance to other or future data.

The RIVET [129] and PROFESSOR [130] software provide a framework in which a systematic tune of any generator can be performed on an short timescale. The interplay of these programs is smooth and easily extendable to new developments on both the generator and experimental side. These software are successfully used in recent tuning efforts and are also used in this work to test whether the presented measurements with CASTOR are able to improve the model predictions of PYTHIA 8.[1]

**RIVET** is a software package that is designed to facilitate the comparison of generator predictions with experimental data. It is independent of event generators by using the common HEPMC [93] format as an interface. The RIVET library includes a large set of analyses from collider experiments such as HERA, LEP, Tevatron, and the LHC, that can be used in a modular way to compare model predictions to the data. Provided externally generated events in the HEPMC format, RIVET produces distributions in the yoda-format following the same event selection and representation of the data. This way, experimental data can also be compared to generators that were not existing at the time of the original publication as well as to modified generators in which, for example, some parameters have been changed during tuning. RIVET therefore provides a perfect tool for systematic model tuning.

**PROFESSOR** provides numerical tools to systematically study distributions in the yoda-format provided by RIVET. The main feature is the ability to interpolate the generator response as a function of varying generator settings. A Goodness-of-fit (Gof) function is defined and minimized in order to find the optimal parameter settings to describe reference

[1] The tuning procedure was developed and validated extensively for PYTHIA 8 but is as well applicable to any other hadronic event generator. Is was for example adapted for HERWIG 7.1, which is summarized in appendix D.5.

data. This functionality is used to find a new parameter tune for the input generator. In the following, the general method used by PROFESSOR shall be briefly explained. A detailed description is given in Ref. [130].

The response of a given generator is evaluated for every bin $b$ of a set of observables and for a large number of different parameter settings $\vec{p}$. The response is afterwards parameterized with a polynomial function of at least second order:

$$\mathrm{MC}_b\left(\vec{p}\right) \approx f^{(b)}\left(\vec{p}\right) = \alpha_0^{(b)} + \sum_i \beta_i^{(b)} p_i' + \sum_{i \leq j} \gamma_{ij}^{(b)} p_i' p_j' + \dots \,. \tag{7.1}$$

The parameters $\alpha_0^{(b)}, \beta_i^{(b)}, \gamma_{ij}^{(b)}, \dots$ are determined with a single value decomposition method that is described in detail in Ref. [130]. Once these parameters are found, a global $\chi^2$ function is defined as a function of the free parameters $\vec{p}$:

$$\chi^2\left(\vec{p}\right) = \sum_b \frac{\left(f^{(b)}\left(\vec{p}\right) - r_b\right)^2}{\sigma_b^2} \,, \tag{7.2}$$

where $r_b$ is the reference data of bin $b$ and $\sigma_b$ the corresponding experimental uncertainty. This $\chi^2$ function is numerically minimized and the resulting parameter vector is estimated to be the best tune of the generator to the chosen observations.

### 7.1.2 New CMS underlying event tunes of PYTHIA 8

Within the CMS collaboration, parameter tunes of the PYTHIA 6, PYTHIA 8 and HERWIG++ event generators were obtained as main references for physics analyses in the beginning of the LHC Run 2 [58]. The charged-particle and $p_{\mathrm{T}}^{sum}$ densities in the TransMax and TransMin regions measured at three energies ($\sqrt{s} = 0.9, 1.96\,\mathrm{TeV}$ by CDF [131] and $\sqrt{s} = 7\,\mathrm{TeV}$ by CMS [46]) were used as reference data. The tuning focused on parameters related to the underlying event, especially MPI, as listed in Tab. 7.1. Only a small set of parameters was tuned, while others, for example concerning the parton showers or fragmentation, remained unchanged with respect to the reference tune Monash. Table 7.1 shows the obtained tuning results for the tune CUETP8M1 of PYTHIA 8 together with the values from the Monash tune on which it is based. Both of these tunes use the NNPDF2.3LO [132, 133] parton distribution function (PDF). It was found with the first results at 13 TeV, that the new tune was performing well but not optimal in some cases. Especially in the measurement of the underlying event at 13 TeV [47], the original Monash tune was found to match better with data than the tune CUETP8M1. With a detailed study of final states with top-pairs $\mathrm{t\bar{t}}$ at $\sqrt{s} = 8$ and 13 TeV [134] it was found that a further tuning of the PYTHIA 8 tune CUETP8M1 is necessary to better describe the observations. The obtained parameter set was called CUETP8M2T in Ref. [134] and will be renamed to CP0 in the future. This lead the path to a new set of tunes CP1–CP4. They all use

the Monash tune as the baseline with an updated PDF set NNPDF3.1 [135] at leading and (next-to-)next-to-leading orders. This is especially important since the updated PDF set includes fits to a large number of LHC precision measurements. Significant changes can especially be seen in the gluon densities at low $x$ and low $Q$ (see Fig. 7.1) which is expected to have a big impact on the underlying event. Therefore the change of the PDF set alone requires a new tune of the parameters related to soft gluon interactions. In the following the tune CP1 will be discussed in more detail, since it was obtained in the scope of the presented work.

The tuning procedure of the tune CUETP8M1 is repeated but this time includes underlying event data measured by CMS at 13 TeV. The generator response of PYTHIA 8 version 8.226 is calculated at 200 points with different parameter values within the tuning range given in Tab. 7.1. A statistics of two million events is chosen at every point in the parameter space to secure a low impact of the generator statistics. The results are interpolated with a third order polynomial according to Eq. (7.1) and the optimal parameter values are obtained by the PROFESSOR software. A journal publication describing the tune procedure and results is being prepared.

The new tune CP1 is the one most similar to the Monash tune using the leading order PDF set. The strong coupling constant $\alpha_\mathrm{s}$ at scale $M_\mathrm{Z}^2$ is kept at the PYTHIA 8 default values for the hard process, the MPI as well as final and initial state parton showers. In Tab. 7.1, the parameter set obtained with the tuning procedure for tune CP1 is shown and compared to the Monash and CUETP8M1 tunes. The most important change with respect to previous tunes – next to the updated PDF – is a much smaller energy scaling factor of the MPI cutoff *MultipartonInteractions:ecmPow*. This leads to a similar $p_\mathrm{T}$-dampening of MPI at the reference energy of 7 TeV compared to previous tunes but a smaller value at 13 TeV. This compensates for the lower gluon density as low $x$ coming from the updated PDF. Furthermore, the impact parameter profile is changed from an exponential profile to a double gaussian, replacing the free parameter *MultipartonInteractions:expPow* with two free parameters *MultipartonInteractions:coreFraction* and *MultipartonInteractions:coreRadius*. Details on the parameters of PYTHIA 8 that are tuned are given in appendix D.1 and the differences between tunes CP1–CP4 are summarized in appendix D.2.

### 7.1.3 Forward energy measurements as additional constraints

It was shown in this thesis that the energy measurements in the very forward region with CASTOR bear the potential to add valuable input for the generator modeling of MPI, parton showers of initial- and final state radiation, and related phenomena. Furthermore, the very forward data is expected to be sensitive to the low $(x, Q^2)$ gluon contribution, which significantly changed with the updated PDF set. It is therefore studied to which extent the tune CP1 would change if the data presented in this work are included in the generator tune performed with PROFESSOR.

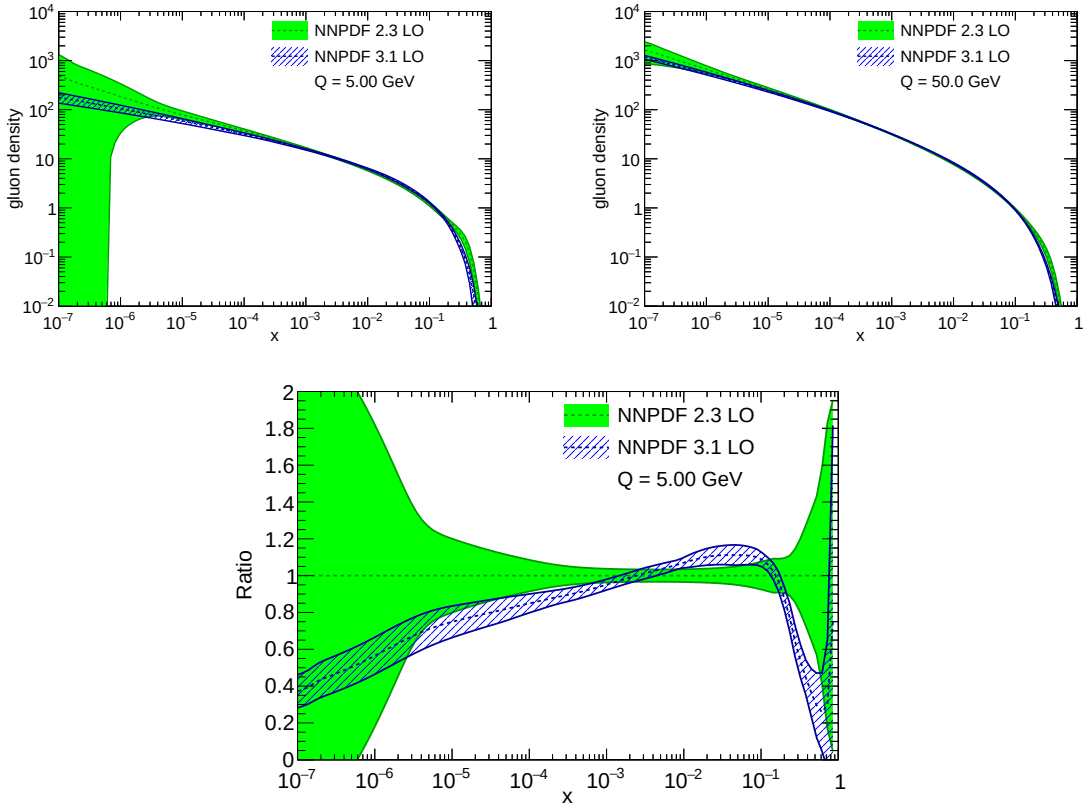



**Fig. 7.1.:** Gluon density as a function of $x$ for the NNPDF 2.3 and NNPDF 3.1 PDF sets at leading order. The gluon density is evaluated at scales $Q = 5\,\text{GeV}$ (top left) and $Q = 50\,\text{GeV}$ (top right). The bottom panel shows the ratio of the two PDF sets at $Q = 5\,\text{GeV}$. The graphs are obtained with the APFEL 2.7.1 WEB tool [136, 137].

**Tab. 7.1.:** Parameters of the PYTHIA 8 tunes Monash and CUETP8M1. The tuning range and obtained best-fit values for the newly derived tune CP1 are given as well.

| PYTHIA 8 Parameter | Monash | CUETP8M1 | Tuning range | CP1 |
|---|---|---|---|---|
| NNPDF Version | 2.3LO | 2.3LO | | 3.1LO |
| MultipartonInteractions:pT0Ref [GeV] | 2.280 | 2.402 | $1.5 - 4.0$ | 2.4 |
| MultipartonInteractions:ecmPow | 0.215 | 0.252 | $0.1 - 0.4$ | 0.154 |
| MultipartonInteractions:expPow | 1.85 | 1.6 | – | – |
| MultipartonInteractions:coreFraction | – | – | $0.1 - 0.95$ | 0.684 |
| MultipartonInteractions:coreRadius | – | – | $0.1 - 0.8$ | 0.544 |
| ColourReconnection:range | 1.80 | 1.80 | $1.0 - 9.0$ | 2.633 |

**Tab. 7.2.:** Reference data used in the tuning procedure for the PYTHIA 8 tunes CP1 (top part) and CP1F (full table).

| Experiment, $\sqrt{s}$ | Observable | RIVET Reference |
|---|---|---|
| Tune CP1 | | |
| CDF, 1.96 TeV | $\langle \mathrm{d}N_{\mathrm{ch}}/\left(\mathrm{d}\eta\mathrm{d}\phi\right)\rangle$ (TransMin) | /CDF_2015_I1388868/d01-x01-y02 |
| | $\langle \mathrm{d}N_{\mathrm{ch}}/\left(\mathrm{d}\eta\mathrm{d}\phi\right)\rangle$ (TransMax) | /CDF_2015_I1388868/d01-x01-y03 |
| | $\langle \sum p_{\mathrm{T}}\rangle/\left(\mathrm{d}\eta\mathrm{d}\phi\right)\rangle$ (TransMin) | /CDF_2015_I1388868/d01-x01-y07 |
| | $\langle \sum p_{\mathrm{T}}\rangle/\left(\mathrm{d}\eta\mathrm{d}\phi\right)\rangle$ (TransMax) | /CDF_2015_I1388868/d01-x01-y08 |
| CMS, 7 TeV | $\langle \mathrm{d}N_{\mathrm{ch}}/\left(\mathrm{d}\eta\mathrm{d}\phi\right)\rangle$ (TransMin) | /CMS_FSQ_12_020/d06-x01-y01 |
| | $\langle \mathrm{d}N_{\mathrm{ch}}/\left(\mathrm{d}\eta\mathrm{d}\phi\right)\rangle$ (TransMax) | /CMS_FSQ_12_020/d05-x01-y01 |
| | $\langle \sum p_{\mathrm{T}}\rangle/\left(\mathrm{d}\eta\mathrm{d}\phi\right)\rangle$ (TransMin) | /CMS_FSQ_12_020/d09-x01-y01 |
| | $\langle \sum p_{\mathrm{T}}\rangle/\left(\mathrm{d}\eta\mathrm{d}\phi\right)\rangle$ (TransMax) | /CMS_FSQ_12_020/d08-x01-y01 |
| CMS, 13 TeV | $\mathrm{d}N/\mathrm{d}\eta$ | /CMS_2015_I1384119/d01-x01-y01 |
| | $\langle \mathrm{d}N_{\mathrm{ch}}/\left(\mathrm{d}\eta\mathrm{d}\phi\right)\rangle$ (TransMin) | /CMS_FSQ_15_007/d05-x01-y01 |
| | $\langle \mathrm{d}N_{\mathrm{ch}}/\left(\mathrm{d}\eta\mathrm{d}\phi\right)\rangle$ (TransMax) | /CMS_FSQ_15_007/d06-x01-y01 |
| | $\langle \sum p_{\mathrm{T}}\rangle/\left(\mathrm{d}\eta\mathrm{d}\phi\right)\rangle$ (TransMin) | /CMS_FSQ_15_007/d01-x01-y01 |
| | $\langle \sum p_{\mathrm{T}}\rangle/\left(\mathrm{d}\eta\mathrm{d}\phi\right)\rangle$ (TransMax) | /CMS_FSQ_15_007/d02-x01-y01 |
| Additionally in tune CP1F | | |
| CMS, 7 TeV | $\left(\mathrm{d}E^{\mathrm{hard}}/\mathrm{d}\eta\right)$ / $\left(\mathrm{d}E^{\mathrm{incl}}/\mathrm{d}\eta\right)$ | /CMS_2013_I1218372/d03-x01-y01 |
| CMS, 13 TeV | $\mathrm{d}E/\mathrm{d}\eta$ | /CMS_FSQ_15_006/d01-x01-y01 |
| | $\mathrm{d}\sigma/\mathrm{d}E_{\mathrm{tot}}$ | /CMS_2017_I1511284/d01-x01-y01 |
| | $\langle E_{\mathrm{tot}}\rangle\,(N_{\mathrm{Ch}})$ | n.a. |

The data considered for the CMS underlying event tunes CP1–CP4 are the charged-particle and $p_{\mathrm{T}}^{sum}$ densities in the TransMax and TransMin regions which have been proven to be sensitive to changes in generator parameters describing MPI and parton showers. The forward energy density $\mathrm{d}E/\mathrm{d}\eta$, the spectrum of the total energy in CASTOR $\mathrm{d}\sigma/\mathrm{d}E_{\mathrm{tot}}$ and the average CASTOR energy as a function of the charged particle multiplicity $\langle E_{\mathrm{tot}}\rangle\,(N_{\mathrm{ch}})$ are added with equal weight to the PROFESSOR tune. Since the energy dependence of MPI is also tuned in CP1, the measurement of the CASTOR energy as a function of the charged jet $p_{\mathrm{T}}$ at $\sqrt{s} = 7\,\mathrm{TeV}$ as published in Ref. [57] is also included in the tune procedure. All observables included in the tune are listed in Tab. 7.2. The reference data of these additional four distributions are shown in Fig. 7.2 together with the envelope of all 200 generator responses within the tuning range. It can be observed that most of the data are covered by the spread of the generator. The rising part of the average CASTOR energy as function of the particle multiplicity and the bump region of the CASTOR energy spectrum are not completely covered. This is due to restrictions of the model itself. These observables are not expected to be perfectly described by tuning only the given parameters.

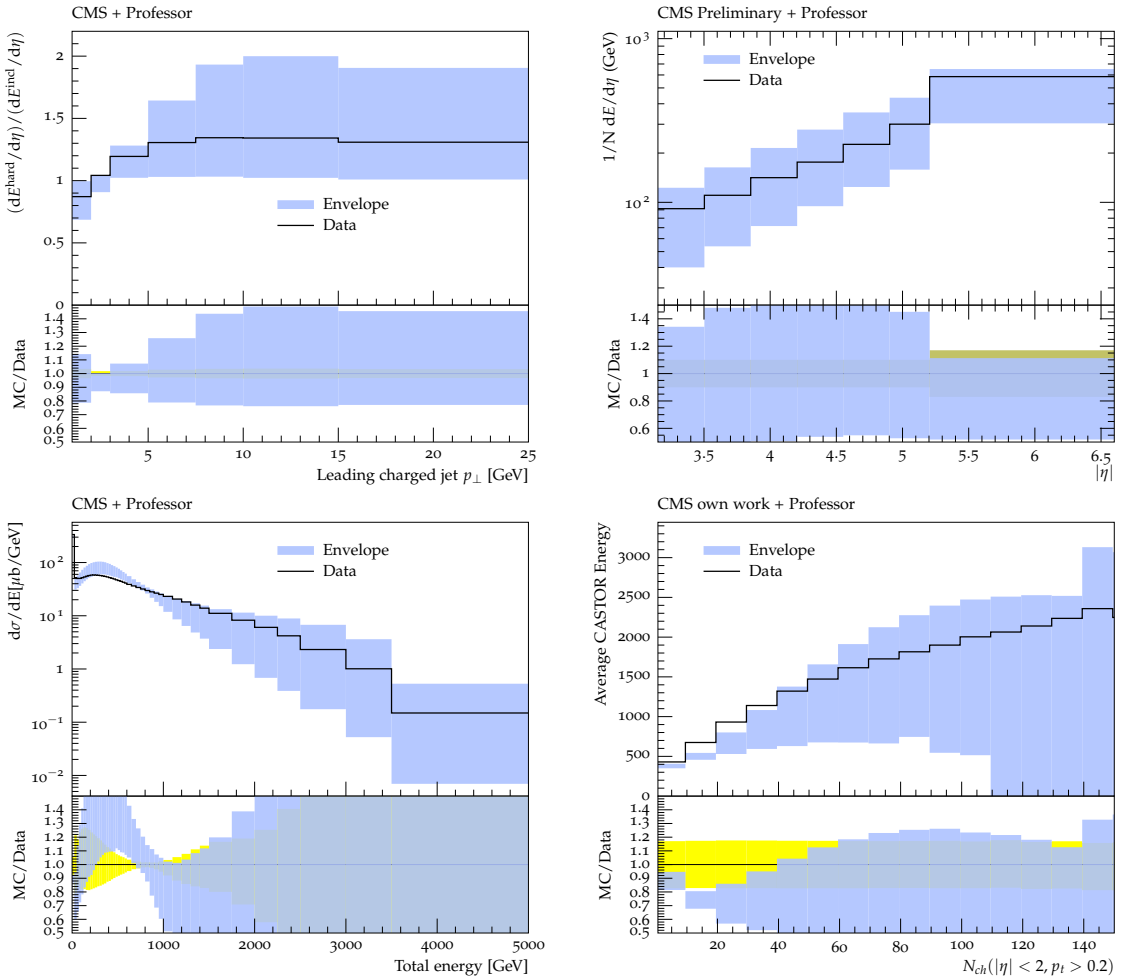



**Fig. 7.2.:** CMS measurements in the very forward region that are used as additional constraints for the PYTHIA 8 tune CP1F. Energy density ratio as a function of leading jet $p_T$ at 7 TeV [57] (top left), energy density in the forward region (top right), total energy spectrum with CASTOR (bottom left), and multiplicity dependent energy in CASTOR (bottom right) all at 13 TeV. The reference data are shown (black markers and yellow band) together with the envelope of the 200 generator generator responses used in the parameter tuning (blue band).

**Tab. 7.3.:** The tuned PYTHIA 8 parameters and their values in the tune CP1 and tune CP1F which includes the energy measurements with CASTOR.

| PYTHIA 8 Parameter | CP1 | CP1F |
|---|---|---|
| NNPDF Version | 3.1LO | |
| MultipartonInteractions:pT0Ref [GeV] | 2.4 | 2.319 |
| MultipartonInteractions:ecmPow | 0.154 | 0.157 |
| MultipartonInteractions:coreFraction | 0.684 | 0.679 |
| MultipartonInteractions:coreRadius | 0.544 | 0.669 |
| ColourReconnection:range | 2.633 | 3.849 |

The same sample points in the parameter space and interpolations as used for the above described tune CP1 are used to determine a second dedicated tune CP1F, indicating the same configuration as tune CP1 but with additional constraints from forward data. The obtained results are given in Tab. 7.3. Most of the parameters receive only small changes with respect to tune CP1. Still it is interesting to note that the parameters *MultipartonInteractions:coreRadius* and *ColourReconnection:range* are noticeably increased. The performance of these newly derived tunes is evaluated by studying their predictions on a large set of observables. Some examples are shown in Fig. 7.3, comparing the prediction to data that were not used directly for the tune. Both tunes CP1 and CP1F perform equally well in describing the underlying event measured in the TransDiff and TransAvg regions. The agreement with the data has significantly improved with respect to the tune CUETP8M1, especially in the plateau region above 5 GeV. The inclusive jet spectrum measured with CASTOR at 13 TeV [138] was already well described by CUETP8M1 within the uncertainties. Still the new tunes agree even better with the central values of the measurement. The bottom right panel of Fig. 7.3 shows the charged particle multiplicity spectrum as measured by ATLAS [139]. It can be seen that the displaced peak of PYTHIA 8 with respect to the data is not corrected for by the tune, but the slope of the distribution is affected. While CP1 is very similar to CUETP8M1, CP1F differs significantly. It describes the fraction of events at intermediate multiplicities between 30 and 100 much better than the other tunes, while the high multiplicity tail is underestimated. Some more figures comparing the model predictions to various measurements can be found in Appendix D.4.

The uncertainties associated to the tuning procedure are commonly estimated with so-called *eigentunes* that are provided by the PROFESSOR software. During the minimization of the interpolated generator response, the covariance matrix between the tuned parameters is calculated at the point of the best tune. PROFESSOR then calculates maximally independent principle directions in the parameter space. The number of these directions and hence the number of obtained eigentunes is twice the number of tuned parameters, ten in the case of CP1(F). The eigentunes themselves are then created by walking out from the best tune point along the principle directions until the calculated $\chi^2$ has increased by

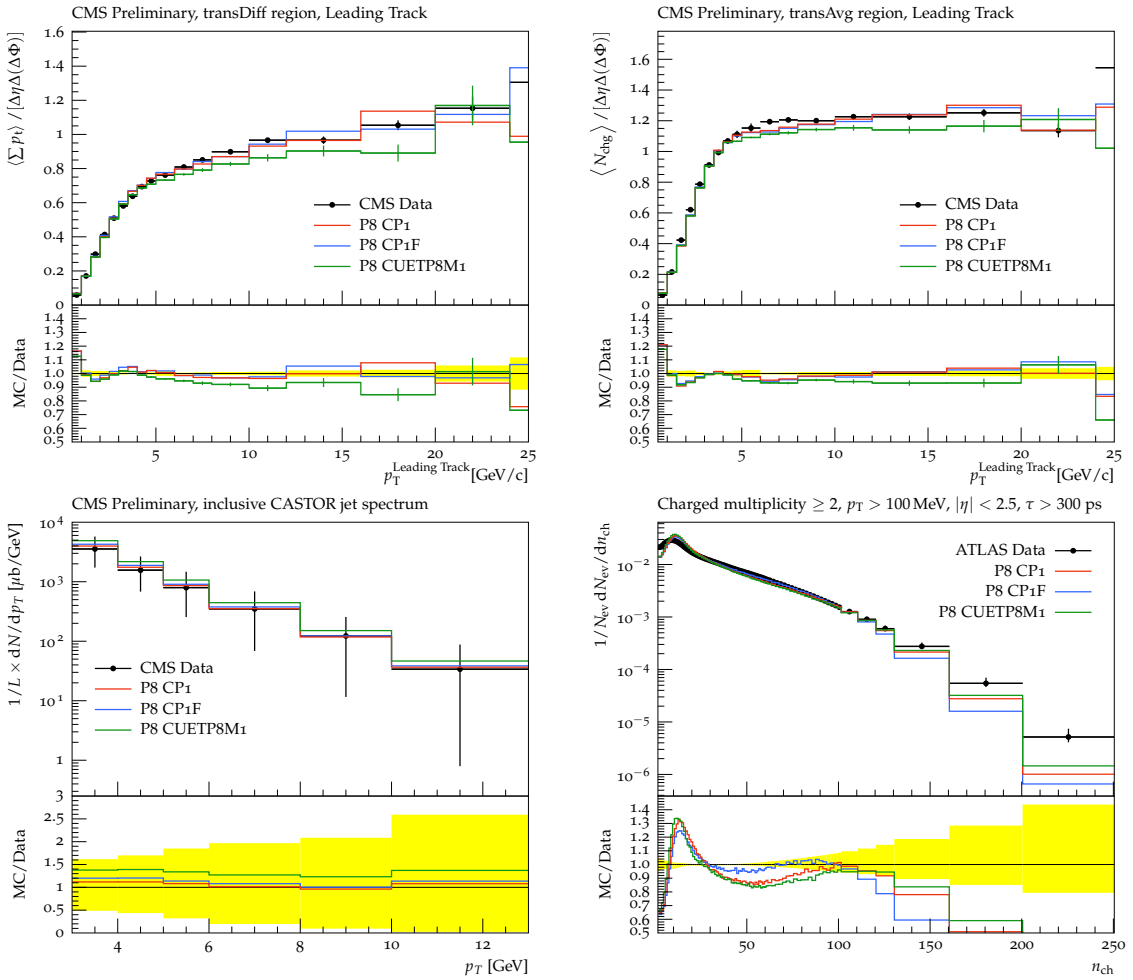



**Fig. 7.3.:** Examples to evaluate the performance of the newly derived tunes of PYTHIA 8 compared to tune CUETP8M1. The top left panel shows the average $p_T$ density measured by CMS in the TransDiff region and the top right the particle density in the TransAvg region as a function of the leading track $p_T$ at $\sqrt{s} = 13$ TeV. The inclusive jet cross section of very forward jets is shown on the bottom left panel, the charged particle multiplicity distribution as measured by ATLAS is shown on the bottom right panel.

one. The obtained eigentunes for tunes CP1 and CP1F are listed in the appendix in Tabs. D.2 and D.3 respectively. In order to study the effective uncertainty of the tuning, the predictions of PYTHIA 8 have to be evaluated with every set of eigentune parameters. This leads to the fact that the uncertainties will be very low for some observables where the parameter change does not affect the model predictions much. In other cases the tune uncertainties will be larger since the parameters that influence the specific observable are less constrained. The tune uncertainties are illustrated in Fig. 7.4 where next to the central tune value, the area enclosed by the eigentunes is illustrated with a shaded band. The left panels show the results for tune CP1 and the right panels for tune CP1F. The uncertainties of tune CP1 and tune CP1F are very similar for a variety of different observables.

Within this work, two sets of new parameter tunes of the PYTHIA 8 event generator have been derived with the PROFESSOR software. The tunes use a state-of-the-art set of parton distribution functions at leading order and focus on the modeling of multiparton interactions. It is found that both tunes CP1 and CP1F improve the description of existing measurements provided by PYTHIA 8. Including additional constraints from forward energy measurements slightly changes the obtained tuning results. In most observables, there is no major difference seen between the two tunes. Only in few cases significant differences are observed. Since the CASTOR measurements are not only sensitive to MPI, the effect in including these measurements in a future generator tune can become more significant, when also other parameters of the generator such as parton showers or fragmentation parameters are included in the study.

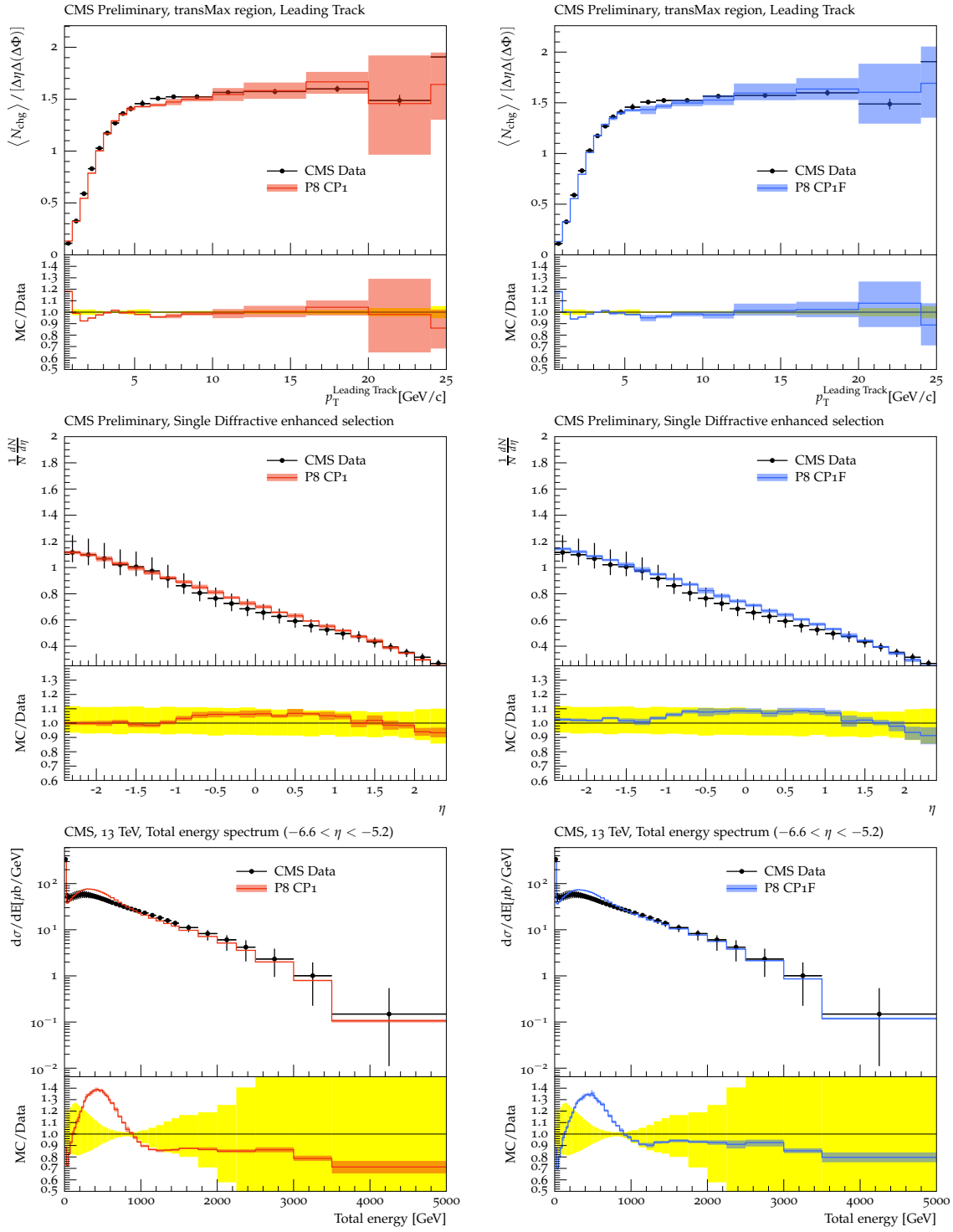



**Fig. 7.4.:** Illustration of the tune uncertainties of CP1 (left panels) and CP1F (right panels) with three example reference measurements: charged particle density in the TransMax region (top), average particle multiplicity $\mathrm{d}N/\mathrm{d}\eta$ with a single diffractive enhanced event selection [140] (middle), and the inclusive CASTOR energy spectrum presented in this work (bottom).

## 7.2 Implications on air shower predictions

As discussed in chapter 2.5, there is a direct link between the energy production in the very forward direction in proton-proton collision and the development of extensive air showers induced by high energy cosmic rays in the Earth's atmosphere. It is therefore of great interest to evaluate to what extend the presented measurements can help to constrain the predictions by commonly used models. While there is no straightforward way to generally apply the data to air shower physics without a complete re-tuning of the used interaction models, some trends can nevertheless be determined.

Ulrich et al. presented simulations studying the influence of some parameters in interaction models on important air shower observables [28]. A ad-hoc technique is used to modify the behavior of hadronic interaction models in air shower simulations. This approach is not equal to a complete tuning of the model since the introduced changes are likely to push the model into a parameter space that is not allowed by the original model. Still, good approximations of how air shower observables change as a function of single modified parameters can be studied. These studies will serve as a baseline to extrapolate the effects seen in proton-proton collisions here.

The uncertainties of model parameters are energy dependent since the models are constrained only at lower energies, leaving some freedom in the extrapolation. Parameters changes can therefore be parameterized in an energy-dependent way as described in Ref. [28]:

$$f\left(E, f_{19}\right) = 1 + \left(f_{19} - 1\right) F\left(E\right) \text{ with} \tag{7.3}$$

$$F\left(E\right) = \begin{cases} 0 & E \leq 10^{15}\text{eV} \\ \frac{\log_{10}\left(E/10^{15}\text{eV}\right)}{\log_{10}(10^{19}\text{eV}/10^{15}\text{eV})} & E > 10^{15}\text{eV} \end{cases} . \tag{7.4}$$

With this definition, the parameter scaling factor $f\left(E, f_{19}\right)$ is unity below $10^{15}\text{eV}$ and reaches $f_{19}$ at $10^{19}\text{eV}$. Proton-proton collision at $\sqrt{s} = 13\,\text{TeV}$ correspond to an energy of $0.9{\times}10^{17}\text{eV}$. With the presented data the corresponding factor $f_{\text{LHC}} = f\left(0.9 \times 10^{17}\text{eV}, f_{19}\right)$ can be determined, from which $f_{19}$ can then be calculated with Eq. (7.3). This way, the results presented in Ref. [28] can be directly applied. Although these reference simulations have been performed with SIBYLL 2.1 it was also shown that the relative changes are similar for various interaction models independent of global offsets. Thus, the results may be applied for different models as well.

### 7.2.1 Shower maximum depth

One of the major open questions in cosmic ray physics is the mass composition of the primary particles at the highest energies. Current measurements by the Pierre Auger

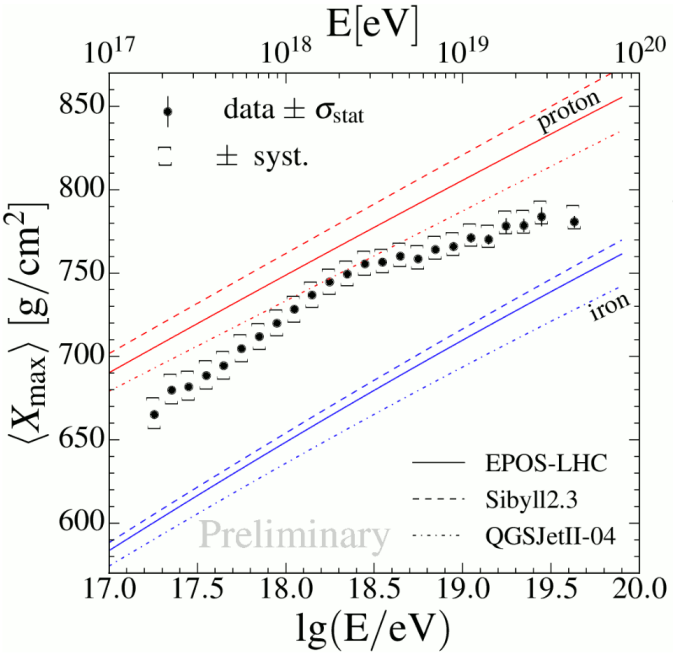



**Fig. 7.5.:** The average of the measured $X_{\max}$ distributions as a function of energy compared to air-shower simulations for proton and iron primaries. Taken from Ref. [141].

Collaboration suggest a proton dominated composition at energies of around $10^{18}$ eV with a tendency towards heavier nuclei at higher energies [141]. The most relevant observable is the depth of the shower maximum in the atmosphere $X_{\max}$ which is expected to be significantly lower for heavy nuclei such as iron. This is shown in Fig. 7.5. The conclusions of the mass composition can only be drawn in comparison to model predictions. Figure 7.5 therefore compares to current interaction models that are tuned to LHC data.

It is one result that EPOS LHC agrees quite well with the measured energy in the forward direction at pseudorapidity $|\eta| \approx 6$. Especially the energy density $\mathrm{d}E/\mathrm{d}\eta$ is in good agreement while the predictions of SIBYLL 2.3 are significantly lower and more importantly in disagreement with the data. This is shown in Fig. 7.6. One reason is that the average collision energy in SIBYLL 2.3 is more concentrated at even higher pseudorapidities $|\eta| > 8$. It is worth noticing that EPOS LHC, SIBYLL 2.1, and QGSJETII.04 all agree well with the data but start to disagree more in the region where no data is available.

The measurement of the CASTOR energy spectrum indicates even further that the fraction of events where no or only very little energy is produced in the direction of CASTOR is significantly overestimated by SIBYLL 2.3. The measurement is shown in Fig. 7.7 with linear scale and zoomed into the low energy region. While EPOS LHC and QGSJETII.04 match the data well, the predictions of SIBYLL 2.3 are about 40-60% too high for $E < 100\,\mathrm{GeV}$.

One possible explanation for these observations is a significantly increased average elasticity (as defined in Eq. (2.10)) in SIBYLL 2.3. This would lead to a particle production that

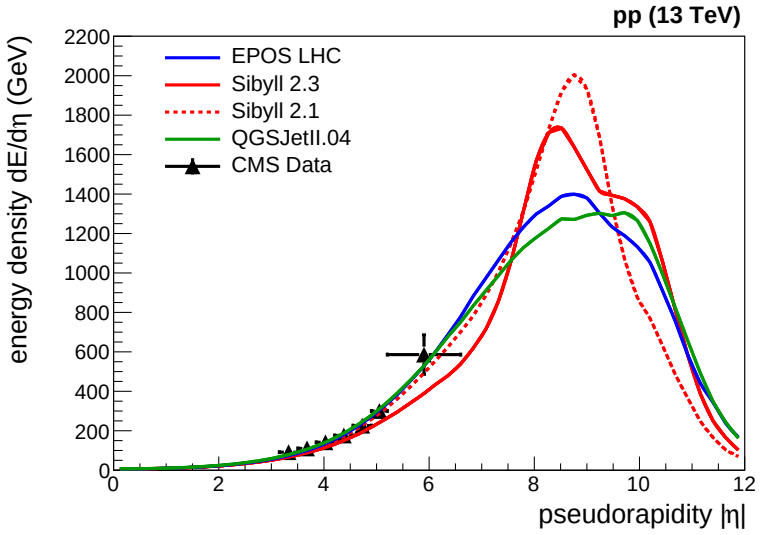



**Fig. 7.6.:** Energy density in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$. The predictions of SIBYLL 2.3 are concentrated at $|\eta| > 8$ leading to a significantly lower energy density at the acceptance of CASTOR at $\eta \approx 6$ compared to EPOS LHC. The CMS data points are taken from Ref. [119].

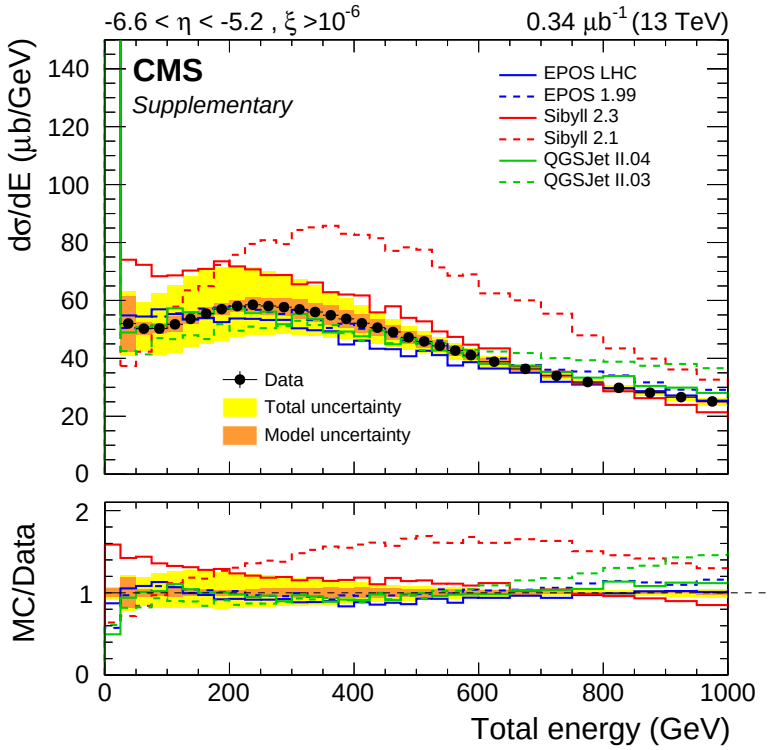



**Fig. 7.7.:** Differential cross section as a function of the total energy in the region $-6.6 < \eta < -5.2$ for events with $\xi > 10^{-6}$ zoomed into the low energy region. Published as supplement material for [118].

is more concentrated in the direction close to the beam rapidity as well as an increased fraction of events with no or only little energy at acceptances shifted away from the beam rapidity such as of CASTOR. Both of these consequences are observed in collision data at the LHC.

These observations indicate therefore that the collision elasticity of SIBYLL 2.3 is significantly higher than that of EPOS LHC (and also QGSJETII.04 and SIBYLL 2.1). This is confirmed by simulations shown in Fig. 7.8, where the elasticity distribution is simulated for proton-proton collisions at $13\,\text{TeV}$ and once for more cosmic ray like collision of a proton with an energy of $10^{19.5}\,\text{eV}$ colliding with a nitrogen target at rest. It can be observed that the elasticity is systematically shifted to larger values for SIBYLL 2.3. Especially, the fraction of events with $\kappa_{\text{el}} \approx 1$ is found to be increased. This corresponds to a larger fraction of events with diffractive or elastic scattering. The average elasticity in proton-proton collision at $\sqrt{s} = 13\,\text{TeV}$ is $\langle \kappa_{\text{el}}^{\text{pp}} \rangle = 0.63$ for SIBYLL 2.3 while it is $0.49$ ($0.57$, $0.53$) for EPOS LHC (SIBYLL 2.1, QGSJETII.04). This increased elasticity however is not supported by the CASTOR data.

The elasticity of the hadronic collisions in an air shower is directly linked to the depth of the shower maximum $X_{\text{max}}$. Under the reasonable assumption that the elasticity of EPOS LHC is confirmed by the data, a decrease of the elasticity by 20% at LHC energies is well justified for SIBYLL 2.3. Although the elasticity of QGSJETII.04 at cosmic ray like energies is also significantly reduced with respect to EPOS LHC, both agree well with the measurement at 13 TeV. The difference between QGSJETII.04 and EPOS LHC is therefore due to different extrapolation to the highest energies, for which no constraints are available. A decrease of 20% corresponds to a scaling factor $f_{\text{LHC}} = 0.8$ which furthermore yields $f_{19} = 0.59$ according to Eq. (7.3). The simulations presented in Ref. [28] demonstrate that the observed value of $f_{19} = 0.59$ for the elasticity corresponds to a decrease of the average shower maximum depth of about $16\,\text{g/cm}^2$. This is more than the difference between EPOS LHC and SIBYLL 2.3 as shown in Fig. 7.5.

In this approach, the presented measurement performed with CASTOR at 13 TeV favor a lower collision elasticity than predicted by SIBYLL 2.3. This implies that the average shower maximum depth $\langle X_{\text{max}} \rangle$ is overestimated. As a consequence, the primary cosmic ray composition is expected to be lighter than what is suggested by comparing air shower measurements to SIBYLL 2.3. The interpretations based on EPOS LHC and QGSJETII.04 on the other hand can not be constrained with the presented data since their agreement is reasonable.

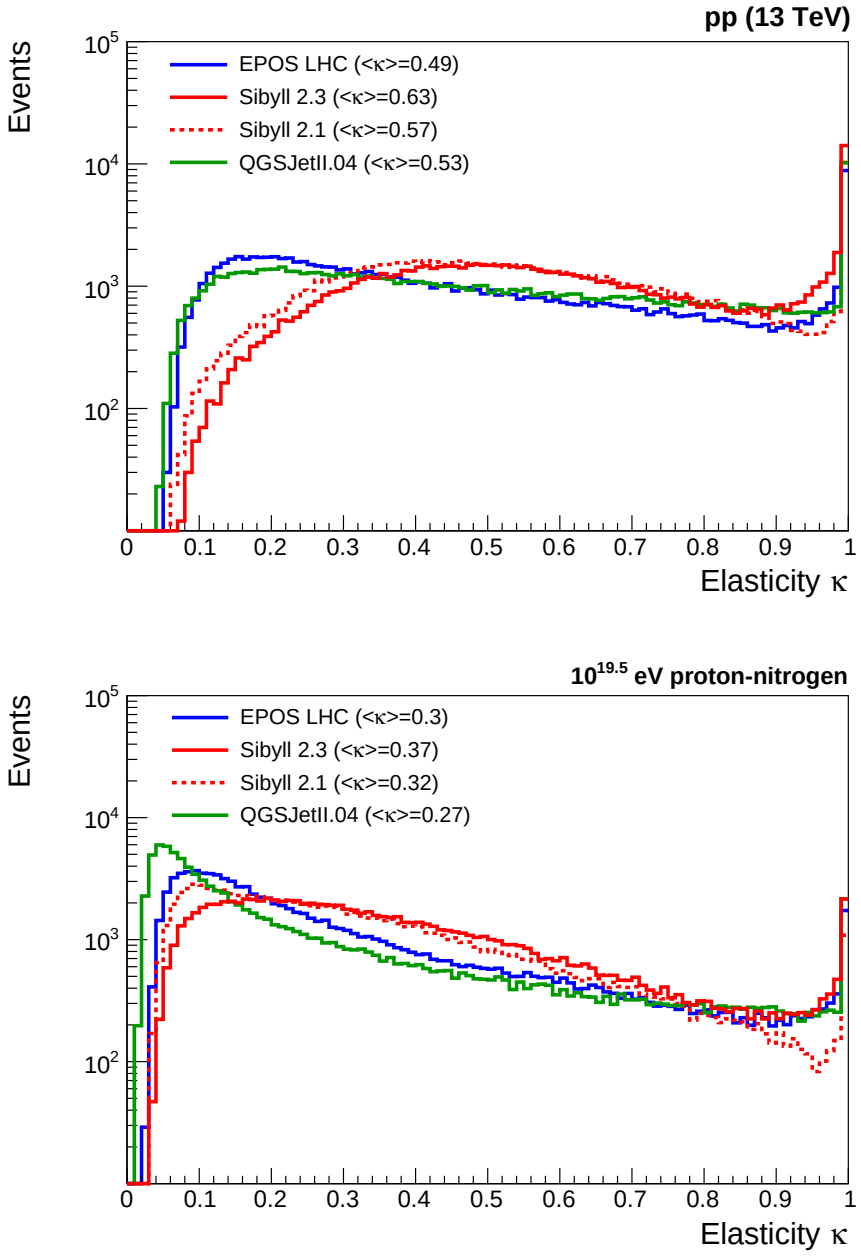



**Fig. 7.8.:** Elasticity distributions for proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ (top) and of protons with energies of $10^{19.5}$ eV with nitrogen at rest (bottom).

### 7.2.2 Muon number in air showers

The muonic component of air showers is until now the least well understood. It is a big challenge for hadronic interaction models to accurately predict the production of muons in a consistent way together with the electromagnetic shower component. For example, a significant excess of muons is observed by the Pierre Auger Observatory at the highest energies that cannot be reproduced by state-of-the art models [23]. A global rescaling of the hadronic shower component (thus indirectly modifying the amount of muons at ground) of 1.33 for EPOS LHC and 1.61 for QGSJETII.04 is needed in order to bring the predictions into agreement with the observed muon abundance [66]. This tension was partly solved with the introduction of SIBYLL 2.3. It incorporates a larger production of baryons and a larger fraction of hadrons that decay into muons, such as the $\rho^0$ [142]. This leads to a significantly increased number of muons in air showers compared to SIBYLL 2.1 and reaches even higher numbers than EPOS LHC and QGSJETII.04, as seen in Fig. 7.9. In contrast to the observations in air showers, so far no such observations were made at colliders, leaving the exact reason for the excess unexplained.

The production of muons in air showers is governed by the production of charged pions and baryons during the primary interaction and during consecutive interactions in later shower stages. At first approximation (as done in the Heitler-Matthews model) this corresponds to the pion charge ratio $c$ as introduced in Eq. (2.11). This is furthermore equivalent to the amount of energy that is given into the production of neutral pions with respect to other hadronic particles. The larger the energy of neutral pions in the interactions, the lower the total amount of muons in the shower. Under the reasonable assumption that the average energy per particle is comparable, this leads to

$$c = \frac{n_{\pi^0}}{n_\pi} \approx \frac{E_{\pi^0}}{E_\pi} \,. \tag{7.5}$$

The CASTOR calorimeter provides an opportunity to study this effect in an unprecedented way. Due to the very forward location a significant part of the total collision energy is measured and can furthermore be separated into the contribution from electromagnetic and hadronic particles. In the presented measurement, the stable particle level was defined in such a way that neutral pions are considered unstable and therefore dominate the electromagnetic energy, while the hadronic energy is dominated by charged pions as well as other stable mesons and baryons. The obtained results can therefore be interpreted in terms of the ratio of average electromagnetic to average hadronic energy per collision (em/had ratio). The average values of the electromagnetic and hadronic CASTOR energy spectra are used to derive this ratio. Based on this measurement, some implications can be drawn on the pion charge ratio $c$.

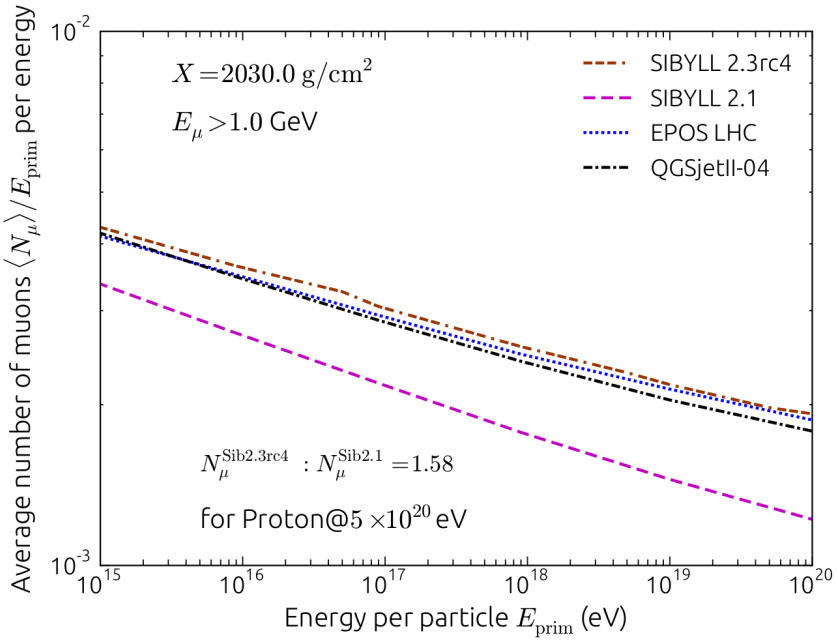



**Fig. 7.9.:** Prediction of the number of muons in proton initiated air showers as a function of the primary energy. Taken from Ref. [142].

For this purpose the observed em/had ratio is extrapolated from the CASTOR acceptance to the full final state of the proton-proton collision.[2] This extrapolation is model dependent and therefore introduces another source of uncertainty. Table 7.4 summarizes the em/had ratio of various event generators in the acceptance of CASTOR as well as the full final state. The derived extrapolation factors are also given therein. Consequently, the em/had ratio for complete proton-proton collisions at 13 TeV can be determined:

$$\frac{\langle E_{\mathrm{em}} \rangle}{\langle E_{\mathrm{had}} \rangle} = 0.276^{+0.022}_{-0.060}\,(\text{experimental}) \pm 0.039\,(\text{extrapolation})\ . \tag{7.6}$$

It is also important to note that the phase space extrapolation uncertainty of 0.039 is driven by pseudorapidities even more forward than CASTOR where no other detectors are available to constrain the measurement. This result is shown in Fig. 7.10 and compared to model predictions. In the limited acceptance of CASTOR it can be observed that all models underestimate the em/had ratio. They predict too small energies by electromagnetic particles compared to hadronic particles. This is especially true for SIBYLL 2.3. The changes introduced with respect to SIBYLL 2.1 that lead to an increased muon number in air showers is also visible at the LHC in the form of a significantly reduced electromagnetic energy. The effect is smaller for EPOS LHC and QGSJETII.04. In the full final state the uncertainties increase due to the model dependent extrapolation such that the predictions of EPOS LHC and SIBYLL 2.1 are compatible with the data within uncertainties, although the general tendency goes towards an underestimation of the electromagnetic energy

[2]Here, the full final state means an acceptance of $-\infty < \eta < \infty$.

**Tab. 7.4.:** Values of the ratio of the average electromagnetic and hadronic energies in CASTOR (left column). Seven interaction models are used to extrapolate the measurement to the full final state of proton-proton collision. This extrapolation is applied to the measurement (bottom row).

| | $\langle E_{\mathrm{em}}\rangle/\langle E_{\mathrm{had}}\rangle$ CASTOR acceptance | Extrapolation coefficient | $\langle E_{\mathrm{em}}\rangle/\langle E_{\mathrm{had}}\rangle$ Full final state |
|---|---|---|---|
| EPOS LHC | 0.386 | 0.587 | 0.227 |
| EPOS 1.99 | 0.357 | 0.653 | 0.233 |
| QGSJETII.04 | 0.41 | 0.495 | 0.203 |
| QGSJETII.03 | 0.424 | 0.536 | 0.228 |
| SIBYLL 2.1 | 0.437 | 0.539 | 0.235 |
| SIBYLL 2.3 | 0.33 | 0.575 | 0.19 |
| PYTHIA 8 CUETP8M1 | 0.385 | 0.618 | 0.238 |
| Data | $0.482\ ^{+0.022}_{-0.060}$ | $0.572 \pm 0.0814$ | $0.276\ ^{+0.041}_{-0.052}$ |

as well. SIBYLL 2.3 and QGSJETII.04 on the other hand undershoot the measurement significantly.

In order to estimate the implications of this finding to the muon number in air showers, a similar method as used for the shower maximum depth is applied. A scaling factor $k$ is derived which would bring in a conservative approach the model prediction to the lower boundary of the experimental uncertainties.[3]

$$k = \left.\frac{\langle E_{\mathrm{em}}\rangle}{\langle E_{\mathrm{had}}\rangle}\right|_{\mathrm{Data}} \Big/ \left.\frac{\langle E_{\mathrm{em}}\rangle}{\langle E_{\mathrm{had}}\rangle}\right|_{\mathrm{Model}} . \tag{7.7}$$

A value of $k = 1.10$ is obtained for QGSJETII.04 and $k = 1.18$ for SIBYLL 2.3. This factor $k$ can be applied to obtain a modified pion charge ratio $c'$ using the assumption of Eq. (7.5):

$$1/c = 1 + \frac{n_{\pi^\pm}}{n_{\pi^0}} \approx 1 + \left.\frac{\langle E_{\mathrm{had}}\rangle}{\langle E_{\mathrm{em}}\rangle}\right|_{\mathrm{Model}} \tag{7.8}$$

$$1/c' = 1 + \left.\frac{\langle E_{\mathrm{had}}\rangle}{\langle E_{\mathrm{em}}\rangle}\right|_{\mathrm{Data}} = 1 + 1/k \times \left.\frac{\langle E_{\mathrm{had}}\rangle}{\langle E_{\mathrm{em}}\rangle}\right|_{\mathrm{Model}} \tag{7.9}$$

$$1/c' = 1 + 1/k\,(1/c - 1) \tag{7.10}$$

[3] One could also argue to bring the model prediction to the central value of the measurement. Since the presented method is rather phenomenological than quantitatively exact, the conservative approach is preferred.

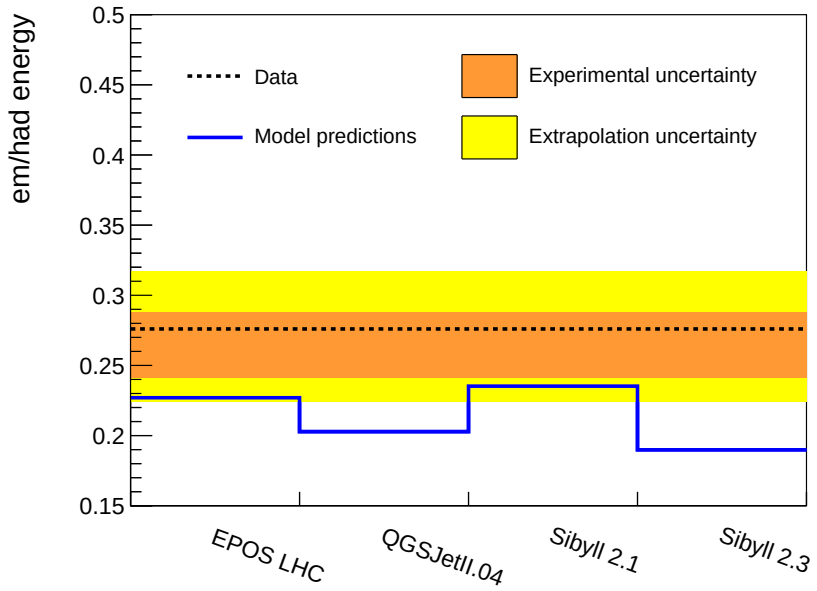



**Fig. 7.10.:** Ratios of the average electromagnetic to hadronic energies in the full final state of proton-proton collision. The data is obtained by extrapolating the measurement with CASTOR. The experimental uncertainty is shown with a yellow band and the extrapolation uncertainty with an orange band. The difference between the model predictions and the lower bound of the uncertainties is used to extract a correction on the pion charge ratio.

Since the exact value of $c$ is model dependent, a value of $c = 1/3$ (which is for example used in the Matthews-Heitler model) is assumed and the modified values of $c'$ are obtained to be:

$$\text{QGSJETII.04}: \; c' = 0.35 = 1.06\, c \tag{7.11}$$

$$\text{SIBYLL 2.3}: \; c' = 0.37 = 1.11\, c\,. \tag{7.12}$$

The simulations presented in Ref. [28] are used to estimate the impact on the average muon number in air showers. The observed increase of the pion charge ratio of 6 (11)% corresponds to modification factors $f_{19} = 1.12(1.23)$ for QGSJETII.04 (SIBYLL 2.3). This change leads to a reduction of the muon content of about 7% and 12% respectively. Given the fact that this estimation is rather conservative, it can be concluded that the muon content of air showers predicted by these models is at least several percent too high. The data even suggests a stronger reduction, but this can at this point not be clearly quantified. The observed muon excess in air showers can therefore not be reconciled with minimum bias data from proton-proton collisions at 13 TeV. They, on the contrary, even indicate that muon numbers predicted by all models are on the upper edge of what the accelerator data supports.

## 7.3 Collective effects in proton-proton collisions with EPOS LHC

Collective effects have been observed in heavy ion collisions with large densities, often referred to as *large* systems, and have also been predicted [143] for and eventually measured in proton-proton collisions (*small* systems) at the LHC (see [144, 145] for detailed reviews). While the existence of collective effects is confirmed in large and small systems, their origin is still unclear. While in large systems the existence of a quark-gluon-plasma (QGP) is commonly assumed, a phase of parton matter where confinement is no longer required [146–148], different explanations such as color reconnection or a hydrodynamic phase are possible for proton-proton collisions [145].

There are multiple different consequences of collective effects such as long-range two-particle correlations, the so-called *ridge* phenomenon [86, 149], jet quenching [150, 151], or enhanced production of strange hadrons [152]. It is also discussed that collective effects can be visible in detailed observation of the underlying event [153]. Studies have therefore been made to check whether or not collective effects might appear in the forward energy measurements performed with CASTOR at 13 TeV.

EPOS LHC is currently the only multipurpose event generator used in high energy physics that includes the description of a QGP-like behavior in high energy collisions. Within EPOS LHC, these effects are realized with a simplified hydrodynamical model in which so-called *clusters* are created in the dense core of the collision that decay afterwards into particles (recall chapter 2.6.2). Within the generator framework of EPOS LHC, all stable ($c\tau > 1\,\text{cm}$) final state particles can be traced back to their origin where they are produced in the collisions: during the fragmentation of strings being created in the hard collision, from the dense core of the collision or the remnant of the beam particles. Figure 7.11 shows the energy density $\mathrm{d}E/\mathrm{d}\eta$ for minimum bias proton-proton collisions at a center-of-mass energy of 13 TeV. The different origin of the energy is indicated by different colors. Next to this, also the relative contribution of the different sources to the total energy density is shown as a function of pseudorapidity. Three regions can be identified:

- The energy density at central pseudorapidities $|\eta| < 5$ is dominated by particles originating in the core,
- at intermediate rapidities $5 < |\eta| < 8$ slightly by particles from strings,
- and at large rapidities $|\eta|$ by the beam remnants.

A more detailed study of the energy produced by the different production mechanisms implemented in EPOS LHC can be performed by separating the energy into the contribution from electromagnetic (only electrons and photons) and hadronic particles. Neutral pions are considered unstable here, thus their decay photons contribute to the electromagnetic component.[4] It is observed that the energies from different production origins also feature

[4] This is the same stable particle definition as used for the data analysis.

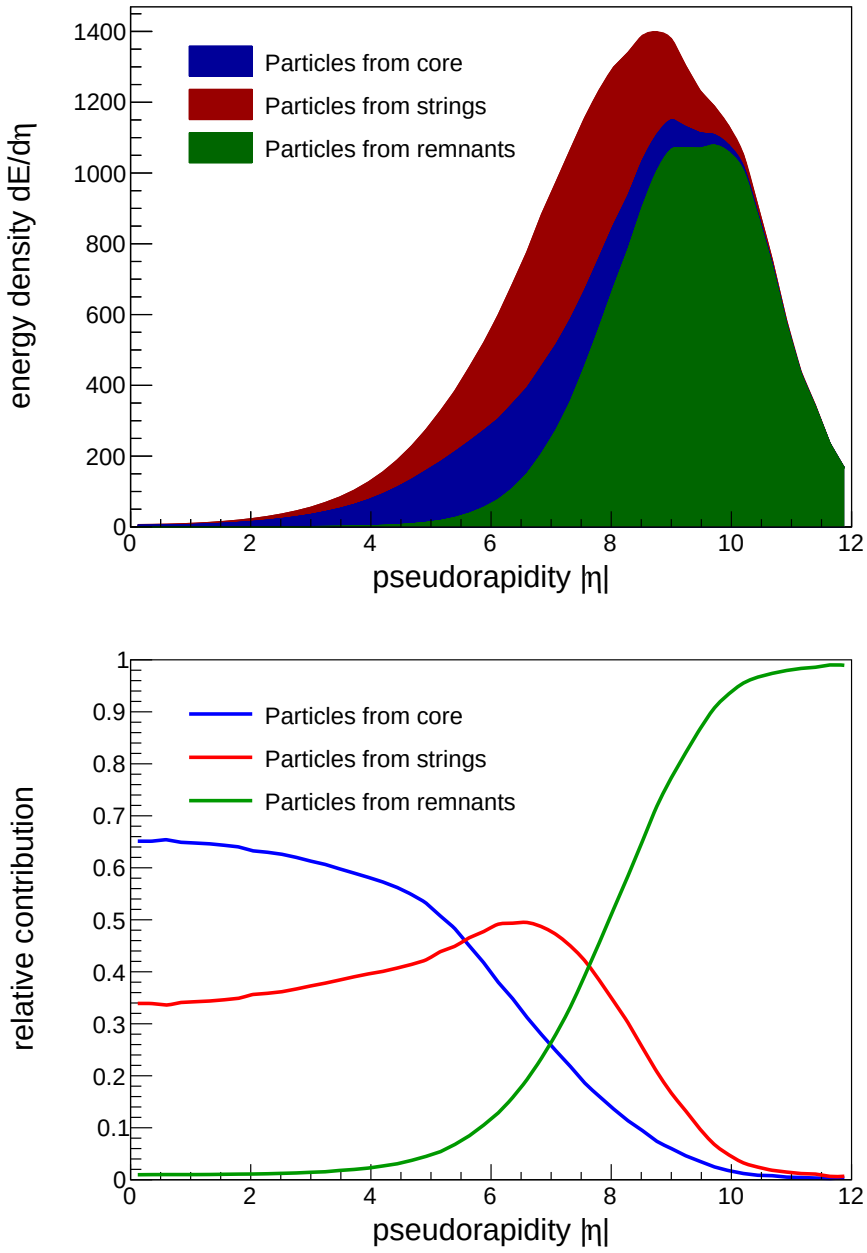



**Fig. 7.11.:** Energy density $\mathrm{d}E/\mathrm{d}\eta$ of proton-proton collisions at 13 TeV simulated with EPOS LHC (top). The contributions from particles originating in the dense core, string fragmentation, beam remnant are indicated with colors. The relative fraction of these contributions to the total energy are shown in the bottom panel.

a different ratio of electromagnetic to hadronic energy. This is shown in Fig. 7.12 for the different production origins as a function of pseudorapidity. Particles from the core, created in the hydrodynamical treatment of the collision, show noticeably less electromagnetic energy than particles originating from string fragmentation. A possible explanations of the reduced ratio is the fact that more baryons and less mesons are produced in the dense environment of the core [84]. Thus, there are significantly less neutral pions and subsequently less electromagnetic energy.

Since the particles are expected to dominate the overall energy density in different regions of phase space, this effect can be seen in the pseudorapidity dependent ratio of electromagnetic to hadronic energy shown in Fig. 7.13. Here for reference also the predictions by other models are shown. At $\eta \approx 0$, the energy density is dominated by the core and therefore the em/had ratio is as low as $0.34$. As the contribution of the core to the total energy decreases with increasing pseudorapidity, also the em/had ratio increases and reaches $0.4$ at $|\eta| \approx 7$ before it decreases rapidly due to the very low electromagnetic contribution in the beam remnants dominating the highest pseudorapidities. Other models that do not consider collective effects such as SIBYLL 2.3 and QGSJETII.04 show a flat em/had ratio over the whole pseudorapidity range. This is confirmed by predictions of EPOS LHC without the collective treatment of the core. Even more striking is the fact that also PYTHIA 8 CUETP8M1 shows this flat behavior. Some of the observables that give a hint towards collective effects in proton-proton collisions can also be described with macroscopic explanations such as the Color Reconnection mechanism in PYTHIA 8. As a consequence the observation of such an increase of the ratio of electromagnetic to hadronic energies would be a strong hint of QGP-like collective effects in proton-proton collisions.

Moreover the described effect should not only be visible as a function of pseudorapidity but also as a function of particle multiplicities. This is due to the fact that collective effects are more pronounced in collisions with high energy and particle densities. This is for example achieved in head-on proton-proton collisions with small impact parameters, which leads to large particle multiplicities. The measurement of the multiplicity dependent average CASTOR energy described in chapter 6.2 provides a possibility to study this effect. The ratio between the distributions of the average electromagnetic and hadronic energies of Figs. 6.27 and 6.28 can be produced to study the ratio of electromagnetic to hadronic energy in the CASTOR acceptance as a function of charged particle multiplicity. The result is shown in Fig. 7.14. The data show that the ratio of the average electromagnetic to the average hadronic energy decreases with multiplicity. This behavior is reproduced very well by EPOS LHC (green line) and PYTHIA 8 CUETP8M1 (red line) although the overall scale is too low. The predictions of QGSJETII.04 are also shown as a reference and do not model the observed decrease correctly, although the overall scale is more consistent with data. The effect of switching off the collective treatment in EPOS LHC is small but still visible. Without collective effects, the decrease is weaker. This follows the expectations assuming that the collective effects become stronger in high multiplicity events leading to a

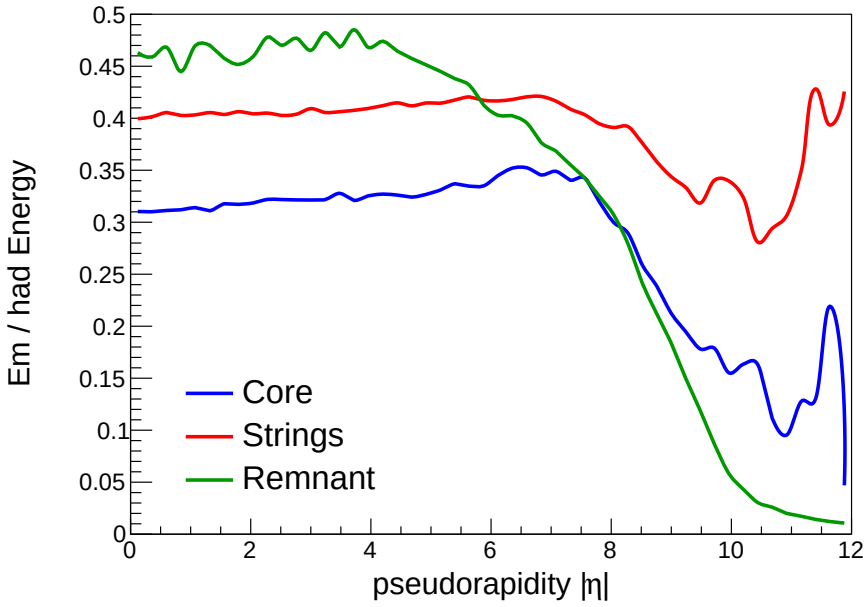



**Fig. 7.12.:** Ratio of electromagnetic to hadronic energy as a function of pseudorapidity simulated with EPOS LHC for proton-proton collisions at 13 TeV. The three production mechanisms in EPOS LHC are shown separately.

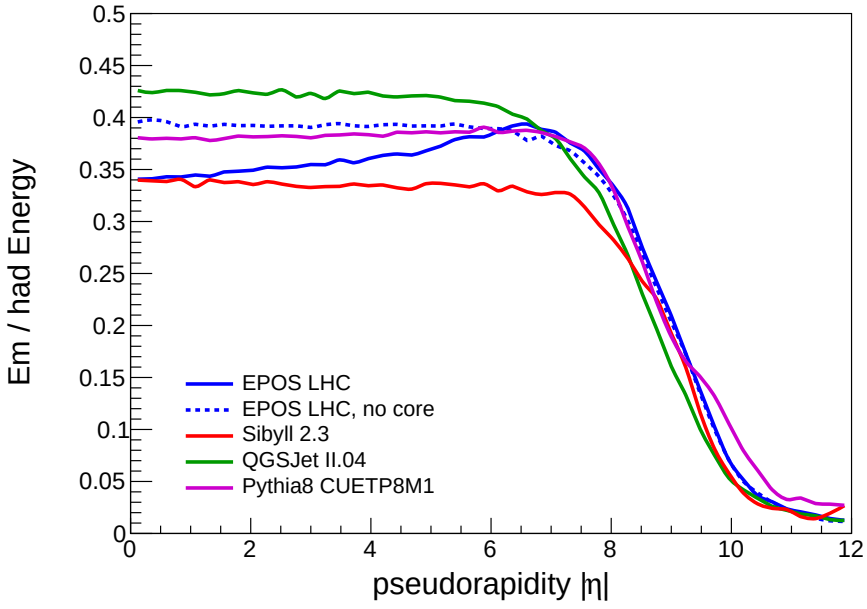



**Fig. 7.13.:** Ratio of electromagnetic to hadronic energy as a function of pseudorapidity for proton-proton collisions at 13 TeV simulated with EPOS LHC (with and without hydrodynamical treatment of the dense core, solid and dashed blue line respectively) as well as SIBYLL 2.3 (red), QGSJETII.04 (green) and PYTHIA 8 tune CUETP8M1 (magenta).

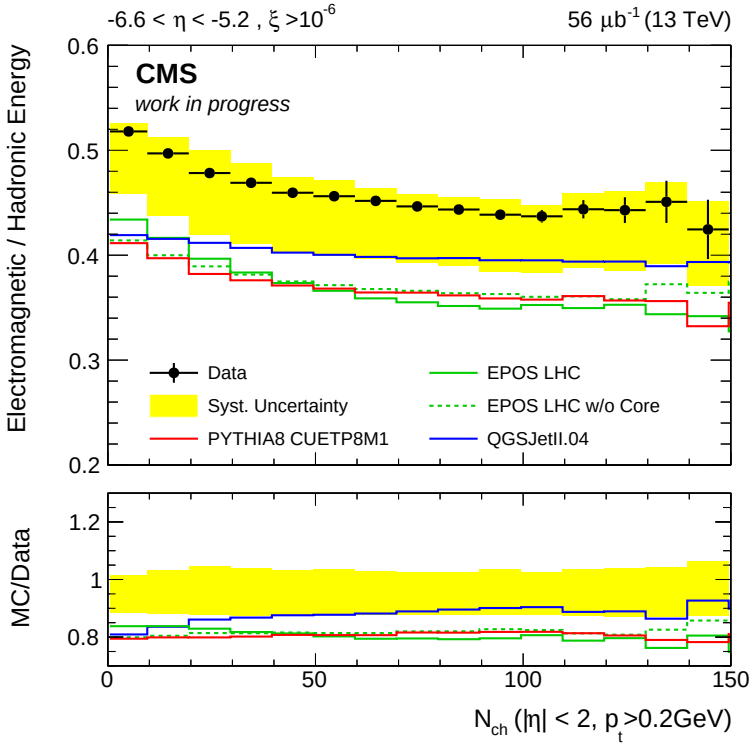



**Fig. 7.14.:** Multiplicity dependent ratio of electromagnetic to hadronic energy. The data points are obtained by dividing the measured average electromagnetic and hadronic energies shown in Figs. 6.27 and 6.28. This is compared to predictions of EPOS LHC (with and without hydrodynamical treatment of the core), PYTHIA 8 tune CUETP8M1 and QGSJETII.04.

smaller em/had ratio. The fact that the observed effect is small agrees with the observation from Fig. 7.11 that the particles produced in the core only partially contribute to the total energy measured in the acceptance of CASTOR. Given the systematic uncertainties, it is not possible to determine whether or not the data supports collective effects as implemented in EPOS LHC.

The presented data alone do not provide enough information to reach a final conclusion on the presence of QGP-like states in proton-proton collisions at 13 TeV and whether or not they are visible in the ratio of electromagnetic to hadronic energy. Nevertheless an interesting new way to probe collective effects is demonstrated. A complementary measurement at central rapidities – which remains to be performed – will reveal the nature of these collisions.

# 8 Summary

In this thesis unique measurements of multiparticle production have been performed with proton-proton collisions in the very forward region at the LHC and implications for hadronic interaction models have been derived. For this purpose data taken with the CASTOR calorimeter of the CMS experiment were analyzed. With an acceptance of $-6.6 < \eta < -5.2$ CASTOR is the most forward calorimeter sensitive to charged as well as neutral particles at the LHC and is a unique tool to study this poorly known phase space.

In preparation of the data taking, measures have been taken to improve the detector performance. It was found in earlier operations that the position of CASTOR with respect to the CMS reference frame needs to be well understood. Uncertainties in the position measurement contribute significantly to the overall experimental uncertainty of the calorimeter. Therefore the alignment system of CASTOR was revisited and improved. A new calibration method was designed for the infrared distance sensors. The main features of this method and improvements to the detector alignment are:

- A realistic mock-up of the LHC beam pipe serves as reference target for the distance sensors. In this way the attenuation of the reflected light as a function of the target curvature is studied.
- A fully automated two-dimensional positioning system is developed to place the sensors in front of the target. The positions of the sensors can be precisely set and individual calibrations of the infrared sensors can be obtained.
- A dedicated software tool provides possibilities to record the response functions of individual sensors in a user-specified as well as fully automatic way. The time to record a full response function was so reduced to about eight hours per sensor.
- Subsequently, a position fit of CASTOR was performed combining data from all available alignment systems.
- With this new calibration, the position of CASTOR was determined with unprecedented accuracy. As a consequence the uncertainty of the CASTOR energy scale due to the position uncertainty was reduced from 16 to 7.5% thus reducing the total uncertainty from 22 to 17%.

In June 2015 a unique dataset at the currently largest accessible center-of-mass energy of 13 TeV was recorded. Due to the special circumstances during this period, extensive

efforts were taken to ensure a good quality of the recorded data. Large sets of Monte Carlo simulations were produced for this purpose, taking the measured position of CASTOR as well as the overall data taking conditions into account. These simulations are widely used and have served as a reference for various physics analyses and publications.

The energy distribution within the acceptance of CASTOR was studied in the scope of this thesis. The total energy deposited in CASTOR was measured for minimally biased events and fully corrected for detector effects with an unfolding technique. The distribution was evaluated in terms of the differential cross section as a function of the energy, $\mathrm{d}\sigma/\mathrm{d}E$. The main results are:

- The spectral shape is sensitive to the modeling of multiparton interactions (MPI). The predictions of PYTHIA 8 change when MPI-related parameters are altered and become incompatible with the data without including MPI.
- Models used for air shower simulations agree generally better with the data in their newer versions that are already tuned to LHC data at lower energies.
- While the general agreement is improved in SIBYLL 2.3 compared to the previous version 2.1, a significantly increase of the soft part of the spectrum is observed.
- The design of CASTOR was furthermore used to separate the electromagnetic from the hadronic component. This was done for the first time and provides important insight into production mechanisms of different particle species.
- The average energy was determined and included in the measurement of the energy density as a function of pseudorapidity $\mathrm{d}E/\mathrm{d}\eta$, extending the forward acceptance up to $|\eta| < 6.6$.
- The main results have been published in the *Journal of High Energy Physics*.

In addition, the average collision energy in CASTOR was correlated to the activity in the central region of CMS and evaluated in terms of the charged particle multiplicity within $|\eta| < 2$. The main features and results are:

- Due to the lack of a magnetic field in the available dataset, a modified track reconstruction algorithm was used to determine the multiplicity of charged particles in the acceptance of the CMS pixel tracker.
- Also in this case the total as well as the electromagnetic and hadronic contributions were evaluated separately.
- A turn-on curve of increasing energy with increasing multiplicity, which is characteristic for the underlying event, was observed.
- The data do not indicate a flat plateau region at high multiplicities, in contrast to the predictions of EPOS LHC and QGSJETII.04.
- SIBYLL 2.3 underestimates significantly the electromagnetic energy.

The energy carried by particles into the acceptance of CASTOR is largely dominated by soft interaction processes such as multiparton interactions or initial state radiation. For this reason, the data offer a unique opportunity to be evaluated in terms of their implications for hadronic interaction models which rely on phenomenological descriptions of these processes and need validation with data. Two new parameter tunes of the PYTHIA 8 event generator were obtained using the PROFESSOR software:

- The tune CP1 continues the tuning effort within the CMS collaboration to improve the description of underlying event measurements. The tune CP1 is expected to become one of the main reference tunes for future physics analyses at 13 TeV in CMS.
- The data obtained with CASTOR are sensitive to some of the tuned parameters and were therefore included in a dedicated tune, CP1F.
- It was found that the description of measurements at 13 TeV was significantly improved with these new tunes.
- The CASTOR data confirm well the parameters found in tune CP1.
- Given the limited number of parameters tuned in this effort, the full potential of the presented data can be exploited in a future, more extended parameter tune or model improvements.

It is known that the forward particle production has crucial impact on the development of cosmic ray induced air showers. The data were therefore also evaluated in terms of their implications for air shower predictions. The difference between the data and the model predictions were quantified and the effect of this discrepancy on air shower observables was estimated based on simulations. The main conclusions from this study are:

- The collision elasticity of SIBYLL 2.3 was found to be too large. This indicates that the average shower depth is also overestimated. In consequence, this translates to a lighter primary cosmic ray mass composition.
- The measured ratio of electromagnetic to hadronic energy indicate that the predicted number of muons in air showers is overestimated by QGSJETII and SIBYLL 2.3. The tension between the measured muon numbers and model predictions remains.
- A reasonable agreement of EPOS LHC with the presented data is found and therefore no immediate constraints can be derived.
- It was shown that the data provide powerful constraints for future cosmic ray generator model improvements.

It was also discovered that the ratio of the electromagnetic to the hadronic energy in CASTOR gives hint to a possible, until now unexplored, consequence of collective effects in proton-proton collisions. EPOS LHC predicts different ratios in different acceptances as well as in different collision topologies. Further measurements are proposed to confirm the possible existence of a quark-gluon-plasma-like state in proton-proton collisions.

In total, this thesis extends the ongoing efforts to measure and constrain theoretically poorly known multiparticle production in dominantly soft interaction processes at the LHC. Dedicated data taken during proton-proton collisions at 13 TeV with the CASTOR calorimeter are analyzed and interpreted for their implications. The full potential of CASTOR is for the first time exploited by providing fully calibrated and corrected energy measurements and also separating the electromagnetic and hadronic contribution. Moreover it was shown that the data has direct impact on a large variety of applications such as the modeling of the underlying event, predictions of air showers, and even collective effects in proton-proton collisions. In this combination, it demonstrates the enormous potential of very forward measurements at the LHC.

# Appendices

# A

# Additional material on the CASTOR alignment

## A.1 List of all calibrated infrared sensors

Below, two list of all infrared sensors is given that have been calibrated with the automated two-dimensional calibration setup. The first list includes the sensors used during the June 2015 data taking period of 13 TeV proton-proton collisions. The second list contains the sensors for the November 2016 data taking period of 5 and 8 TeV proton-lead collisions.

| sensor name | installation location | comment |
|---|---|---|
| IR633 | IP near top | |
| IR637 | IP near bottom | |
| IR648 | IP far top | |
| IR638 | IR far bottom | |
| IR612 | non-IP near top | |
| IR616 | non-IP near center | |
| IR620 | non-IP near bottom | |
| IR630 | non-IP far top | |
| IR625 | non-IP far center | |
| IR621 | non-IR far bottom | |
| IR611 | RS near top | not used for alignment |
| IR603 | RS near center | not used for alignment |
| IR605 | RS near bottom | not used for alignment |
| IR606 | RS far top | not used for alignment |
| IR608 | RS far center | not used for alignment |
| IR607 | RS far bottom | not used for alignment |
| IR511 | not installed | spare sensor |

| sensor name | installation location | comment |
|---|---|---|
| IR788 | IP near top | |
| IR765 | IP near bottom | |
| IR792 | IP far top | |
| IR772 | IR far bottom | |
| IR769 | non-IP near top | |
| IR758 | non-IP near center | |
| IR800 | non-IP near bottom | |
| IR786 | non-IP far top | |
| IR796 | non-IP far center | |
| IR770 | non-IR far bottom | |
| IR789 | RS near top | not used for alignment |
| IR787 | RS near center | not used for alignment |
| IR755 | RS near bottom | not used for alignment |
| IR790 | RS far top | not used for alignment |
| IR795 | RS far center | not used for alignment |
| IR798 | RS far bottom | not used for alignment |
| IR797 | not installed | spare sensor |
| IR756 | not installed | spare sensor |

## A.2 CASTOR alignment for the 2015 and 2016 lead-lead and proton-lead data taking

The position of the CASTOR face is fitted relative to the LHC beam pipe using the infrared sensors. The results are shown in Fig. A.1. The IR sensors for the 2015 lead-lead data taking period have not been calibrated with the automated calibration setup, therefore the obtained results are less precise than the ones for the 2015 proton-proton data taking. During the installation of CASTOR for the 2016 proton-lead runs, the readout electronics for the sensors were faulty, the obtained results therefore need to be considered with care.

The estimation of the position of CASTOR with respect to the CMS reference frame is obtained by combining the results from all different sensor technologies and are found to be:

$$\begin{aligned} \text{2015 lead-lead:}\ \ \text{Near Half } (x,y)/\,\text{mm} &= (-4.90 \pm 2.0, -1.97 \pm 2.0) \\ \text{Far Half } (x,y)/\,\text{mm} &= (-7.45 \pm 2.0, -5.50 \pm 2.0) \end{aligned} \tag{A.1}$$

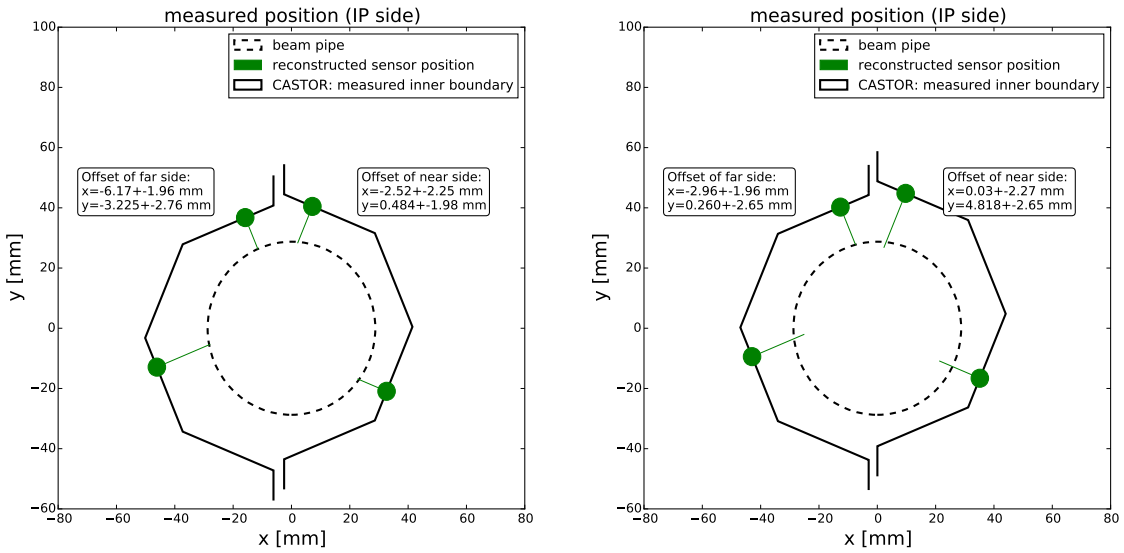



**Fig. A.1.:** Fit results of the IP side of CASTOR with respect to the beampipe for the 2015 lead-lead data taking period (left) and the 2016 leadl-lead data taking (right). The fits are obtained by using the IR sensors only.

$$\begin{aligned} \text{2016 proton-lead: Near Half } (x,y)/\,\text{mm} &= (-3.88 \pm 2.9, -1.44 \pm 2.9) \\ \text{Far Half } (x,y)/\,\text{mm} &= (-6.33 \pm 2.9, -6.03 \pm 2.9) \end{aligned} \tag{A.2}$$

# Additional material for the 2016 CASTOR data certification

B

The below table summarizes the results of the data certification campaign during the 2016 proton-lead data taking period. Only runs are shown where CASTOR and the rest of CMS are certified to have good data quality.

| Era | Run number | energy | beam to CASTOR | luminosity [1/mub] |
|---|---|---|---|---|
| HICollisions2016B | 285369 | 5TeV | p | 44.9 |
| | 285371 | 5TeV | p | 28.6 |
| | 285374 | 5TeV | p | 3.6 |
| | 285383 | 5TeV | p | 69.3 |
| HICollisions2016C | 285505 | 8TeV | p | 1714 |
| | 285517 | 8TeV | p | 9889 |
| | 285530 | 8TeV | p | 12154 |
| | 285537 | 8TeV | p | 3071 |
| | 285538 | 8TeV | p | 1245 |
| | 285539 | 8TeV | p | 5833 |
| | 285549 | 8TeV | p | 3727 |
| | 285684 | 8TeV | p | 5193 |
| | 285718 | 8TeV | p | 2540 |
| | 285726 | 8TeV | p | 3653 |
| | 285750 | 8TeV | p | 10829 |
| | 285759 | 8TeV | p | 4271 |
| | 285832 | 8TeV | p | 641 |
| HICollisions2016C | 285956 | 8TeV | Pb | 2.6 |
| | 285975 | 8TeV | Pb | 1232 |
| | 285993 | 8TeV | Pb | 6898 |
| | 285994 | 8TeV | Pb | 104 |
| | 285995 | 8TeV | Pb | 267 |
| | 286009 | 8TeV | Pb | 2365 |
| | 286010 | 8TeV | Pb | 6340 |
| | 286023 | 8TeV | Pb | 7028 |
| | 286031 | 8TeV | Pb | 2861 |

| | | | | |
|---|---|---|---|---|
| | 286033 | 8TeV | Pb | 868 |
| | 286034 | 8TeV | Pb | 4187 |
| | 286051 | 8TeV | Pb | 3874 |
| | 286054 | 8TeV | Pb | 4161 |
| | 286069 | 8TeV | Pb | 3841 |
| | 286070 | 8TeV | Pb | 782 |
| | 286178 | 8TeV | Pb | 5147 |
| | 286200 | 8TeV | Pb | 1666 |
| | 286201 | 8TeV | Pb | 2231 |
| | 286288 | 8TeV | Pb | 632 |
| | 286301 | 8TeV | Pb | 5799 |
| | 286302 | 8TeV | Pb | 2981 |
| | 286309 | 8TeV | Pb | 85 |
| | 286314 | 8TeV | Pb | 673 |
| | 286327 | 8TeV | Pb | 393 |
| | 286329 | 8TeV | Pb | 1923 |
| | 286365 | 8TeV | Pb | 7781 |
| | 286420 | 8TeV | Pb | 182 |
| | 286422 | 8TeV | Pb | 190 |
| | 286424 | 8TeV | Pb | 0.8 |
| | 286425 | 8TeV | Pb | 9175 |
| | 286441 | 8TeV | Pb | 125 |
| | 286442 | 8TeV | Pb | 6536 |
| | 286450 | 8TeV | Pb | 10642 |
| | 286471 | 8TeV | Pb | 10066 |
| | 286496 | 8TeV | Pb | 5053 |
| HICollisions2016D | 286516 | 5TeV | p | 26.3 |
| | 286517 | 5TeV | p | 8.3 |
| | 286518 | 5TeV | p | 4.7 |
| | 286519 | 5TeV | p | 30.9 |
| | 286520 | 5TeV | p | 25.7 |

# C Additional material for the CASTOR energy measurements

## C.1 Inclusive CASTOR energy spectra at low energies

The inclusive energy spectra recorded with CASTOR are shown in Figs. 6.10–6.11 with a logarithmic scale. The same spectra are shown below in Fig. C.1 with linear scale and zoomed into the low energy region.

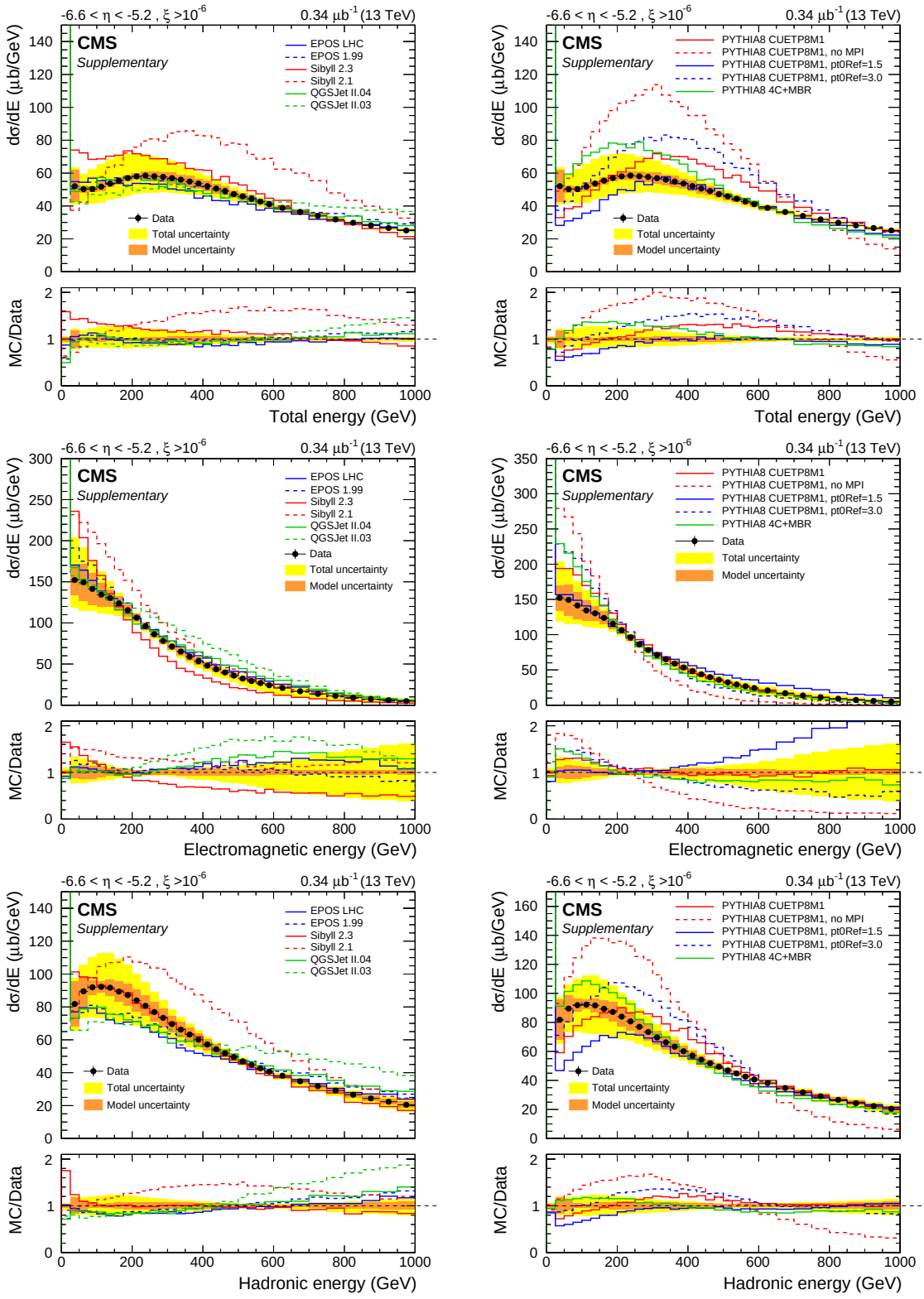



**Fig. C.1.:** Differential cross section as a function of the total (top) electromagnetic (middle) and hadronic energy (bottom) in the region $-6.6 < \eta < -5.2$ for events with $\xi > 10^{-6}$. The left panel shows the data compared to MC event generators mostly developed for cosmic ray induced air showers, and the right panel to different PYTHIA 8 tunes. Zoomed into the low energy region. Published as supplement material for Ref. [118].

## C.2 Validation of the line tracking algorithm

The line tracking algorithm described in chapter 6.2.1 was validated with Monte Carlo simulations based on PYTHIA 8 CUETP8M1. In Figs. C.2–C.7 a variety of distributions is shown that demonstrate that the presented procedure works well to identify charged particle tracks even without magnetic field. The distributions of reconstructed tracks as a function of the azimuth $\phi$ or the pseudorapidity $\eta$ match well with data and simulations. Only at higher values of $|\eta|$, the agreement deteriorates. This is due to the fact that only one layer of the pixel barrel covers the region beyond $|\eta| < 2$ (see Fig. C.2). As a consequence, the tracking algorithm is considered stable and well understood only for $|\eta| < 2$. The distributions of the reconstructed vertices (Figs. C.5–C.7) show reasonable agreement between data and simulations. A slight shift in x is found of the order of 0.1 mm. This effect is small and does not impact the tracking performance at an observable level.

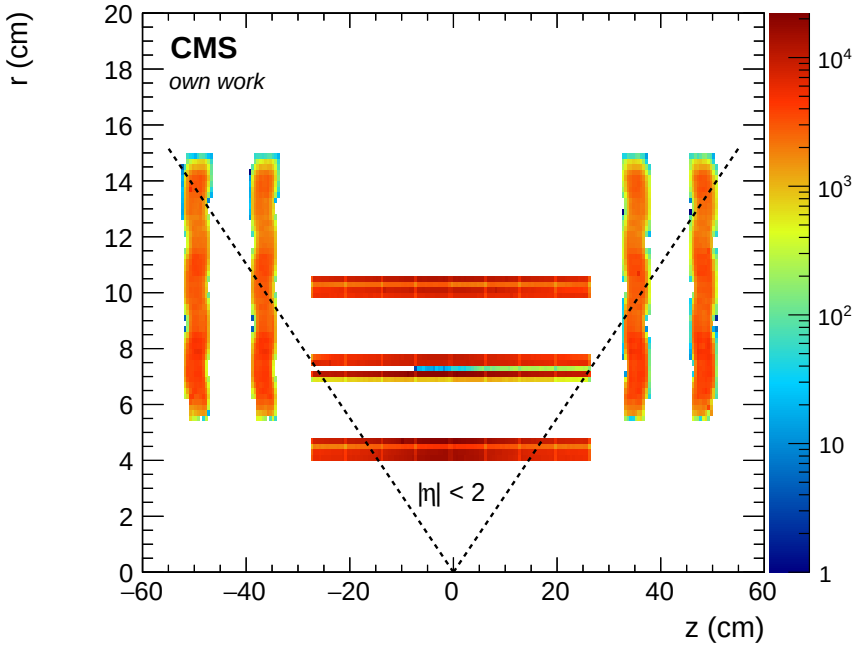



**Fig. C.2.:** Location of reconstructed pixel hits in the $r - z$-plane of CMS. The layered structure of the pixel detector as shown in Fig. 3.3 can be identified.

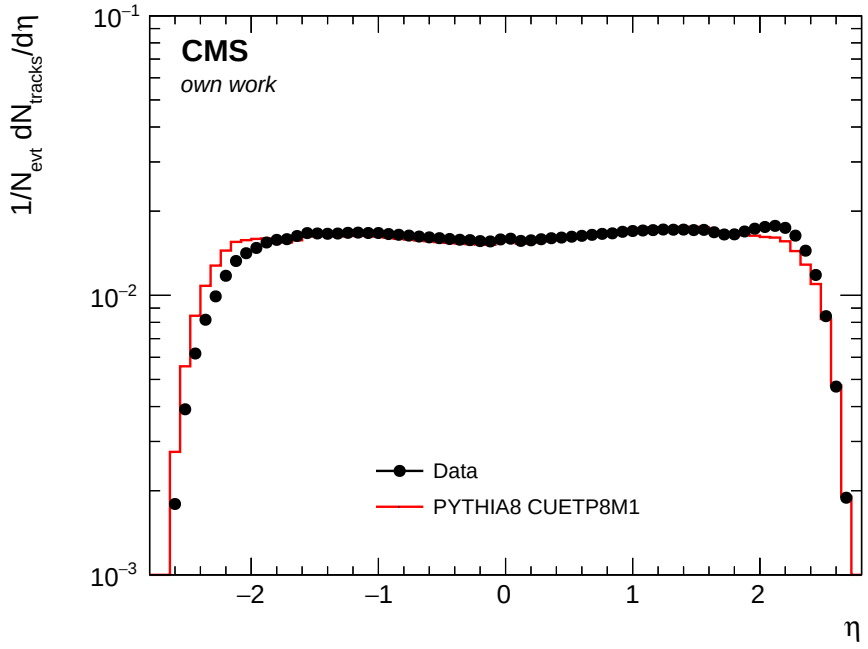



**Fig. C.3.:** Pseudorapidity distribution of particle tracks reconstructed with the CMS pixel detector for both data (solid black markers) and Monte Carlo simulations (red line). A good agreement is observed for tracks with $|\eta| < 2$.

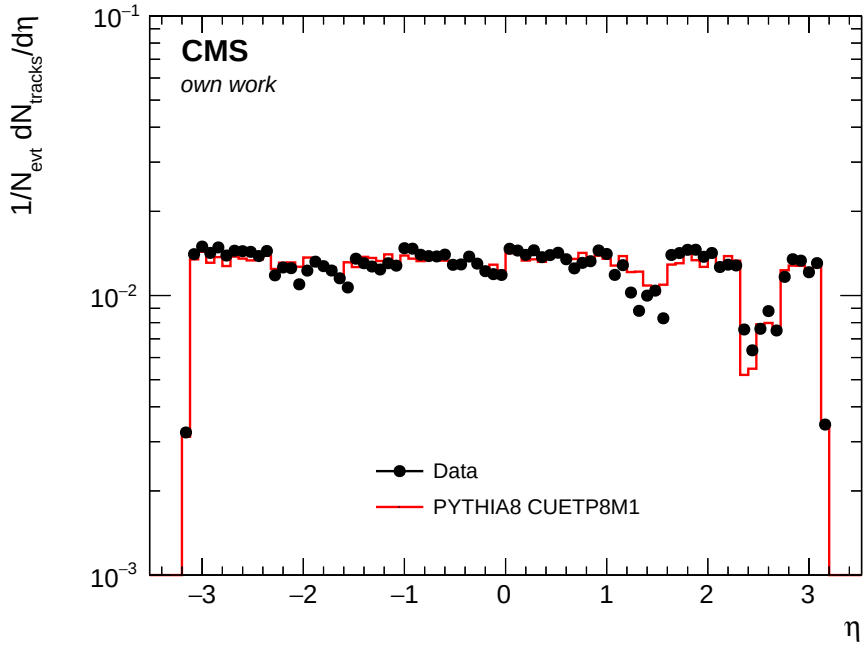



**Fig. C.4.:** Azimuthal distribution of particle tracks reconstructed with the CMS pixel detector for both data (solid black markers) and Monte Carlo simulations (red line). A good agreement is observed over the whole azimuthal range.

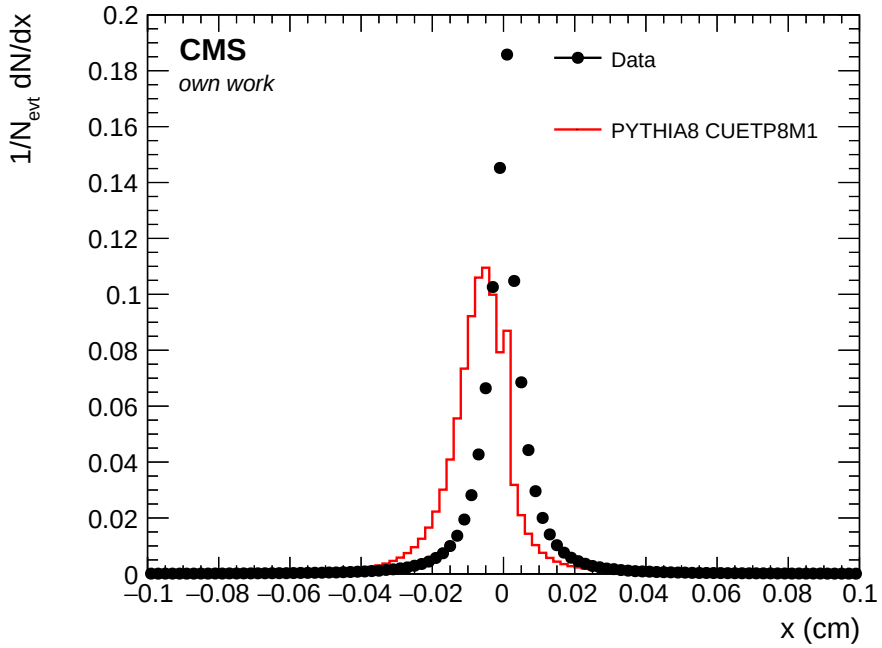



**Fig. C.5.:** Position of the reconstructed vertices in the x coordinate of the CMS reference frame for both data (solid black markers) and Monte Carlo simulations (red line). A slight mismatch of about 0.1 mm is observed. This mismatch does not impact the tracking efficiency in any significant way.

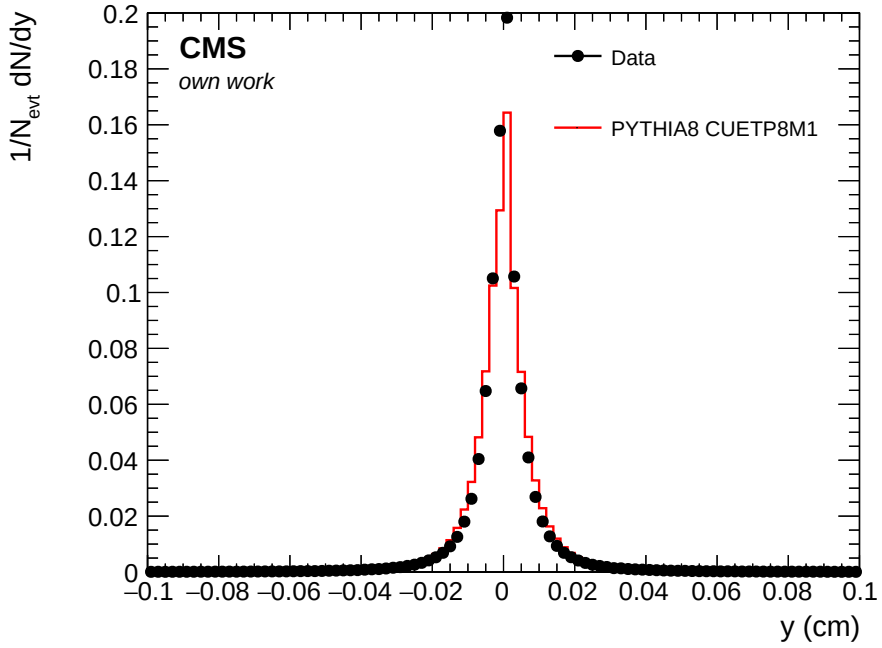



**Fig. C.6.:** Position of the reconstructed vertices in the y coordinate of the CMS reference frame for both data (solid black markers) and Monte Carlo simulations (red line). A very good agreement between data and simulations is observed.

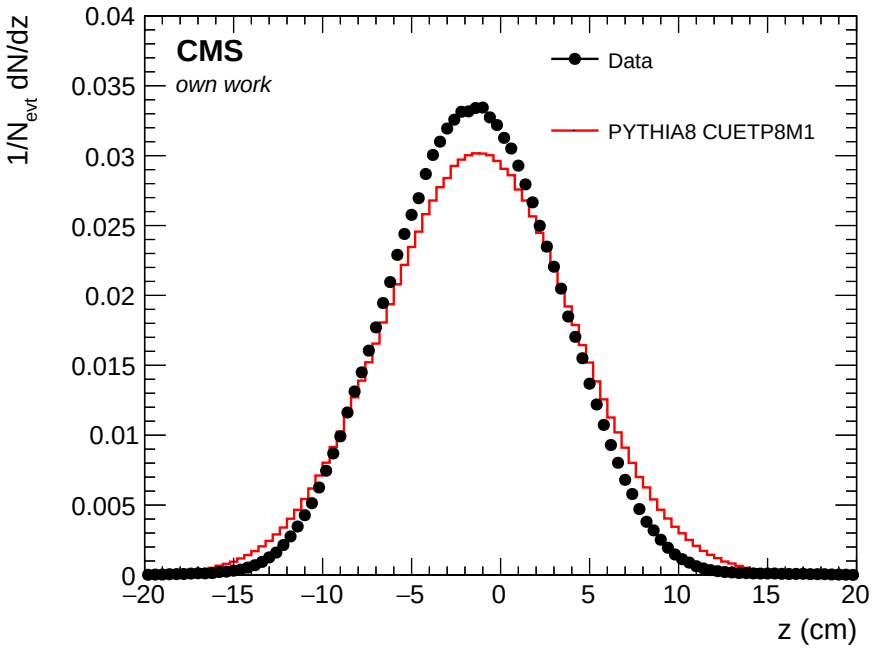



**Fig. C.7.:** Position of the reconstructed vertices in the z coordinate of the CMS reference frame for both data (solid black markers) and Monte Carlo simulations (red line). A very good agreement between data and simulations is observed.

## C.3 Systematic uncertainties of the multiplicity dependent CASTOR energy

Figures C.8–C.10 show the details of single contributions to the total systematic uncertainty on the measurement of the multiplicity dependent CASTOR energy for the total, electromagnetic and hadronic energies respectively.

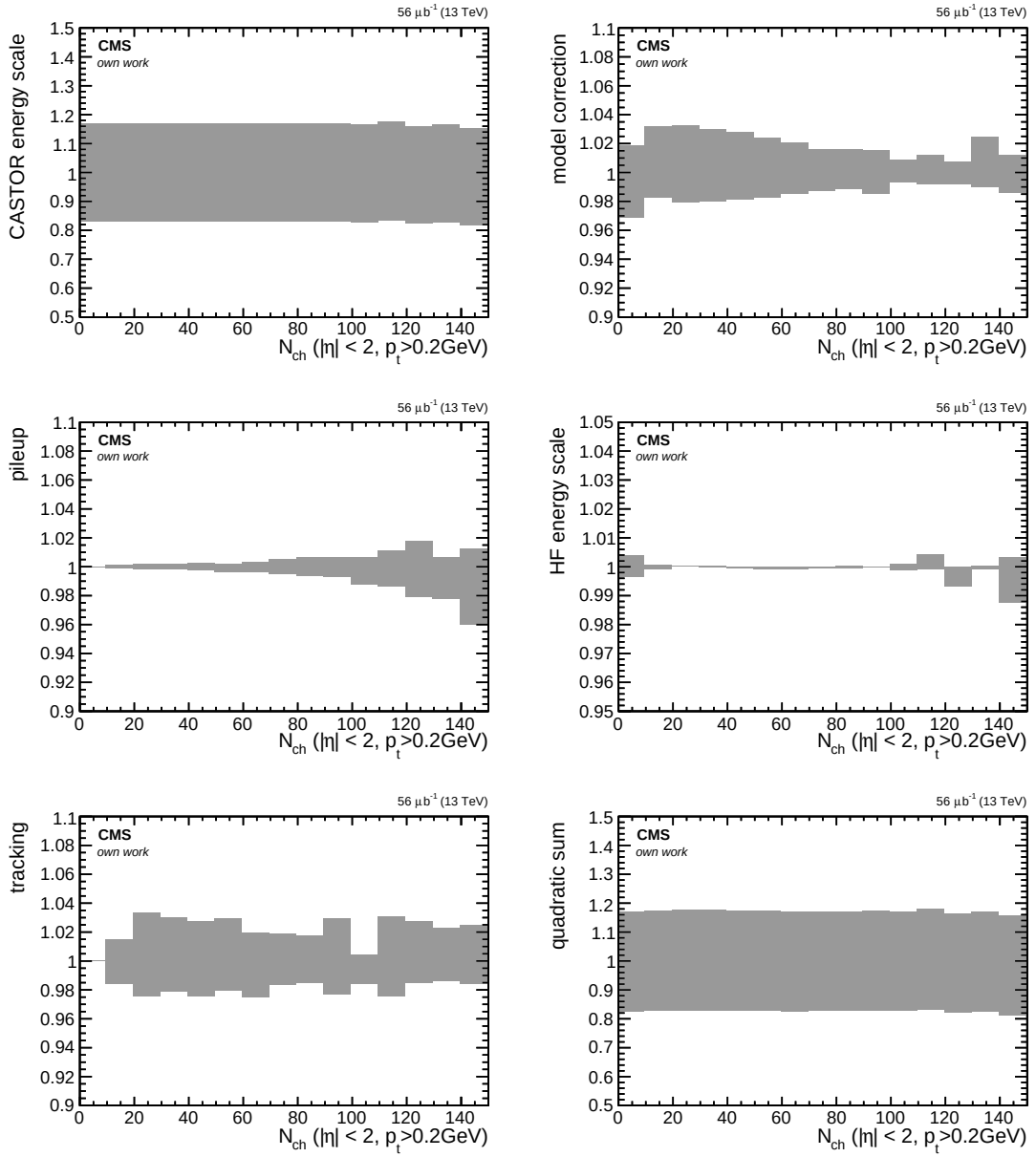



**Fig. C.8.:** Contributions to the total systematic uncertainty of the total multiplicity dependent CASTOR energy.

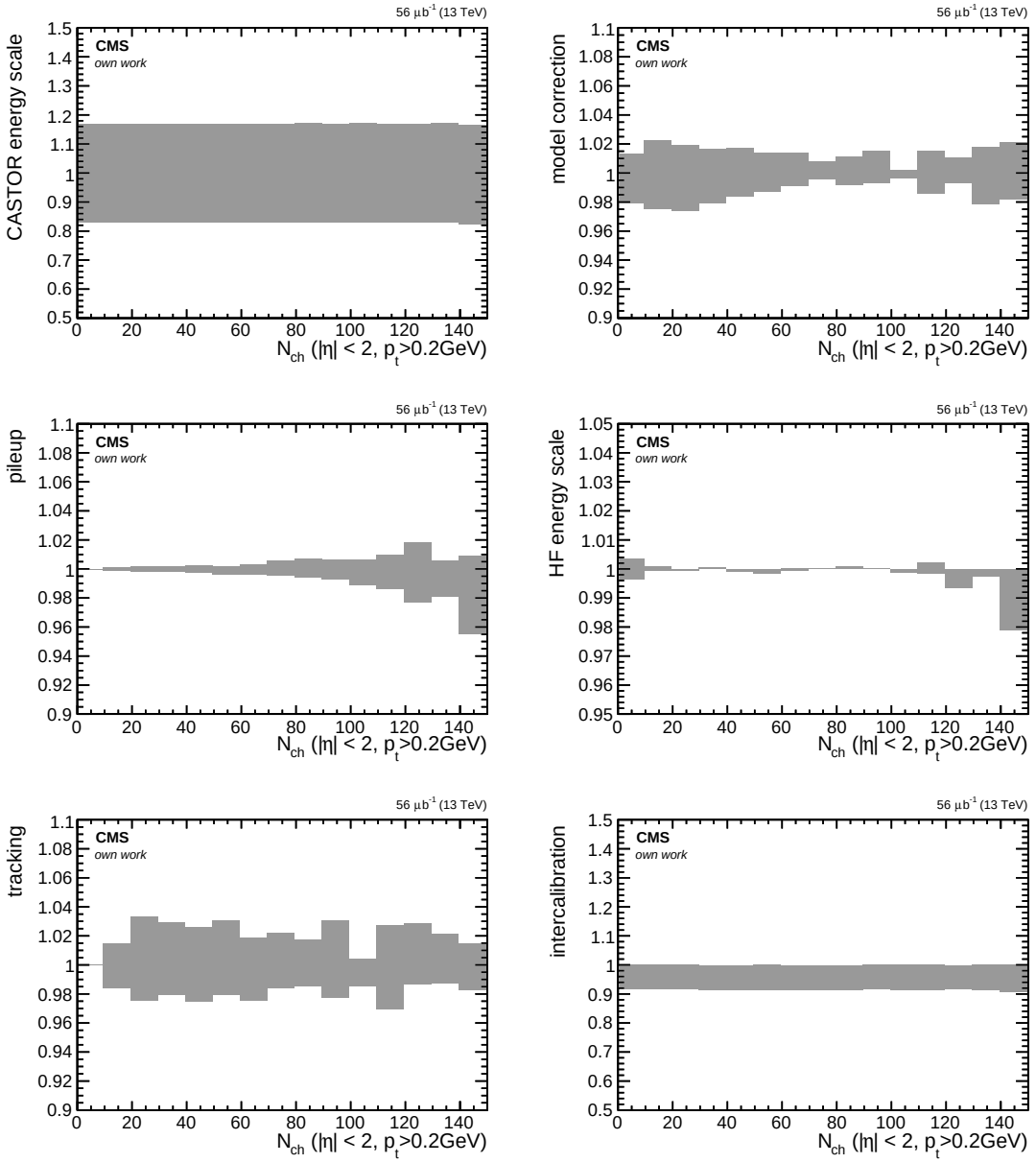



**Fig. C.9.:** Contributions to the total systematic uncertainty of the electromagnetic multiplicity dependent CASTOR energy.

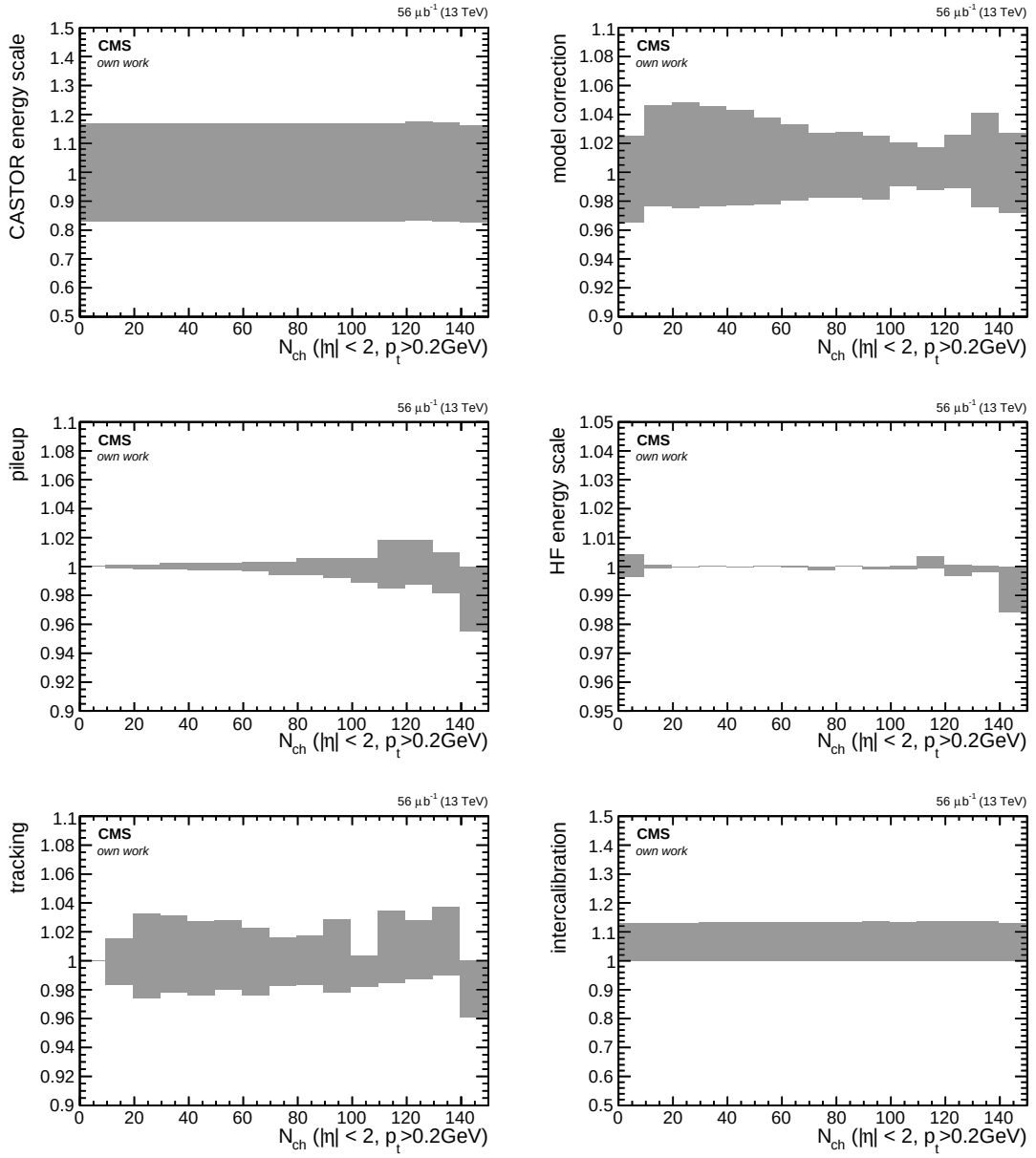



**Fig. C.10.:** Contributions to the total systematic uncertainty of the hadronic multiplicity dependent CASTOR energy.

## C.4 Multiplicity dependent CASTOR energy, normalized to the inclusive energy

The measurements of the CASTOR energy spectra and the multiplicity dependent CASTOR energy share a common definition of the stable particle level based on the minimal proton momentum loss $\xi$. The multiplicity dependent distribution can therefore be scaled by the average of the spectra. This way, many uncertainties, especially the CASTOR energy scale uncertainty, cancel and a shape comparison is possible with reduced uncertainties. The obtained distributions are shown in Figs C.11–C.13. It can be observed that the rising part of the distributions is well described by all of the models. EPOS LHC and QGSJETII.04 start to deviate from the measurement by flattening too soon to a plateau region.

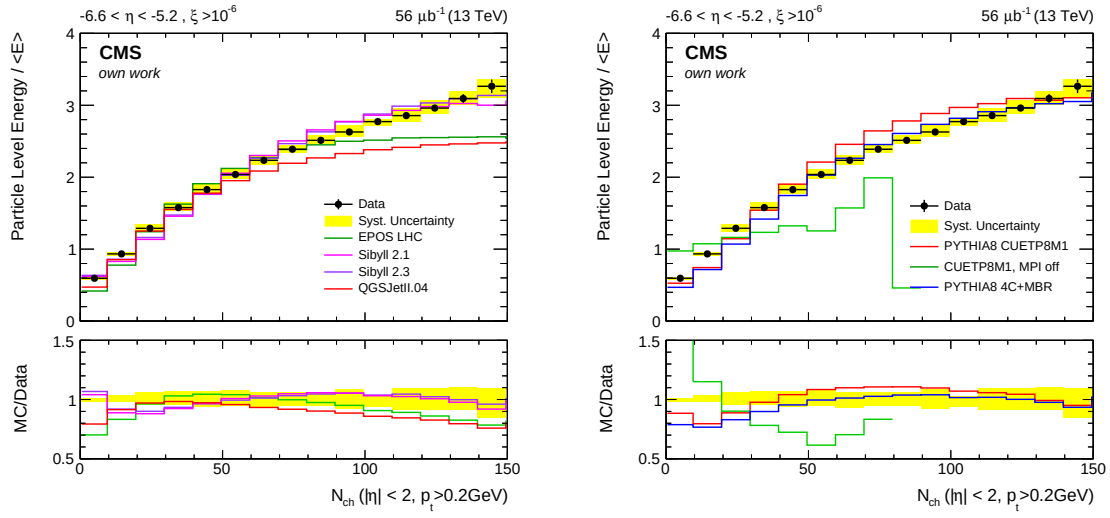



**Fig. C.11.:** Average total energy in the acceptance of CASTOR at $-6.6 < \eta < -5.2$ for events with $\xi > 10^{-6}$ as a function of the charged particle multiplicity $N_{\mathrm{ch}}$ ($p_{\mathrm{T}} > 0.2\,\mathrm{GeV}, |\eta| < 2$) and scaled by the inclusive average energy. The data is compared to predictions by hadronic event generators commonly used in high energy cosmic ray physics (left panel) and variations of PYTHIA 8 (right panel).

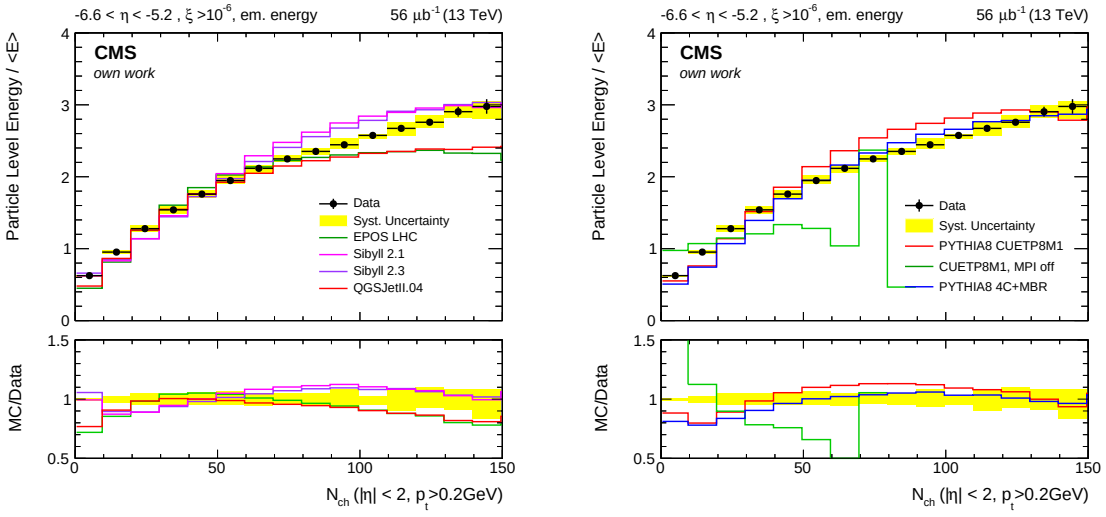



**Fig. C.12.:** Average electromagnetic energy in the acceptance of CASTOR at $-6.6 < \eta < -5.2$ for events with $\xi > 10^{-6}$ as a function of the charged particle multiplicity $N_{\mathrm{ch}}\,(p_{\mathrm{T}} > 0.2\,\mathrm{GeV}, |\eta| < 2)$ and scaled by the inclusive average energy. The data is compared to predictions by hadronic event generators commonly used in high energy cosmic ray physics (left panel) and variations of PYTHIA 8 (right panel).

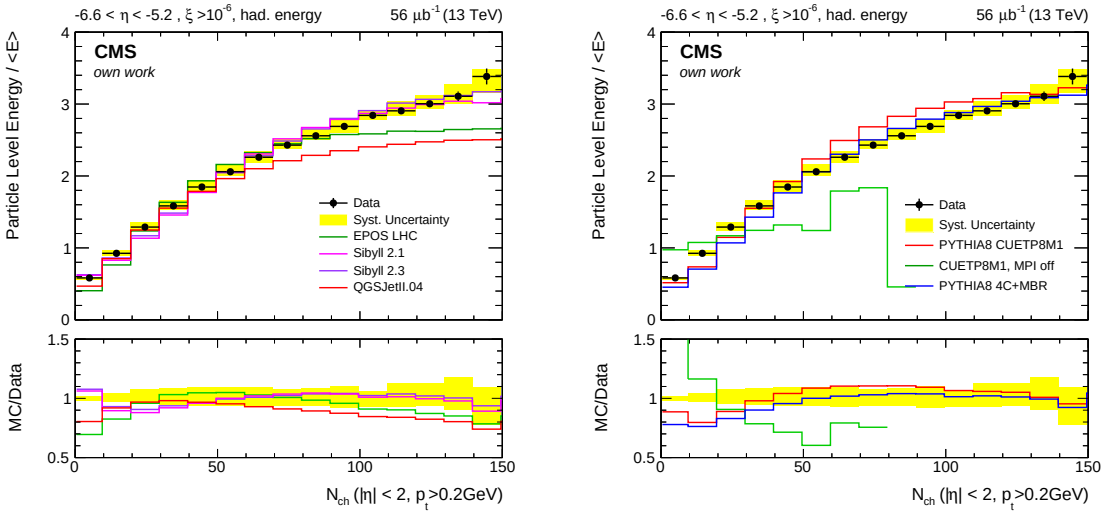



**Fig. C.13.:** Average hadronic energy in the acceptance of CASTOR at $-6.6 < \eta < -5.2$ for events with $\xi > 10^{-6}$ as a function of the charged particle multiplicity $N_{\mathrm{ch}}\,(p_{\mathrm{T}} > 0.2\,\mathrm{GeV}, |\eta| < 2)$ and scaled by the inclusive average energy. The data is compared to predictions by hadronic event generators commonly used in high energy cosmic ray physics (left panel) and variations of PYTHIA 8 (right panel).

# D Additional material on the CMS generator tunes

## D.1 Detailed description of the PYTHIA 8 parameters used in the CMS tunes

- `MultipartonInteractions:PT0Ref`
  The regularization parameter of the partonic cross section for $p_{\mathrm{T}} \to 0$ at the reference energy of 7 TeV, see Eqs. (2.13) and (2.14). The allowed range is $[0.5 - 10]$.
- `MultipartonInteractions:ecmPow`
  Energy scaling of the MPI dampening parameter, see Eq. (2.14). The allowed range is $[0 - 0.5]$.
- `MultipartonInteractions:coreFraction`
  The fraction of parton matter contained within the inner radius of the proton. This corresponds to the parameter $\beta$ in Eq. (2.16). The allowed range is $[0 - 1]$.
- `MultipartonInteractions:coreRadius` The radius fraction in which the above defined matter is located. This corresponds to the parameter $\hat{r}_{\beta}$ in Eq. (2.16). The allowed range is $[0.1 - 0]$.
- `ColorReconnection:range` Probability parameter for the merging of systems in the MPI-based color reconnection model. Corresponds to $R$ in Eq. (2.17). The allowed range is $[0 - 10]$.

## D.2 Full parameter sets of the CMS tunes CP1–CP4

Table D.1 summarizes the setting of the PYTHIA 8 event generator as obtained for the tunes CP1–CP4.

**Tab. D.1.:** Setting for the CMS underlying event tunes CP1–CP4. Tunes use PYTHIA 8 diffraction.

| Parameter | CP1 | CP2 | CP3 | CP4 |
|---|---|---|---|---|
| PDF Set | NNPDF3.1 | NNPDF3.1 | NNPDF3.1 | NNPDF3.1 |
| PDF Order | LO | LO | NLO | NNLO |
| $\alpha_s(M_Z)$ | 0.130 | 0.130 | 0.118 | 0.118 |
| Central diffraction | on | on | on | on |
| SigmaTotal:zeroAXB | off | off | off | off |
| SpaceShower:rapidityOrder | off | off | off | off |
| bProfile | 2 | 2 | 2 | 2 |
| $\alpha_s^{ISR}$ value/order | 0.1365/LO | 0.130/LO | 0.118/NLO | 0.118/NLO |
| $\alpha_s^{FSR}$ value/order | 0.1365/LO | 0.130/LO | 0.118/NLO | 0.118/NLO |
| $\alpha_s^{MPI}$ value/order | 0.130 /LO | 0.130/LO | 0.118/NLO | 0.118/NLO |
| $\alpha_s^{ME}$ value/order | 0.130 /LO | 0.130/LO | 0.118/NLO | 0.118/NLO |
| EcmRef (GeV) | 7000* | 7000* | 7000* | 7000* |
| PT0Ref (GeV) | 2.4 | 2.306 | 1.516 | 1.483 |
| ecmPow | 0.154 | 0.1391 | 0.0227 | 0.0201 |
| coreFraction | 0.684 | 0.3269 | 0.3869 | 0.3053 |
| coreRadius | 0.544 | 0.3755 | 0.5396 | 0.5971 |
| ColorReconnection:range | 2.633 | 2.323 | 4.727 | 5.613 |

## D.3 Parameters of the eigentunes of CP1 and CP1F

So-called Eigentunes are provided by the PROFESSOR framework and estimate the tune uncertainties by varying the tuned parameters in a way that the fit $\chi^2/ndof$ increases by one. The parameters are varied along maximally independent principle directions in the parameter space. The obtained parameter sets of the eigentunes are given in Tab. D.2 and Tab. D.3 for the tunes CP1 and CP1F respectively.

**Tab. D.2.:** Parameters of the eigentunes for tune CP1 as obtained with the PROFESSOR software.

| | | MultipartonInteractions | | | ColourReconnection |
|---|---|---|---|---|---|
| Tune | pT0Ref | ecmPow | coreFraction | coreRadius | range |
| Central | 2.39995 | 0.154346 | 0.683556 | 0.543576 | 2.63253 |
| 1+ | 2.39995 | 0.154323 | 0.683558 | 0.543609 | 2.63254 |
| 1- | 2.39995 | 0.154339 | 0.683557 | 0.543585 | 2.63253 |
| 2+ | 2.39995 | 0.154346 | 0.683556 | 0.543573 | 2.63253 |
| 2- | 2.39995 | 0.154347 | 0.683555 | 0.543594 | 2.63249 |
| 3+ | 2.39995 | 0.154336 | 0.683496 | 0.543573 | 2.63253 |
| 3- | 2.39995 | 0.154343 | 0.683535 | 0.543575 | 2.63253 |
| 4+ | 2.39995 | 0.154343 | 0.683557 | 0.543574 | 2.63252 |
| 4- | 2.39995 | 0.154336 | 0.683558 | 0.543571 | 2.63252 |
| 5+ | 2.39995 | 0.154346 | 0.683556 | 0.543576 | 2.63253 |
| 5- | 2.39997 | 0.154344 | 0.683556 | 0.543576 | 2.63253 |

**Tab. D.3.:** Parameters of the eigentunes for tune CP1F as obtained with the PROFESSOR software.

| | | MultipartonInteractions | | | ColourReconnection |
|---|---|---|---|---|---|
| Tune | pT0Ref | ecmPow | coreFraction | coreRadius | range |
| Central | 2.31905 | 0.157023 | 0.678655 | 0.668801 | 3.84906 |
| 1+ | 2.31905 | 0.157024 | 0.678646 | 0.668818 | 3.84906 |
| 1- | 2.31905 | 0.157024 | 0.678649 | 0.668811 | 3.84906 |
| 2+ | 2.31905 | 0.157023 | 0.678647 | 0.668844 | 3.84891 |
| 2- | 2.31905 | 0.157023 | 0.678656 | 0.668796 | 3.84908 |
| 3+ | 2.31905 | 0.157029 | 0.678656 | 0.668801 | 3.84906 |
| 3- | 2.31905 | 0.157038 | 0.678657 | 0.668801 | 3.84906 |
| 4+ | 2.31905 | 0.157010 | 0.678753 | 0.668853 | 3.84907 |
| 4- | 2.31905 | 0.157023 | 0.678653 | 0.668800 | 3.84906 |
| 5+ | 2.31906 | 0.157023 | 0.678655 | 0.668801 | 3.84906 |
| 5- | 2.31916 | 0.157023 | 0.678655 | 0.668802 | 3.84906 |

## D.4 Tune validations

Figures D.1 – D.3 illustrate the performance of the newly derived tunes CP1 and CP1F of PYTHIA 8 compared to the older tune CUETP8M1. The underlying event measurement at 13 TeV is consistently better described by the new tunes. The same is valid for the CASTOR measurements that correlate the central to forward activity. No such improvement is observed in the inclusive measurements. The underlying event measurement at 7 TeV is equally well described by all tunes. This is expected since these measurements were also included in the fit of tune CUETP8M1. IN general no big difference is seen between tunes CP1 and CP1F.

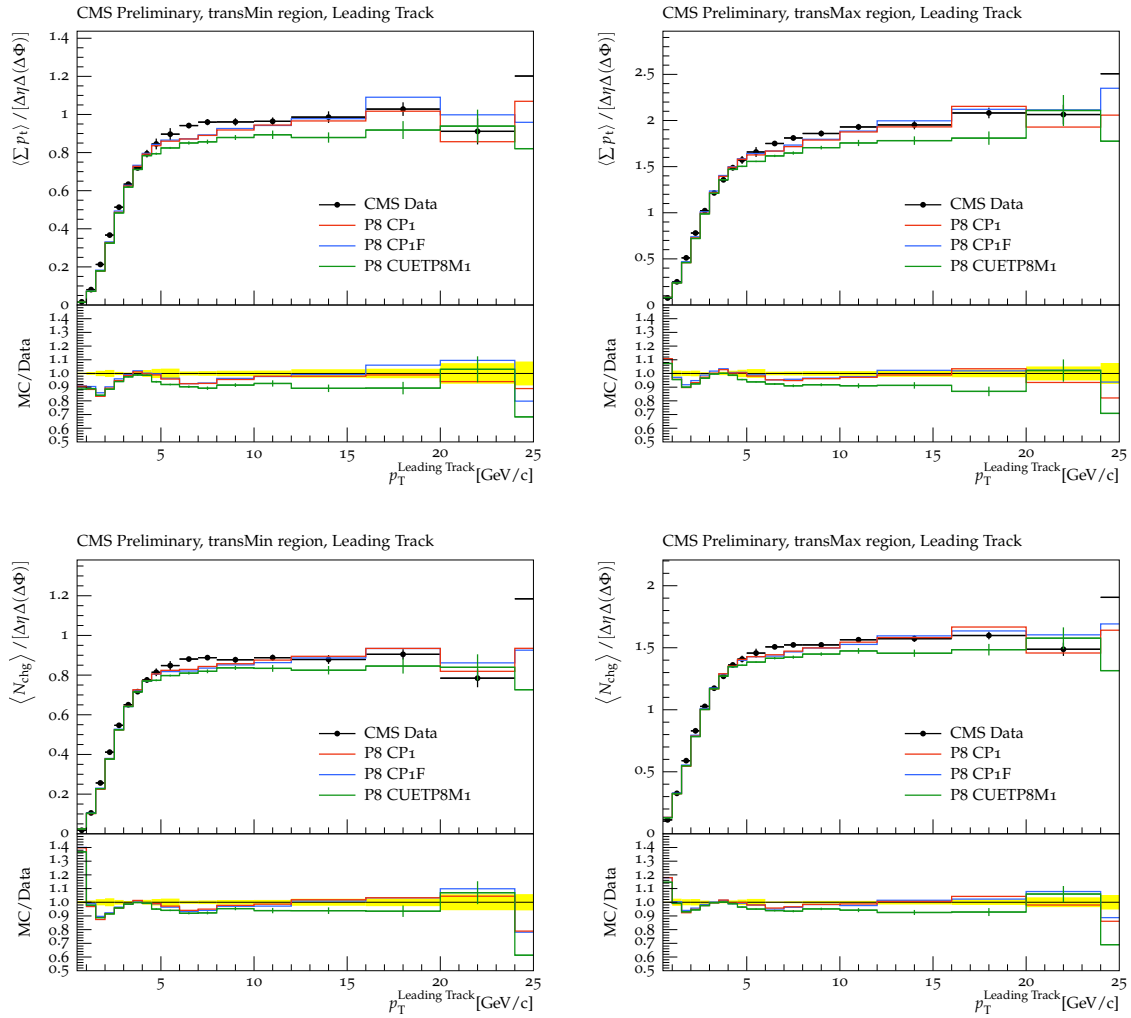



**Fig. D.1.:** Performance of the newly derived tunes compared to tune CUETP8M1: Underlying event measurements in the TransMin (left) and TransMax (right) regions at $\sqrt{s} = 13\,\text{TeV}$. These distributions were used to derive the tunes CP1 and CP1F.

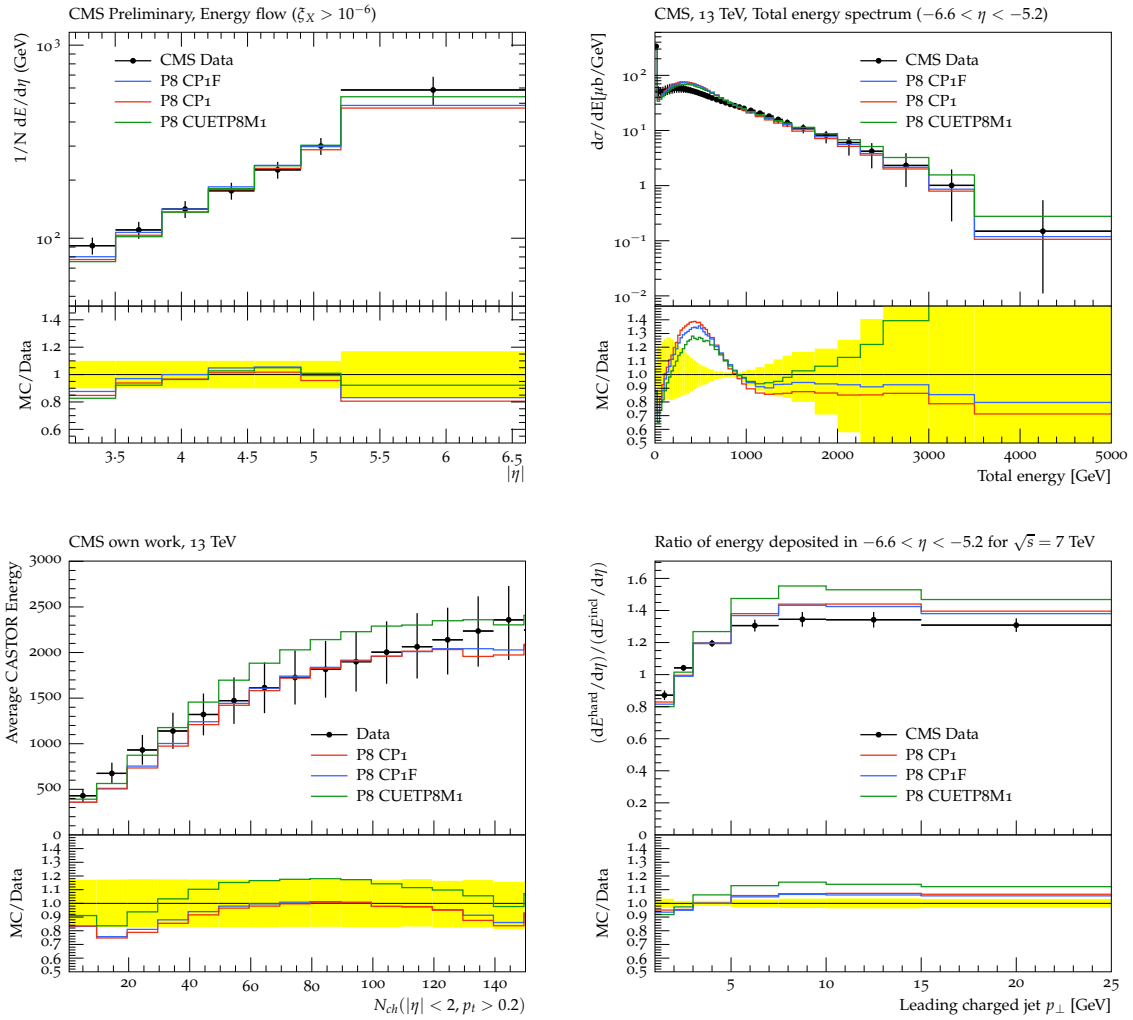



**Fig. D.2.:** Performance of the newly derived tunes compared to tune CUETP8M1: Forward energy measurements at $\sqrt{s} = 7$ and 13 TeV. Inclusive energy density $\mathrm{d}E/\mathrm{d}\eta$ (top left), CASTOR energy spectrum $\mathrm{d}\sigma/\mathrm{d}E$ (top right), multiplicity dependent CASTOR energy $\langle E\,(N_{\mathrm{ch}})\rangle$ (bottom left), and CASTOR energy ratio as function of leading charged jet $p_{\mathrm{T}}$. These distributions were used to derive the tune CP1F.

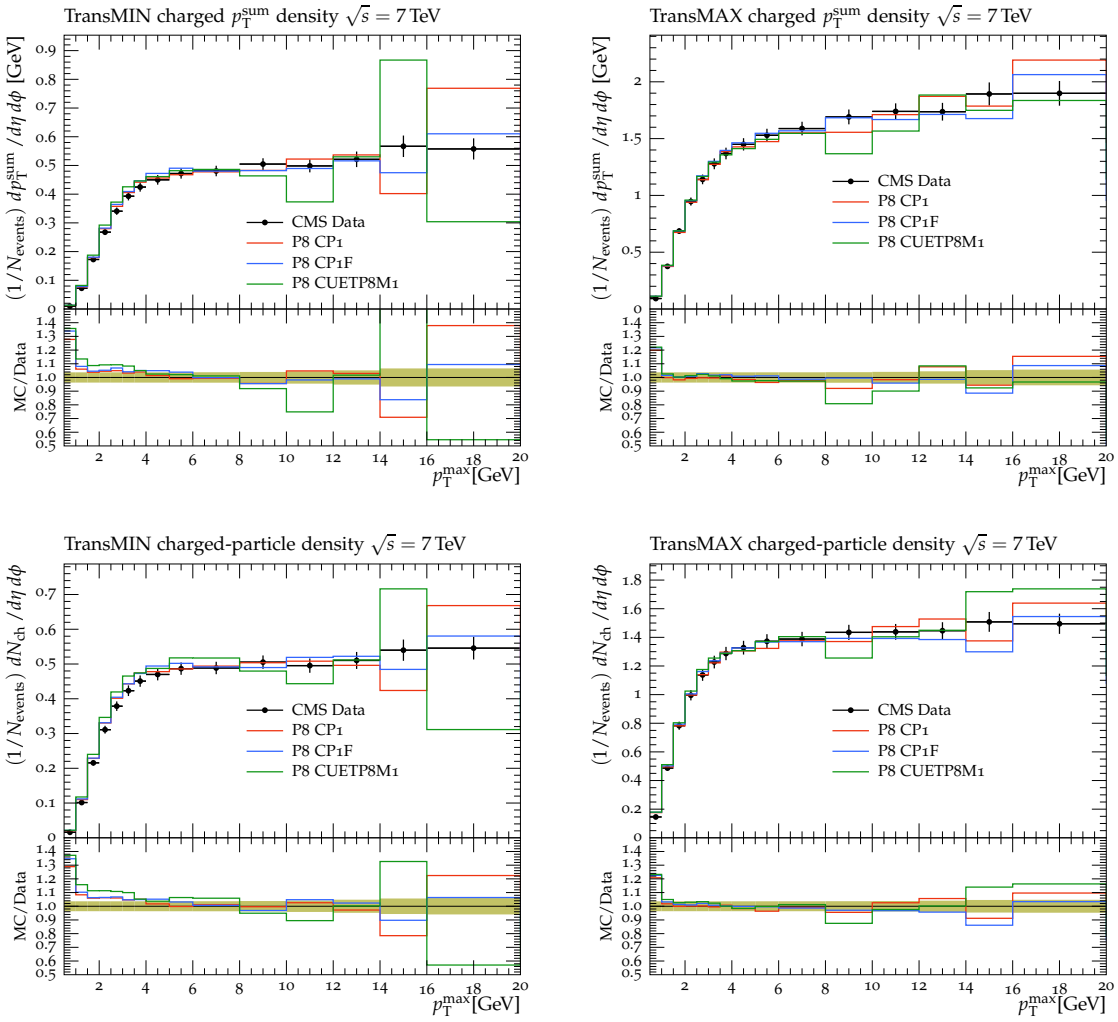



**Fig. D.3.:** Performance of the newly derived tunes compared to tune CUETP8M1: Underlying event measurements in the TransMin (left) and TransMax (right) regions at $\sqrt{s} = 7$ TeV. These distributions were used to derive the tunes CP1 and CP1F.

## D.5 Adaption for HERWIG 7.1

The multi-purpose event generator HERWIG [154] is a widely used alternative for PYTHIA. In this work, it is considered in the latest version HERWIG 7.1 [155] which includes a new model for soft and diffractive scattering [156]. Both are similar in the way the full collision final state is composed of different components (hard scattering, MPI, parton showers, etc) that are calculated independently and combined afterwards. Still, the implementation of these processes is significantly different. The most striking difference is the hadronization, which is realized through *clusters* in HERWIG instead of strings. The cluster model is based on the QCD confinement, assuming that partons in the final state are grouped to form color-neutral clusters. These clusters decay into pairs of color-singlet hadrons until the mass of the cluster is below the hadron mass. For a more detailed description of the HERWIG model see [154–156].

With the latest update the default settings of HERWIG 7.1 provide a very good description of minimum bias (MB) and underlying event (UE) measurements. The parameter sets were derived by tuning to data of the ATLAS experiment at center-of-mass energies of 900 GeV and 7 TeV. In order to demonstrate the power of the presented measurements and tuning efforts, the procedure developed for PYTHIA 8 that resulted in the tune CP1F is adapted to obtain a new tune of HERWIG 7.1. Here, the energy dependent parameters are not tunes and only reference data at $\sqrt{s} = 13$ TeV is used. The resulting tune is called *CH1F*. Table D.4 summarizes the tuned parameters with their default values, tuning ranges, and new tuned values.

The predictions of the new tune are compared to the default tune of HERWIG 7.1 and PYTHIA 8 tune CP1F in Figs. D.4–D.6. It can be observed that the agreement of HERWIG 7.1 with data at 13 TeV is improved for the new tune CH1F with respect to the default tune. This is especially true for the measurement in the forward region shown in Fig. D.5. Still, the level of PYTHIA 8 is not reached.

**Tab. D.4.:** Parameters and values of default HERWIG 7.1 as well as for the newly derived tune CH1F

| HERWIG 7.1 Parameter | default | Tuning range | CH1F |
|---|---|---|---|
| MPIHandler:pTmin0 [GeV] | 3.503 | $2.0 - 5.0$ | 2.854 |
| MPIHandler:InvRadius | 1.402 | $1.0 - 2.0$ | 1.083 |
| RemnantDecayer:ladderNorm | 0.95 | $0.8 - 1.2$ | 1.035 |
| MEMinBias:csNorm | 4.558 | $3.0 - 6.0$ | 2.47 |
| ColourReconnector:ReconnectionProbability | 0.5 | $0.1 - 0.9$ | 0.407 |

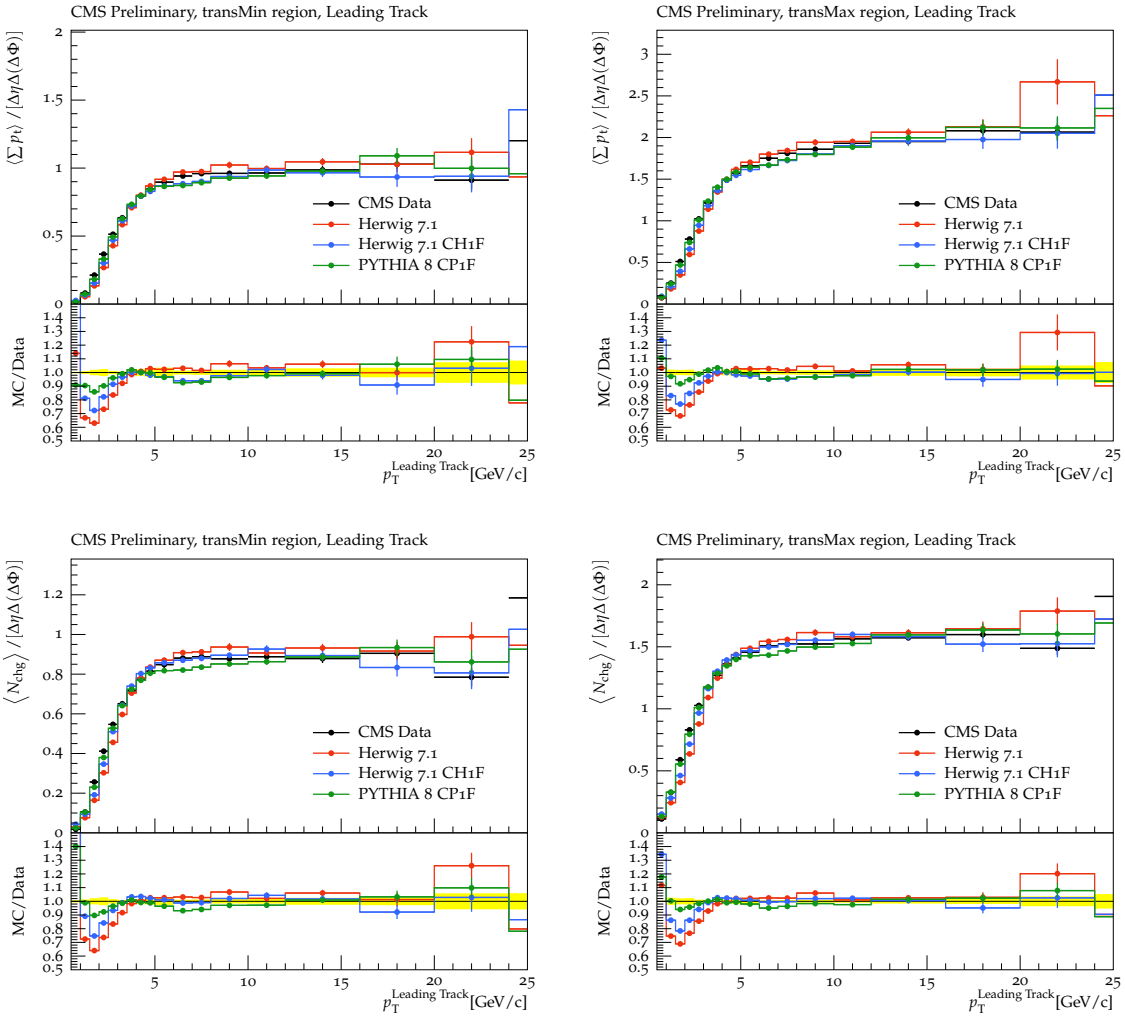



**Fig. D.4.:** Performance of the newly derived tune CH1F of HERWIG 7.1 compared to the default tune as well as PYTHIA 8 tune CP1F: Underlying event measurements in the TransMin (left) and TransMax (right) regions at $\sqrt{s} = 13\,\text{TeV}$ [47]. These distributions were used to derive the tunes CH1F and CP1F.

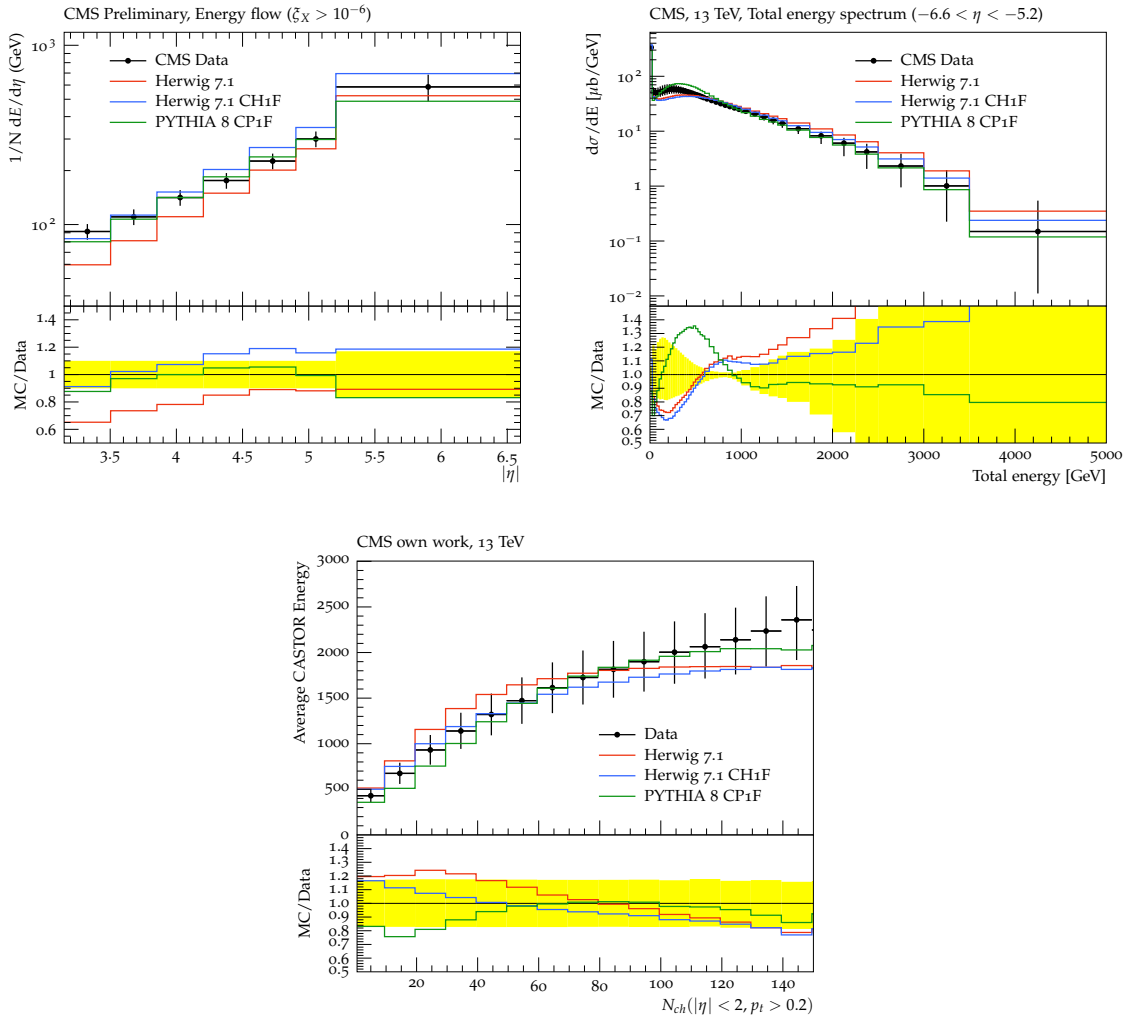



**Fig. D.5.:** Performance of the newly derived tune CH1F of HERWIG 7.1 compared to the default tune as well as PYTHIA 8 tune CP1F: Forward energy measurements at 13 TeV. Inclusive energy density $\mathrm{d}E/\mathrm{d}\eta$ [119] (top left), CASTOR energy spectrum $\mathrm{d}\sigma/\mathrm{d}E$ [118] (top right), and multiplicity dependent CASTOR energy $\langle E\,(N_{\mathrm{ch}})\rangle$ (bottom). These distributions were used to derive the tune CP1F.

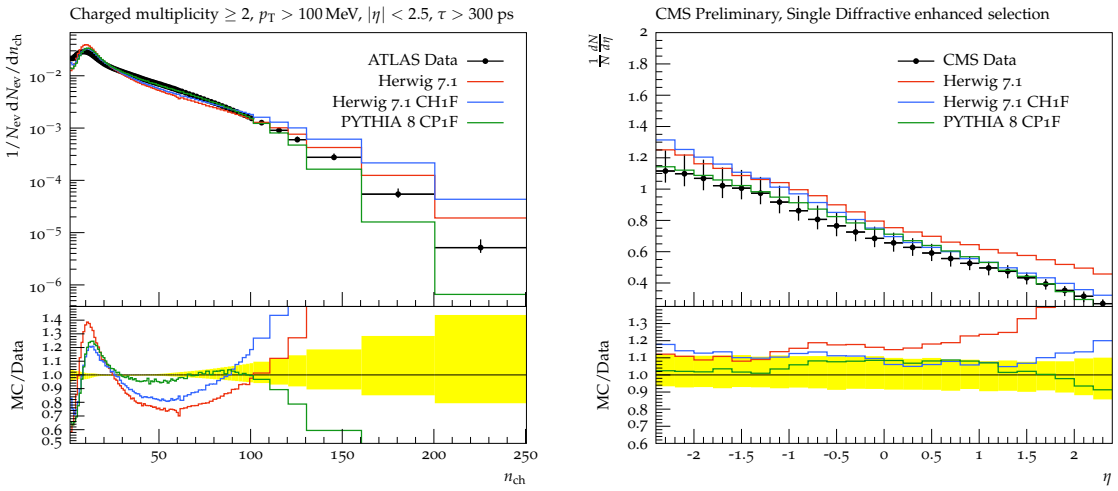



**Fig. D.6.:** Performance of the newly derived tune CH1F of HERWIG 7.1 compared to the default tune as well as PYTHIA 8 tune CP1F: charged particle multiplicity distribution as measured by ATLAS [139] (left) and average particle multiplicity $\mathrm{d}N/\mathrm{d}\eta$ with a single diffractive enhanced event selection [140] (right).

# Additional material regarding Sibyll 2.3c

E

Shortly before the submission of this thesis, an updated version of SIBYLL 2.3 was provided by the authors. The new version is called SIBYLL 2.3c and is available through the CRMC package but currently not publicly available. For consistency reasons, all results and conclusions were provided for SIBYLL 2.3 only. It was nevertheless investigated whether the changes introduced in the updated versions would influence the results or conclusions. The predictions of SIBYLL 2.3c are compared to SIBYLL 2.3 in Figs. E.1–E.3 for the measurements presented in this work. The data points are not shown for simplicity. It can be seen that the differences between the original and updated version of SIBYLL 2.3 are consistent on the level of few percent. Only the high energy tail in the energy spectra is noticeably reduced. The same is true for the elasticity distribution shown in Fig. E.4. Therefore the comparison to the data as well as all implications found for SIBYLL 2.3 are as well valid for the updated version SIBYLL 2.3c.

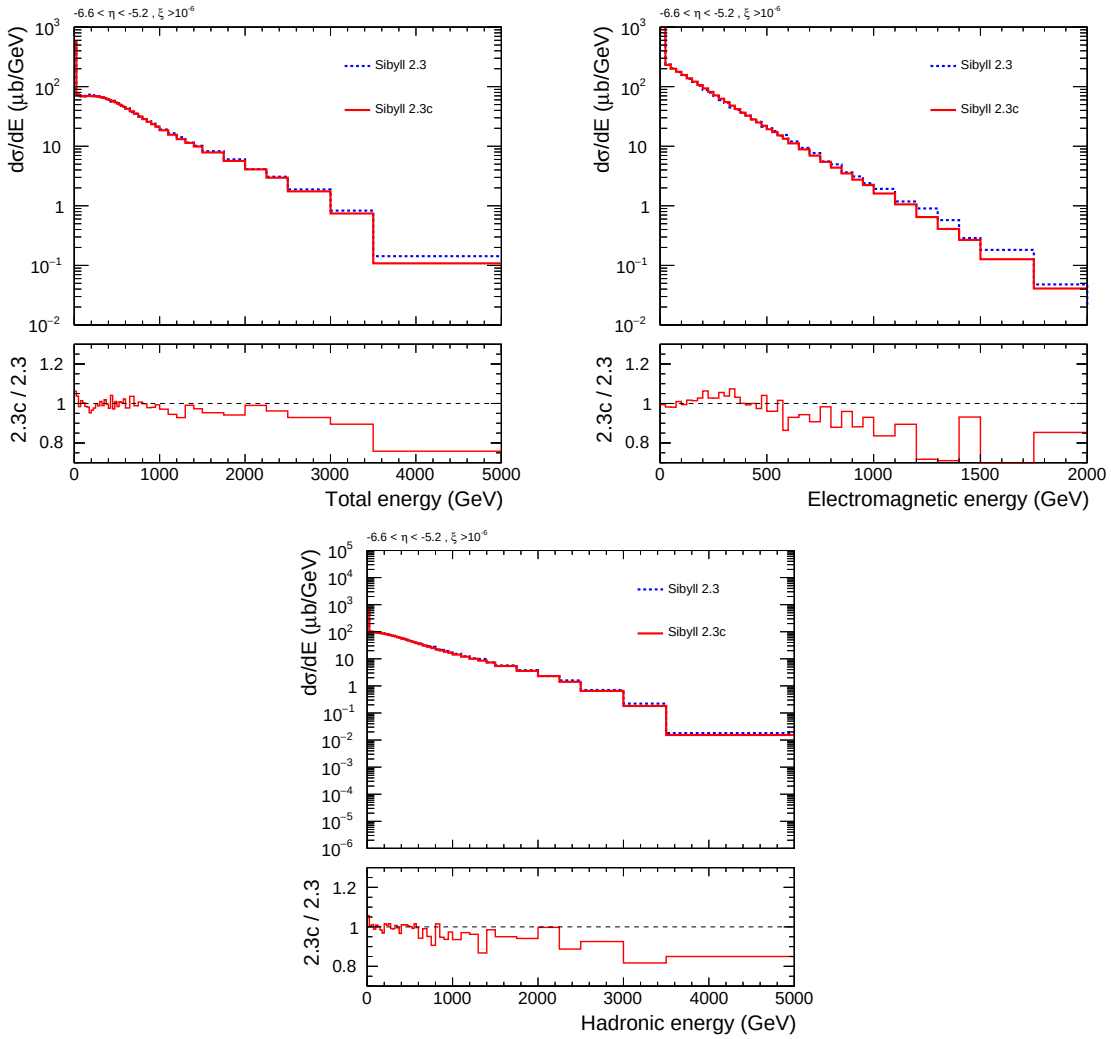



**Fig. E.1.:** Predictions of SIBYLL 2.3c compared to SIBYLL 2.3 for the inclusive energy spectra measured with CASTOR. The spectrum of the total energy is shown in the top left, the electromagnetic in the top right and the hadronic spectrum in the bottom panel. The data are not shown for simplicity.

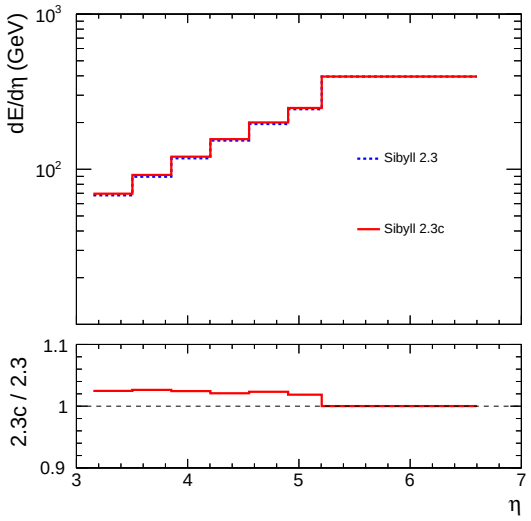



**Fig. E.2.:** Predictions of SIBYLL 2.3c compared to SIBYLL 2.3 for the forward energy density as function of the pseudorapidity. The data point are not shown for simplicity.

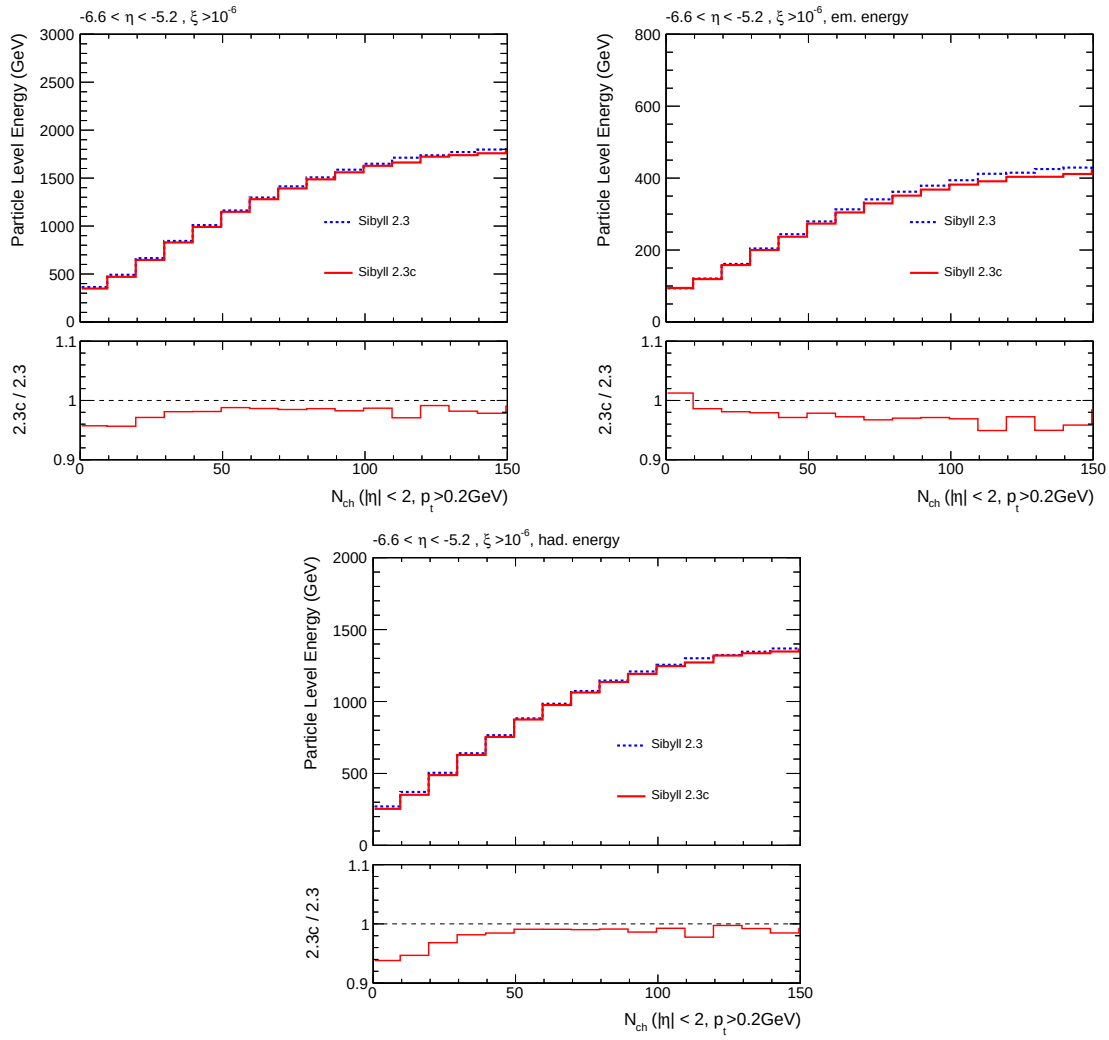



**Fig. E.3.:** Predictions of SIBYLL 2.3c compared to SIBYLL 2.3 for the multiplicity dependent energy measured with CASTOR. The distribution of the total energy is shown in the top left, the electromagnetic in the top right and the hadronic energy in the bottom panel. The data are not shown for simplicity.

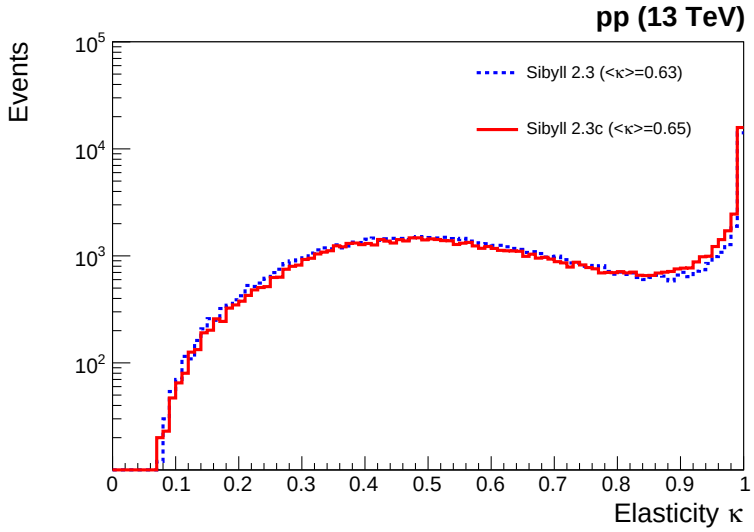



**Fig. E.4.:** Elasticity distribution for proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ of SIBYLL 2.3c compared to SIBYLL 2.3.

# References


[1] L. Evans and P. Bryant, "LHC Machine", *JINST*, vol. 3, S08001, 2008. DOI: 10.1088/1748-0221/3/08/S08001.

[2] CDF Collaboration, "Observation of top quark production in $\bar{p}p$ collisions", *Phys. Rev. Lett.*, vol. 74, p. 2626, 1995. DOI: 10.1103/PhysRevLett.74.2626.

[3] D0 Collaboration, "Observation of the top quark", *Phys. Rev. Lett.*, vol. 74, p. 2632, 1995. DOI: 10.1103/PhysRevLett.74.2632.

[4] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B*, vol. 716, p. 1, 2012. DOI: 10.1016/j.physletb.2012.08.020.

[5] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B*, vol. 716, p. 30, 2012. DOI: 10.1016/j.physletb.2012.08.021.

[6] E. Halkiadakis, G. Redlinger, and D. Shih, "Status and Implications of Beyond-the-Standard-Model Searches at the LHC", *Ann. Rev. Nucl. Part. Sci.*, vol. 64, p. 319, 2014. DOI: 10.1146/annurev-nucl-102313-025632. arXiv: 1411.1427 [hep-ex].

[7] R. Pasechnik and M. Šumbera, "Phenomenological Review on Quark–Gluon Plasma: Concepts vs. Observations", *Universe*, vol. 3, p. 7, 2017. DOI: 10.3390/universe3010007.

[8] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST*, vol. 3, S08002, 2008. DOI: 10.1088/1748-0221/3/08/S08002.

[9] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST*, vol. 3, S08003, 2008. DOI: 10.1088/1748-0221/3/08/S08003.

[10] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST*, vol. 3, S08004, 2008. DOI: 10.1088/1748-0221/3/08/S08004.

[11] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST*, vol. 3, S08005, 2008. DOI: 10.1088/1748-0221/3/08/S08005.

[12] LHCf Collaboration, "The LHCf detector at the CERN Large Hadron Collider", *JINST*, vol. 3, S08006, 2008. DOI: 10.1088/1748-0221/3/08/S08006.

[13] TOTEM Collaboration, "The TOTEM experiment at the CERN Large Hadron Collider", *JINST*, vol. 3, S08007, 2008. DOI: 10.1088/1748-0221/3/08/S08007.

[14] P. Nason and B. Webber, "Next-to-leading-order event generators", *Ann. Rev. Nucl. Part. Sci.*, vol. 62, p. 187, 2012. DOI: 10.1146/annurev-nucl-102711-094928.

[15] C. D. Anderson, "The Positive Electron", *Phys. Rev.*, vol. 43, pp. 491–494, 1933. DOI: 10.1103/PhysRev.43.491.

[16] S. H. Neddermeyer and C. D. Anderson, "Note on the Nature of Cosmic Ray Particles", *Phys. Rev.*, vol. 51, pp. 884–886, 1937. DOI: 10.1103/PhysRev.51.884.

[17] C. M. G. Lattes, H. Muirhead, G. P. S. Occhialini, and C. F. Powell, "PROCESSES INVOLVING CHARGED MESONS", *Nature*, vol. 159, p. 694, 1947, [,42(1947)]. DOI: 10.1038/159694a0.

[18] J. Blümer, R. Engel, and J. R. Horandel, "Cosmic Rays from the Knee to the Highest Energies", *Prog. Part. Nucl. Phys.*, vol. 63, p. 293, 2009. DOI: 10.1016/j.ppnp.2009.05.002.

[19] Pierre Auger Collaboration, "The Fluorescence Detector of the Pierre Auger Observatory", *Nucl. Instrum. Meth. A*, vol. 620, p. 227, 2010. DOI: 10.1016/j.nima.2010.04.023.

[20] Pierre Auger Collaboration, "The Pierre Auger Cosmic Ray Observatory", *Nucl. Instrum. Meth. A*, vol. 798, p. 172, 2015. DOI: 10.1016/j.nima.2015.06.058.

[21] K. Akiba *et al.*, "LHC forward physics", *J. Phys. G*, vol. 43, p. 110 201, 2016. DOI: 10.1088/0954-3899/43/11/110201.

[22] Pierre Auger Collaboration, "Muons in air showers at the Pierre Auger Observatory: Measurement of atmospheric production depth", *Phys. Rev. D*, vol. 90, p. 012 012, 2014, [Erratum: Phys. Rev.D92,019903(2015)]. DOI: 10.1103/PhysRevD.90.012012.

[23] Pierre Auger Collaboration, "Muons in air showers at the Pierre Auger Observatory: Mean number in highly inclined events", *Phys. Rev. D*, vol. 91, p. 032 003, 2015, [Erratum: PhysRevD91,059901]. DOI: 10.1103/PhysRevD.91.032003.

[24] Pierre Auger Collaboration, "Depth of maximum of air-shower profiles at the Pierre Auger Observatory. I. Measurements at energies above $10^{17.8}$ eV", *Phys. Rev. D*, vol. 90, p. 122 005, 2014. DOI: 10.1103/PhysRevD.90.122005.

[25] Pierre Auger Collaboration, "Depth of maximum of air-shower profiles at the Pierre Auger Observatory. II. Composition implications", *Phys. Rev. D*, vol. 90, p. 122 006, 2014. DOI: 10.1103/PhysRevD.90.122006.

[26] R. U. Abbasi *et al.*, "Study of Ultra-High Energy Cosmic Ray composition using Telescope Array's Middle Drum detector and surface array in hybrid mode", *Astropart. Phys.*, vol. 64, pp. 49–62, 2015. DOI: 10.1016/j.astropartphys.2014.11.004.

[27] T. Sjöstrand, S. Ask, J. R. Christiansen, *et al.*, "An introduction to PYTHIA 8.2", *Comput. Phys. Commun.*, vol. 191, p. 159, 2015. DOI: 10.1016/j.cpc.2015.01.024.

[28] R. Ulrich, R. Engel, and M. Unger, "Hadronic Multiparticle Production at Ultra-High Energies and Extensive Air Showers", *Phys. Rev. D*, vol. 83, p. 054 026, 2011. DOI: 10.1103/PhysRevD.83.054026.

[29] Wikimedia and MissMJ, *The standard model of elementary particles*, 2017. Online: https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg.

[30] CMS Collaboration, "Determination of the strong coupling constant from the measurement of inclusive multijet event cross sections in pp collisions at $\sqrt{s} = 8$ TeV", CMS Physics Analysis Summary, 2017. Online: http://cds.cern.ch/record/2253091.

[31] E. D. Bloom, D. H. Coward, H. DeStaebler, *et al.*, "High-Energy Inelastic $e - p$ Scattering at $6^\circ$ and $10^\circ$", *Phys. Rev. Lett.*, vol. 23, p. 930, Oct. 1969. DOI: 10.1103/PhysRevLett.23.930.

[32] M. Breidenbach, J. I. Friedman, H. W. Kendall, *et al.*, "Observed behavior of highly inelastic electron-proton scattering", *Phys. Rev. Lett.*, vol. 23, p. 935, Oct. 1969. DOI: 10.1103/PhysRevLett.23.935.

[33] ZEUS, H1 Collaboration, "Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD anysis of HERA data", *Eur. Phys. J. C*, vol. 75, p. 580, 2015. DOI: 10.1140/epjc/s10052-015-3710-4.

[34] T. Muta, *Foundations of Quantum Chromodynamics: An Introduction to Perturbative Methods in Gauge Theories, (3rd ed.)* Ser. World scientific Lecture Notes in Physics. Hackensack, N.J.: World Scientific, 2010, vol. 78.

[35] R. Corke and T. Sjöstrand, "Multiparton Interactions and Rescattering", *JHEP*, vol. 01, p. 035, 2010. DOI: 10.1007/JHEP01(2010)035.

[36] T. Sjöstrand, "The Development of MPI Modelling in PYTHIA", 2017. arXiv: 1706.02166 [hep-ph].

[37] M. Diehl, D. Ostermeier, and A. Schafer, "Elements of a theory for multiparton interactions in QCD", *JHEP*, vol. 03, p. 089, 2012, [Erratum: JHEP03,001(2016)]. DOI: 10.1007/JHEP03(2012)089.

[38] B. Blok, Yu. Dokshitser, L. Frankfurt, and M. Strikman, "pQCD physics of multiparton interactions", *Eur. Phys. J. C*, vol. 72, p. 1963, 2012. DOI: 10.1140/epjc/s10052-012-1963-8.

[39]P. Bartalini and L. Fano, Eds., *Proceedings, 1st International Workshop on Multiple Partonic Interactions at the LHC (MPI08)*. 2010. arXiv: 1003.4220 [hep-ex].

[40] T. Sjöstrand and M. van Zijl, "A Multiple Interaction Model for the Event Structure in Hadron Collisions", *Phys. Rev. D*, vol. 36, p. 2019, 1987. DOI: 10.1103/PhysRevD.36.2019.

[41] CMS Collaboration, "Multiple Parton Interactions Studies at CMS", 2011. arXiv: 1103.6201 [hep-ex].

[42] Amber Harmon, *Sherpa and Open Science Grid*, 2014. Online: https://sciencenode.org/feature/sherpa-and-open-science-grid-predicting-emergence-jets.php.

[43] CDF Collaboration, R. Field, "The Underlying event in hard scattering processes", vol. C010630, p. 501, 2001. arXiv: HEP-PH/0201192 [hep-ph].

[44] CDF Collaboration, "Charged jet evolution and the underlying event in $p\bar{p}$ collisions at 1.8 TeV", *Phys. Rev. D*, vol. 65, p. 092 002, 2002. DOI: 10.1103/PhysRevD.65.092002.

[45] CDF Collaboration, "Studying the Underlying Event in Drell-Yan and High Transverse Momentum Jet Production at the Tevatron", *Phys. Rev. D*, vol. 82, p. 034 001, 2010. DOI: 10.1103/PhysRevD.82.034001.

[46] CMS Collaboration, "Measurement of the Underlying Event Activity at the LHC at 7 TeV and Comparison with 0.9 TeV", CMS Physics Analysis Summary, 2012. Online: http://cds.cern.ch/record/1478982.

[47] CMS Collaboration, "Underlying Event Measurements with Leading Particles and Jets in pp collisions at $\sqrt{s} = 13$ TeV", CMS Physics Analysis Summary, 2015. Online: http://cds.cern.ch/record/2104473.

[48] ATLAS Collaboration, "Measurement of charged-particle distributions sensitive to the underlying event in $\sqrt{s} = 13$ TeV proton-proton collisions with the ATLAS detector at the LHC", *JHEP*, vol. 03, p. 157, 2017. DOI: 10.1007/JHEP03(2017)157.

[49] Rick Field, Richard Haas, and David Stuart, "Defining Maximum and Minimum Transverse Regions to Study the Underlying Event in Hard Scattering Processes", *CDF Note CDF/ANAL/MIN BIAS/CDFR/5626,*, 2001. Online: http://www.phys.ufl.edu/~rfield/cdf/CDFnote_maxmin.pdf.

[50] V. N. Gribov, "A REGGEON DIAGRAM TECHNIQUE", *Sov. Phys. JETP*, vol. 26, p. 414, 1968, [Zh. Eksp. Teor. Fiz.53,654(1967)].

[51] N. Cartiglia, "Measurement of the proton-proton total cross section at 2, 7, 8 and 57 TeV", in *Proceedings, 32nd International Symposium on Physics in Collision (PIC 2012)*, 2013, p. 55. arXiv: 1303.2927 [hep-ex].

[52] S. Donnachie, H. G. Dosch, O. Nachtmann, and P. Landshoff, "Pomeron physics and QCD", *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.*, vol. 19, pp. 1–347, 2002.

[53] U. Amaldi, R. Biancastelli, C. Bosio, *et al.*, "The Energy dependence of the proton proton total cross-section for center-of-mass energies between 23 and 53 GeV", *Phys. Lett.*, vol. 44B, pp. 112–118, 1973. DOI: 10.1016/0370-2693(73)90315-8.

[54] S. R. Amendolia *et al.*, "Measurement of the total proton proton cross-section at the ISR", *Phys. Lett.*, vol. 44B, pp. 119–124, 1973. DOI: 10.1016/0370-2693(73)90316-X.

[55] P. D. B. Collins, *An Introduction to Regge Theory and High-Energy Physics*, ser. Cambridge Monographs on Mathematical Physics. Cambridge, UK: Cambridge Univ. Press, 2009.

[56] H. J. Drescher, M. Hladik, S. Ostapchenko, T. Pierog, and K. Werner, "Parton based Gribov-Regge theory", *Phys. Rept.*, vol. 350, p. 93, 2001. DOI: 10.1016/S0370-1573(00)00122-8.

[57] CMS Collaboration, "Study of the underlying event at forward rapidity in pp collisions at $\sqrt{s}$ = 0.9, 2.76, and 7 TeV", *JHEP*, vol. 04, p. 072, 2013. DOI: 10.1007/JHEP04(2013)072.

[58] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J. C*, vol. 76, p. 155, 2016. DOI: 10.1140/epjc/s10052-016-3988-x.

[59] Pierre Auger Collaboration, "Observation of a Large-scale Anisotropy in the Arrival Directions of Cosmic Rays above $8 \times 10^{18}$ eV", *Science*, vol. 357, no. 6537, p. 1266, 2017. DOI: 10.1126/science.aan4338.

[60] Telescope Array Collaboration, "The surface detector array of the Telescope Array experiment", *Nucl. Instrum. Meth. A*, vol. 689, p. 87, 2013. DOI: 10.1016/j.nima.2012.05.079.

[61] H. Tokuno *et al.*, "New air fluorescence detectors employed in the Telescope Array experiment", *Nucl. Instrum. Meth. A*, vol. 676, p. 54, 2012. DOI: 10.1016/j.nima.2012.02.044.

[62] K.-H. Kampert and M. Unger, "Measurements of the cosmic ray composition with air shower experiments", *Astropart. Phys.*, vol. 35, p. 660, 2012. DOI: http://dx.doi.org/10.1016/j.astropartphys.2012.02.004.

[63] R. Engel, D. Heck, and T. Pierog, "Extensive air showers and hadronic interactions at high energy", *Ann. Rev. Nucl. Part. Sci.*, vol. 61, p. 467, 2011. DOI: 10.1146/annurev.nucl.012809.104544.

[64] J. Matthews, "A Heitler model of extensive air showers", *Astropart. Phys.*, vol. 22, p. 387, 2005. DOI: 10.1016/j.astropartphys.2004.09.003.

[65] A. Haungs, H. Rebel, and M. Roth, "Energy spectrum and mass composition of high-energy cosmic rays", *Rept. Prog. Phys.*, vol. 66, p. 1145, 2003. DOI: 10.1088/0034-4885/66/7/202.

[66] Pierre Auger Collaboration, "Testing Hadronic Interactions at Ultrahigh Energies with Air Showers Measured by the Pierre Auger Observatory", *Phys. Rev. Lett.*, vol. 117, p. 192 001, 2016. DOI: 10.1103/PhysRevLett.117.192001.

[67] D. d'Enterria, R. Engel, T. Pierog, S. Ostapchenko, and K. Werner, "Constraints from the first LHC data on hadronic event generators for ultra-high energy cosmic-ray physics", *Astropart. Phys.*, vol. 35, p. 98, 2011. DOI: 10.1016/j.astropartphys.2011.05.002.

[68] T. Sjöstrand, "The Lund Monte Carlo for Jet Fragmentation", *Comput. Phys. Commun.*, vol. 27, p. 243, 1982. DOI: 10.1016/0010-4655(82)90175-8.

[69] H. U. Bengtsson, "The Lund Monte Carlo for High $p_T$ Physics", *Comput. Phys. Commun.*, vol. 31, p. 323, 1984. DOI: 10.1016/0010-4655(84)90018-3.

[70] T. Sjöstrand, "A Model for Initial State Parton Showers", *Phys. Lett. B*, vol. 157, p. 321, 1985. DOI: 10.1016/0370-2693(85)90674-4.

[71] H.-U. Bengtsson and T. Sjöstrand, "The Lund Monte Carlo for Hadronic Processes: Pythia Version 4.8", *Comput. Phys. Commun.*, vol. 46, p. 43, 1987. DOI: 10.1016/0010-4655(87)90036-1.

[72] T. Sjöstrand, P. Eden, C. Friberg, *et al.*, "High-energy physics event generation with PYTHIA 6.1", *Comput. Phys. Commun.*, vol. 135, p. 238, 2001. DOI: 10.1016/S0010-4655(00)00236-8.

[73] A. Buckley, J. Ferrando, S. Lloyd, *et al.*, "LHAPDF6: parton density access in the LHC precision era", *Eur. Phys. J. C*, vol. 75, p. 132, 2015. DOI: 10.1140/epjc/s10052-015-3318-8.

[74] J. Alwall, R. Frederix, S. Frixione, *et al.*, "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP*, vol. 07, p. 079, 2014. DOI: 10.1007/JHEP07(2014)079.

[75] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP*, vol. 11, p. 070, 2007. DOI: 10.1088/1126-6708/2007/11/070.

[76] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, "Parton Fragmentation and String Dynamics", *Phys. Rept.*, vol. 97, p. 31, 1983. DOI: 10.1016/0370-1573(83)90080-7.

[77] B. Andersson, G. Gustafson, and B. Söderberg, "A general model for jet fragmentation", *Zeitschrift für Physik C Particles and Fields*, vol. 20, p. 317, 1, 1983. DOI: 10.1007/BF01407824.

[78] B. Andersson, *The Lund Model*, ser. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 1998.

[79] T. Sjöstrand and P. Z. Skands, "Multiple interactions and the structure of beam remnants", *JHEP*, vol. 03, p. 053, 2004. DOI: 10.1088/1126-6708/2004/03/053.

[80] R. Ciesielski and K. Goulianos, "MBR Monte Carlo Simulation in PYTHIA8", *PoS (ICHEP2012)*, p. 301, 2013. arXiv: 1205.1446 [hep-ph].

[81] 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", *Phys. Rev. C*, vol. 74, p. 044 902, 2006. DOI: 10.1103/PhysRevC.74.044902.

[82] K. Werner, T. Hirano, I. Karpenko, *et al.*, "Gribov-Regge theory, partons, remnants, strings - and the EPOS model for hadronic interactions", *Nucl. Phys. B - Proc. Suppl.*, vol. 196, p. 36, 2009. DOI: 10.1016/j.nuclphysbps.2009.09.006.

[83] T. Pierog and K. Werner, "EPOS model and ultra high energy cosmic rays", *Nucl. Phys. B - Proc. Suppl.*, vol. 196, p. 102, 2009. DOI: 10.1016/j.nuclphysbps.2009.09.017.

[84] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko, and K. Werner, "EPOS LHC: test of collective hadronization with data measured at the CERN Large Hadron Collider", *Phys. Rev. C*, vol. 92, p. 034 906, 2015. DOI: 10.1103/PhysRevC.92.034906.

[85] K. Werner, "Core-corona separation in ultra-relativistic heavy ion collisions", *Phys. Rev. Lett.*, vol. 98, p. 152 301, 2007. DOI: 10.1103/PhysRevLett.98.152301.

[86] CMS Collaboration, "Observation of Long-Range Near-Side Angular Correlations in Proton-Proton Collisions at the LHC", *JHEP*, vol. 09, p. 091, 2010. DOI: 10.1007/JHEP09(2010)091.

[87] S. Ostapchenko, "Monte carlo treatment of hadronic interactions in enhanced pomeron scheme: QGSJET-II model", *Phys. Rev. D*, vol. 83, p. 014 018, 2011. DOI: 10.1103/PhysRevD.83.014018.

[88] E.-J. Ahn, R. Engel, T. K. Gaisser, P. Lipari, and T. Stanev, "Cosmic ray interaction event generator sibyll 2.1", *Phys. Rev. D*, vol. 80, p. 094 003, 2009. DOI: 10.1103/PhysRevD.80.094003.

[89] F. Riehn, R. Engel, A. Fedynitch, T. K. Gaisser, and T. Stanev, "A new version of the event generator Sibyll", *PoS (ICRC2015)*, p. 558, 2016. arXiv: 1510.00568 [hep-ph].

[90] T. Pierog, R. Ulrich, and C. Baus, *Cosmic Ray Monte Carlo (CRMC) package*, 2016. Online: https://web.ikp.kit.edu/rulrich/crmc.html.

[91] C. Baus, "Measurements in the Forward Phase-Space with the CMS Experiment and their Impact on Physics of Extensive Air Showers", PhD thesis, 2015. Online: http://cds.cern.ch/record/2109249.

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

[93] M. Dobbs and J. B. Hansen, "The HepMC C++ Monte Carlo event record for High Energy Physics", *Comput. Phys. Commun.*, vol. 134, pp. 41–46, 2001. DOI: 10.1016/S0010-4655(00)00189-2.

[94] G. Brianti, W. Hoogland, M. R. M. Jacob, *et al.*, *Workshop on the Feasibility of Hadron Colliders in the LEP Tunnel, Summary Report*. Geneva: CERN, 1984. Online: https://cds.cern.ch/record/683131.

[95] 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, 2012. Online: http://cds.cern.ch/record/1472732.

[96] C. Lefèvre, "The CERN accelerator complex. Complexe des accélérateurs du CERN", 2008, Online: http://cds.cern.ch/record/1260465.

[97] G. Bayatian *et al.*, "Design, performance and calibration of the CMS forward calorimeter wedges", *Eur. Phys. J. C*, vol. 53, p. 139, 2008. DOI: 10.1140/epjc/s10052-007-0459-4.

[98] Mans, J. and Hanson, J., "Calibration of the HF Absolute Energy Scale with Z Di-Electron Events Using 1 Inv. fb of Data in 2011", CMS Detector Note DN-11-012, 2011. Online: cms.cern.ch/iCMS/jsp/openfile.jsp?type=DN%5C&year=2011%5C&files=DN2011_012.pdf.

[99] CMS Collaboration, "The CMS trigger system", *JINST*, vol. 12, P01020, 2017. DOI: 10.1088/1748-0221/12/01/P01020.

[100] A. Heering, J. Rohlf, J. Freeman, *et al.*, "Performance of silicon photomultiplers with the CMS HCAL front-end electronics", *Nucl. Instrum. Meth. A*, vol. 576, p. 341, 2007. DOI: http://dx.doi.org/10.1016/j.nima.2006.11.049.

[101] CMS Collaboration, "Results on CASTOR Performance during LHC Run 2", Tech. Rep., 2016. Online: http://cds.cern.ch/record/2133152.

[102] V. Andreev *et al.*, "Performance studies of a full-length prototype for the CASTOR forward calorimeter at the CMS experiment", *Eur. Phys. J. C*, vol. 67, p. 601, 2010. DOI: 10.1140/epjc/s10052-010-1316-4.

[103] CMS Collaboration, "Measurement of energy flow at large pseudorapidities in pp collisions at $\sqrt{s} = 0.9$ and 7 TeV", *JHEP*, vol. 11, p. 148, 2011, [Erratum: JHEP02,055(2012)]. DOI: 10.1007/JHEP11(2011)148.

[104] CMS Collaboration, "Alignment of the CMS tracker with LHC and cosmic ray data", *JINST*, vol. 9, P06009, 2014. DOI: 10.1088/1748-0221/9/06/P06009.

[105]"Alignment of the CMS Tracking-Detector with First 2015 Cosmic-Ray and Collision Data", 2015. Online: http://cds.cern.ch/record/2041841.

[106] J. F. Rodrigues Antunes, "Design, implementation, and performance of a distributed and scalable sensor system for critical distance measurements in the CMS detector at LHC", PhD thesis, 2015. Online: https://cds.cern.ch/record/1999498.

[107] A. Sensors, *Linear position sensors - Active Sensors*. Online: http://www.activesensors.com/products/linear-potentiometer-position-sensors.

[108]"Alignment of CASTOR with 2013 pA TOTEM-Data", 2014. Online: https://cds.cern.ch/record/1697694.

[109] PerkinElmer, *Long Range Retro-reflective Sensor VTR24F1H*. Online: http://www.perkinelmer.com/CMSResources/Images/DTS_vtr24f1.pdf.

[110] GEANT4 Collaboration, "GEANT4—a simulation toolkit", *Nucl. Instrum. Meth. A*, vol. 506, p. 250, 2003. DOI: 10.1016/S0168-9002(03)01368-8.

[111] CMS-CASTOR Collaboration, P. Göttlicher, "Design and test beam studies for the CASTOR calorimeter of the CMS experiment", *Nucl. Instrum. Meth. A*, vol. 623, p. 225, 2010. DOI: 10.1016/j.nima.2010.02.203.

[112] CMS Collaboration, "Measurement of the energy distribution in the very forward direction at 13 TeV with CMS", CMS Physics Analysis Summary, 2016. Online: http://cds.cern.ch/record/2145374.

[113] CMS, TOTEM Collaboration, "Measurement of pseudorapidity distributions of charged particles in proton-proton collisions at $\sqrt{s} = 8$ TeV by the CMS and TOTEM experiments", *Eur. Phys. J. C*, vol. 74, p. 3053, 2014. DOI: 10.1140/epjc/s10052-014-3053-6.

[114] TOTEM Collaboration, "Measurement of the forward charged particle pseudorapidity density in $pp$ collisions at $\sqrt{s} = 7$ TeV with the TOTEM experiment", *Europhys. Lett.*, vol. 98, p. 31 002, 2012. DOI: 10.1209/0295-5075/98/31002.

[115] TOTEM Collaboration, "Measurement of the forward charged particle pseudorapidity density in pp collisions at $\sqrt{s} = 8$TeV using a displaced interaction point", *Eur. Phys. J. C*, vol. 75, p. 126, 2015. DOI: 10.1140/epjc/s10052-015-3343-7.

[116] LHCf Collaboration, "Measurement of zero degree single photon energy spectra for $\sqrt{s} = 7$ TeV proton-proton collisions at LHC", *Phys. Lett. B*, vol. 703, p. 128, 2011. DOI: 10.1016/j.physletb.2011.07.077.

[117] LHCf Collaboration, "Measurement of very forward neutron energy spectra for 7 TeV proton-proton Collisions at the Large Hadron Collider", *Phys. Lett. B*, vol. 750, p. 360, 2015. DOI: 10.1016/j.physletb.2015.09.041.

[118] CMS Collaboration, "Measurement of the inclusive energy spectrum in the very forward direction in proton-proton collisions at sqrt(s) = 13 TeV", *JHEP*, vol. 08, p. 046, 2017. DOI: 10.1007/JHEP08(2017)046.

[119] CMS Collaboration, "Measurement of the pseudorapidity dependence of the energy and transverse energy density in pp collisions at $\sqrt{s} = 13$ TeV with CMS", CMS Physics Analysis Summary. Online: http://cds.cern.ch/record/2146007.

[120] CMS Collaboration, "Pseudorapidity distribution of charged hadrons in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Lett. B*, vol. 751, p. 143, 2015. DOI: 10.1016/j.physletb.2015.10.004.

[121] CMS Collaboration, "CMS Luminosity Measurement for the 2015 Data Taking Period", CMS Physics Analysis Summary, 2016. Online: http://cds.cern.ch/record/2138682.

[122] T. Adye, "Unfolding algorithms and tests using RooUnfold", *Proceedings of the PHYSTAT 2011 Workshop*, p. 313, 2011. arXiv: 1105.1160 [physics.data-an].

[123] G. D'Agostini, "A multidimensional unfolding method based on Bayes' theorem", *Nucl. Instrum. Meth. A*, vol. 362, p. 487, 1995. DOI: 10.1016/0168-9002(95)00274-X.

[124] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST*, vol. 9, P10009, 2014. DOI: 10.1088/1748-0221/9/10/P10009.

[125] P. Z. Skands, "Tuning Monte Carlo Generators: The Perugia Tunes", *Phys. Rev. D*, vol. 82, p. 074 018, 2010. DOI: 10.1103/PhysRevD.82.074018.

[126] R. Corke and T. J. Sjöstrand, "Interleaved parton showers and tuning prospects", *JHEP*, vol. 03, p. 032, 2011. DOI: 10.1007/JHEP03(2011)032.

[127] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 Tune", *Eur. Phys. J. C*, vol. 74, p. 3024, 2014. DOI: 10.1140/epjc/s10052-014-3024-y.

[128] ATLAS Collaboration, "Summary of ATLAS Pythia 8 tunes", ATLAS Note, 2012. Online: http://cds.cern.ch/record/1474107.

[129] A. Buckley, J. Butterworth, D. Grellscheid, *et al.*, "Rivet user manual", *Comput. Phys. Commun.*, vol. 184, p. 2803, 2013. DOI: http://dx.doi.org/10.1016/j.cpc.2013.05.021.

[130] A. Buckley, H. Hoeth, H. Lacker, H. Schulz, and J. E. von Seggern, "Systematic event generator tuning for the LHC", *Eur. Phys. J. C*, vol. 65, p. 331, 2010. DOI: 10.1140/epjc/s10052-009-1196-7.

[131] CDF Collaboration, "Study of the energy dependence of the underlying event in proton-antiproton collisions", *Phys. Rev. D*, vol. 92, p. 092 009, 2015. DOI: 10.1103/PhysRevD.92.092009.

[132] NNPDF Collaboration, "Parton distributions with QED corrections", *Nucl. Phys. B*, vol. 877, p. 290, 2013. DOI: 10.1016/j.nuclphysb.2013.10.010.

[133] NNPDF Collaboration, "Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO", *Nucl. Phys. B*, vol. 855, p. 153, 2012. DOI: 10.1016/j.nuclphysb.2011.09.024.

[134] CMS Collaboration, "Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $\mathrm{t\bar{t}}$ at $\sqrt{s} = 8$ and 13 TeV", CMS Physics Analysis Summary, 2016. Online: http://cds.cern.ch/record/2235192.

[135] NNPDF Collaboration, "Parton distributions from high-precision collider data", 2017. arXiv: 1706.00428 [hep-ph].

[136] V. Bertone, S. Carrazza, and J. Rojo, "APFEL: A PDF Evolution Library with QED corrections", *Comput. Phys. Commun.*, vol. 185, pp. 1647–1668, 2014. DOI: 10.1016/j.cpc.2014.03.007.

[137] S. Carrazza, A. Ferrara, D. Palazzo, and J. Rojo, "APFEL Web", *J. Phys. G*, vol. 42, p. 057 001, 2015. DOI: 10.1088/0954-3899/42/5/057001.

[138] CMS Collaboration, "Measurement of the very forward inclusive jet cross section in pp collisions at $\sqrt{s} = 13$ TeV with CMS", CMS Physics Analysis Summary, 2016. Online: http://cds.cern.ch/record/2146006.

[139] ATLAS Collaboration, "Charged-particle distributions at low transverse momentum in $\sqrt{s} = 13$ TeV $pp$ interactions measured with the ATLAS detector at the LHC", *Eur. Phys. J. C*, vol. 76, p. 502, 2016. DOI: 10.1140/epjc/s10052-016-4335-y.

[140] CMS Collaboration, "Measurement of pseudorapidity distributions of charged particles in proton-proton collisions at $\sqrt{s} = 13$ TeV by the CMS experiment.", CMS Physics Analysis Summary, 2016. Online: http://cds.cern.ch/record/2145373.

[141] Pierre Auger Collaboration, "Depth of maximum of air-shower profiles at the Pierre Auger Observatory: Measurements above $10^{17.2}$ eV and Composition Implications", in *Contributions to the 35th International Cosmic Ray Conference (ICRC 2017)*, 2017, p. 40. arXiv: 1708.06592 [astro-ph.HE].

[142] F. Riehn, "Hadronic multiparticle production with sibyll", PhD thesis, 2015. DOI: 10.5445/IR/1000052699.

[143] K. Werner, I. Karpenko, and T. Pierog, "Collective effects in proton proton and heavy ion scattering, and the "ridge"at RHIC", *Journal of Physics: Conference Series*, vol. 230, p. 012 026, 2010. Online: http://stacks.iop.org/1742-6596/230/i=1/a=012026.

[144] K. Dusling, W. Li, and B. Schenke, "Novel collective phenomena in high-energy proton–proton and proton–nucleus collisions", *Int. J. Mod. Phys. E*, vol. 25, p. 1 630 002, 2016. DOI: 10.1142/S0218301316300022.

[145] C. Loizides, "Experimental overview on small collision systems at the LHC", *Nucl. Phys. A*, vol. 956, p. 200, 2016. DOI: 10.1016/j.nuclphysa.2016.04.022.

[146] E. V. Shuryak, "Quantum Chromodynamics and the Theory of Superdense Matter", *Phys. Rept.*, vol. 61, p. 71, 1980. DOI: 10.1016/0370-1573(80)90105-2.

[147] H. Stoecker and W. Greiner, "High-Energy Heavy Ion Collisions: Probing the Equation of State of Highly Excited Hadronic Matter", *Phys. Rept.*, vol. 137, p. 277, 1986. DOI: 10.1016/0370-1573(86)90131-6.

[148] P. F. Kolb and U. W. Heinz, "Hydrodynamic description of ultrarelativistic heavy ion collisions", 2003. arXiv: NUCL-TH/0305084 [nucl-th].

[149] STAR Collaboration, "Long range rapidity correlations and jet production in high energy nuclear collisions", *Phys. Rev. C*, vol. 80, p. 064 912, 2009. DOI: 10.1103/PhysRevC.80.064912.

[150] ATLAS Collaboration, "Observation of a Centrality-Dependent Dijet Asymmetry in Lead-Lead Collisions at $\sqrt{s_{NN}} = 2.77$ TeV with the ATLAS Detector at the LHC", *Phys. Rev. Lett.*, vol. 105, p. 252 303, 2010. DOI: 10.1103/PhysRevLett.105.252303.

[151] CMS Collaboration, "Observation and studies of jet quenching in PbPb collisions at nucleon-nucleon center-of-mass energy = 2.76 TeV", *Phys. Rev. C*, vol. 84, p. 024 906, 2011. DOI: 10.1103/PhysRevC.84.024906.

[152] ALICE Collaboration, "Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions", *Nature Phys.*, vol. 13, pp. 535–539, 2017. DOI: 10.1038/nphys4111.

[153] T. Martin, P. Skands, and S. Farrington, "Probing Collective Effects in Hadronisation with the Extremes of the Underlying Event", *Eur. Phys. J. C*, vol. 76, p. 299, 2016. DOI: 10.1140/epjc/s10052-016-4135-4.

[154] M. Bahr *et al.*, "Herwig++ Physics and Manual", *Eur. Phys. J.*, vol. C58, pp. 639–707, 2008. DOI: 10.1140/epjc/s10052-008-0798-9.

[155] J. Bellm *et al.*, "Herwig 7.0/Herwig++ 3.0 release note", *Eur. Phys. J.*, vol. C76, no. 4, p. 196, 2016. DOI: 10.1140/epjc/s10052-016-4018-8.

[156] S. Gieseke, F. Loshaj, and P. Kirchgaeßer, "Soft and diffractive scattering with the cluster model in Herwig", *Eur. Phys. J.*, vol. C77, no. 3, p. 156, 2017. DOI: 10.1140/epjc/s10052-017-4727-7.

# Acknowledgements

I would like to thank Prof. J. Blümer and Prof. T. Müller for serving as the referees of this thesis as well as providing the frame in which I was able to work the last few years. At the same time, I am grateful for the the doctoral school KSETA that provided the financial resources for my work as well as an enriching scientific environment.
I am especially thankful for the encouraging support of my supervisor, Dr. Ralf Ulrich. He provided the unique opportunity to work with the CASTOR calorimeter, from data-taking preparation, installation and commissioning, to data analysis and publication – and finally the support to conduct this thesis.
I would like to thank my colleagues in Karlsruhe and beyond who shared the fascination (and sometimes frustration) of CASTOR: Melike Akbiyik, Colin Baus, Igor Katkov, Hauke Wöhrman, Wim Beaumont, Hans van Haevermaet, and Merijn von de Klundert.
Special thanks goes to those who made life and work at the institute a lot more productive and fun, especially Sabine Bucher, the 10:20 coffee-gang and the WAB-crowd.
I am thankful to Tanguy Pierog for many stimulating conversations and the idea to study collective effects with EPOS LHC.
The construction of the two-dimensional calibration setup would not have been possible without the help of Norbert Barenthien, Heike Bolz, Kai Daumiller, and Günther Wörner.
I owe thanks to Ferencz Sikler who provided me with the line tracking algorithm and Sergey Ostapchenko for the idea to study the multiplicity dependent CASTOR energy.
I furthermore would like to thank Paolo Gunnellini for his help with PROFESSOR and generator tuning in general.

More than I could ever express, I owe thanks to my wife Daniela. I am deeply indebted for your love and support, for your patience and prayers. It is the greatest privilege to spend my life with you!