First measurement of very forward jets with CASTOR in CMS to probe the structure of the proton down to $\mathbf{x_{Bjorken} \approx 10^{-6}}$

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

genehmigte
DISSERTATION

Hauke Wöhrmann
aus München

Tag der mündlichen Prüfung: 21.07.2017
Referent: Prof. Dr. Thomas Müller
Korreferent: Prof. Dr. Dr. h.c. Johannes Blümer
Betreuer: Dr. Ralf Ulrich

Abstract

In this thesis very forward ($-6.6 < \eta < -5.2$) jets are introduced and measured for the first time in proton-proton collisions at 13 TeV. In high energy particle collisions the modeling of final state hadron production over a wide range in phase space is still an unsolved problem in theory. The effects known as "underlying event" make a significant contribution to the physics in such collisions. The main manifestations of this are significant difficulties in estimating the backgrounds in LHC collisions, and to simulate high energy cosmic ray air showers. For a study of hadronic particle production, a jet measurement with the CASTOR calorimeter at the CMS experiment is performed using data recorded during the low intensity startup of LHC Run-2. Due to the position of the detector in the very forward phase space it is possible to probe down to very low-x values of the proton momentum fraction. This unique measurement is sensitive to extreme kinematic regions in QCD collisions, providing an excellent opportunity to study the impact of the underlying event in this phase space. To achieve this challenging measurement the detector reconstruction software is improved and a new alignment procedure for the calorimeter is developed. For the first time a joint TOTEM-CMS trigger is introduced, and the data recorded with this trigger is used for the most precise detector alignment of CASTOR so far. Finally it is shown that the measured results are very sensitive to multi-parton interactions, and that exclusive dijet events in CASTOR can be used to maximize the sensitivity to low-x parton kinematics.

Zusammenfassung

Diese Arbeit präsentiert stark vorwärts gerichtete Jets, die zum ersten Mal bei 13 TeV in Proton-Proton Kollisionen gemessen wurden. In hochenergetischen Teilchenkollisionen ist das Berechnen und Simulieren von hadronischen Endzuständen ein in der Theorie immer noch ungelöstes Problem. Effekte wie das sognannte "underlying event" (UE) haben einen nicht unerheblichen Anteil an der Physik bei solchen Kollisionen. Dies ist dafür verantwortlich, dass der Untergrund bei Beschleunigerexperimenten sowie die Entwicklung von Luftschauern erzeugt durch Höhenstrahlung nur schwer abzuschätzen sind. Um die hadronische Teilchenproduktion zu studieren, wurden hier Jets bei niedriger Strahlintensität während des LHC Run-2 aufgezeichnet. Wegen der günstigen Position des Kalorimeters in der Vorwärtsrichtung liegend ist man in der Lage, bis zu sehr kleinen Werten von x_{Bjorken} im Proton zu messen. Diese Messung ist sehr empfindlich auf diesen kinematischen Bereich der QCD und eine Möglichkeit die Effekte des UE zu untersuchen. Damit diese Messung gelingt, ist es nötig, die Rekonstruktionssoftware zu verbessern und die genaue Position des Kalorimeters innerhalb von CMS zu bestimmen. Hierfür wurden Daten zum erstem Mal kollaborationsübergreifend mit TOTEM und CMS gemeinsam aufgenommen und ein spezieller TOTEM-CMS Trigger entwickelt. Schlussendlich zeigen die aufgenommenen Daten, dass die Physik im untersuchten Bereich sehr sensitiv auf Effekte wie Multiparton-Kollisionen reagiert und dass Ereignisse mit zwei Jets in CASTOR dazu geeignet sind, die Kinematik von Partonen bei besonders niedrigen Werten von x_{Bjorken} zu studieren.

Contents

1 Introduction

A detailed understanding of hadronic multiparticle production induced by quantum chromodynamics is of paramount importance for any precision measurement and search for new physics at the LHC. An important contribution to multiparticle production is related to the low-x parton kinematics which is in particular poorly known. Therefore, a very good understanding of the underlying event [1] as well as multi-parton interaction [2] are necessary since they represent a huge background for all processes investigated at the LHC (see Section 2). Only with a precise modeling of these effects is any relevant LHC study possible. Measurement and tuning of the underlying event activity at the LHC is a core task of the experiments [3–7]. In general this physics is most dominant at small-x and soft p_{T}. Typically the underlying event is studied by measurements in the phase space region which is most separated from the central high-p_{T} collision. Here the CASTOR calorimeter offers the most forward acceptance at the LHC, which has already been exploited to measure the underlying event activity at 0.9, 2.76 and 7 TeV [8]. A corresponding study at 13 TeV is one of the results of this thesis.

Furthermore, for the description of ultra-high energy cosmic ray air showers one needs to extrapolate measurements performed at accelerators to the forward phase space where most center-of-mass energy is transported to. This can only be done with phenomenological models since no fundamental theoretical description is possible. The problem to consistently describe air showers is very persistent [9–11] and the most recent manifestation of this is the large muon overabundance observed by the Pierre Auger observatory [12, 13]. In order to ultimately advance the field of ultra-high energy cosmic ray physics it is understood that a precise mass composition analysis at the highest energies is needed. This is not possible at the moment due to the large uncertainties in the extrapolations done by the models [14, 15]. It is important to realize that it is not only the highest center-of-mass energies at the LHC but in particular also the extrapolation into the forward phase space, and nuclear effects, that are most relevant in this context. In this thesis I present results of unique measurements at the most forward region accessible at the LHC for full hadron calorimetry.

The main mechanism of hadronic particle production at high energies is the fragmentation of partons leading to the formation of *jets*. The ideal detector to study small-x and low p_{T} is the CASTOR calorimeter in the very forward region of CMS at a pseudorapidity of $-6.6 \leq \eta \leq -5.2$ (see Section 4). For this purpose the data taken at $\sqrt{s} = 13\,\mathrm{TeV}$ in 2015 are used in this thesis. Very forward jets are reconstructed and analyzed as a proxy of

small-x parton dynamics inside highly energetic protons. It is found that the inclusive very forward jet production spectra are an excellent tool to study multi-parton interactions.

To perform a precise measurement, the CASTOR calorimeter needs to reconstruct very energetic jets. In the recording and reconstruction of such objects with individual energies up to the TeV level it needs to be assured that over the whole dynamic range the detector can perform a valid measurement. One typical problem occurring in such events is signal saturation. It is one task of this work to develop an algorithm using the property of a stable pulse shape of the detector to recover the saturated signal by using the pulse tail (see Section 4.2).

Due to the very forward location of the calorimeter the exact position of the detector represents a major uncertainty. Small changes of the position can result in big changes of the response since the energy flow, $\mathrm{d}E/\mathrm{d}\eta$, distribution is very steep in this acceptance range. Here the TOTEM detector is ideal to use for a detector alignment. The T2 tracking detector of TOTEM has the same acceptance as CASTOR and is placed directly in front of the calorimeter. By using single charged particles, such as electrons, a precise alignment depending on the observed tracks is performed (see Section 4.4). Since TOTEM and CMS are different collaborations, common data taking is not trivial but has been done during this thesis. For the recording of CMS events with combined TOTEM data the first joint TOTEM-CASTOR trigger is developed and commissioned (see Section 4.3).

The primary physics goal is measuring the inclusive jet spectrum with CASTOR. It is found that the CASTOR energy scale uncertainty has a major limiting impact on the precision of the measurement. Thus, it is proposed that not only the jet production cross section is important but also the shape of the spectrum. Furthermore, specific back-to-back dijet events in CASTOR are finally exploited to maximize the sensitivity with respect to the very small-x gluon parton densities down to $x_{\mathrm{Bjorken}} \approx 10^{-6}$. The correlation between very forward jets and central particle production is used as a tool to study the impact of multi-parton interactions.

It is found that on the level of the inclusive jet cross section spectrum all models can describe the data inside the experimental uncertainties. But the models show major difficulties in predicting the shape of the distributions for dijet events. None of the models is found to describe all aspects of the data (see Section 5).

The measurements reported in this thesis are unique at LHC and help to better understand the very low-x parton dynamics in hadrons as well as the relative importance of the underlying event with respect to low p_{T} jet particle production.

2 Description of hadron collisions

2.1 Quantum chromodynamics

The strong interactions between quarks are described by the theory of quantum chromodynamics (QCD). It is part of the standard model which results from the unification of the electromagnetic, weak and strong interactions.

The standard model of particle physics is the theory which describes quarks and leptons and their interactions. It is a relativistic field theory based on the symmetry groups $\mathrm{SU(3)_C} \times \mathrm{SU(2)_L} \times \mathrm{U(1)_Y}$ with gauge bosons (photons, Z, W gluon) resulting from the local symmetries. The standard model describes fermions, which are quarks and leptons with half integer spin, and the already mentioned bosons with integer spin. Elementary particles are characterized by their electric charge, mass, flavor and color.

The QCD itself results from the non-abelian gauge theory $\mathrm{SU(3)_C}$ and introduces three colors with gluons as force carrier. The dynamics of quarks and gluons are expressed by the Lagrangian

$$\mathcal{L}_{\mathrm{QCD}} = -\frac{1}{4} F^A_{\mu\nu} F^{\mu\nu}_A + \sum_q \bar{\psi}_{q,a} \left(i\gamma^\mu \partial_\mu \delta_\mu - g_s \gamma^\mu t^C_{ab} \mathcal{A}^C_\mu - m_q \delta_{ab} \right) \psi_{q,b}, \tag{2.1}$$

which sums over repeated indices [16]. The quark field $\psi_{q,a}$ represents a quark with the flavor q, as well as color indices a and b running from 1 to 3 and with a mass m_q while $\mathcal{A}^C_\mu$ corresponds to the gluon fields with C that runs from 1 to 8 for the eight color degrees of freedom of the gluon field. The strength of the interaction between the standard generator matrices of the $\mathrm{SU(3)_C}$ group t^C_{ab}, and the quark and gluon fields is determined by the coupling g_s (or $\alpha_s = g_s^2/4\pi$). The gluon fields $\mathcal{A}^C_\mu$ and the coupling g_s are also represented in the Lagrangian field tensor $F^A_{\mu\nu}$ given by

$$F^A_{\mu\nu} = \partial_\mu \mathcal{A}^A_\nu - \partial_\nu \mathcal{A}^A_\mu + g_s f_{ABC} \mathcal{A}^B_\mu \mathcal{A}^C_\nu, \tag{2.2}$$

where f_{ABC} are the structure constants of the $\mathrm{SU(3)_C}$ group and defined such that $[t^A, t^B] = i f_{ABC} t^C$. The last term in the field tensor $F^A_{\mu\nu}$ gives also rise to the triplet and quartic gluon self interaction since gluons also carry a color charge. This is particularly different from all the other fundamental forces [16].

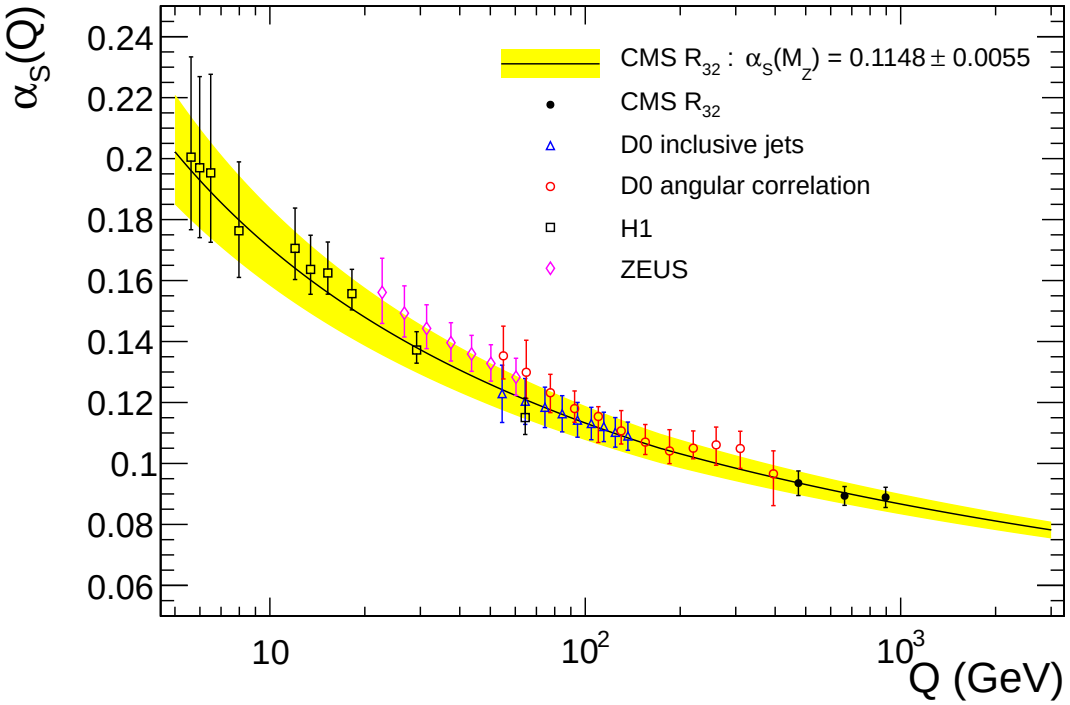


Figure 2.1: Compilation of measurements of the dependence of α_s on Q. [17]

The strong coupling constant α_s depends on the scale $Q^2 = -q^2$, where q is the momentum transfer. The dependence of α_s on Q is shown in figure 2.1 and is well approximated by the empirical formula

$$\alpha_S(Q) = \frac{12\pi}{(33 - 2n_f)\log(Q^2/\Lambda^2)}, \tag{2.3}$$

with n_f as number of active quark flavors and where Λ is a scale parameter below which the perturbatively defined coupling starts to diverge [18].

This behaviour of the coupling constant is responsible for two major properties of QCD. Processes characterized by a high value of Q can be calculated correctly by the perturbation theory since the coupling is weak. For processes which are characterized by a small value of Q the perturbative approach of calculation is not usable because α_s gets too high. The scale of Q where this transition occurs is represented by Λ. In particular the calculation of bound states of quarks into hadrons is therefore very difficult and non perturbative.

To describe the structure of hadrons and how they interact Feynman proposed the parton model [19], according to a parton can be both quarks and gluons, carrying the momentum fraction x of the hadron. The fundamental consequence of this model is that the interaction between hadrons must be described by the interaction of their involved partons. Hence the hadron can be described by the so-called parton distribution function (PDF) which represents the probability of a parton to carry a fraction x of the total hadron momentum.

The final hadron scattering cross section can thus be factorized into parton level cross sections σ_i, convoluted with the PDF f_i for parton i according to:

$$f_i(\mu_F^2) \otimes \sigma_i(\mu_F^2, \mu_R^2). \tag{2.4}$$

The factorization scale μ_F is introduced here to distinguish between these two regimes and μ_R is the renormalization scale. In particular the inelastic cross section for the collision of two hadrons is given by

$$\sigma_{\text{inel.}} = \sum_{i,j} \int \mathrm{d}x_1 \mathrm{d}x_2 f_i(x_1, \mu_F) f_j(x_2, \mu_F) \sigma_{ij}(x_1, x_2, \mu_F, \mu_R) \tag{2.5}$$

where x_i represents the momentum fraction carried by parton i of the proton. The σ_{ij} represents the hard scattering cross section between two partons [20, 21].

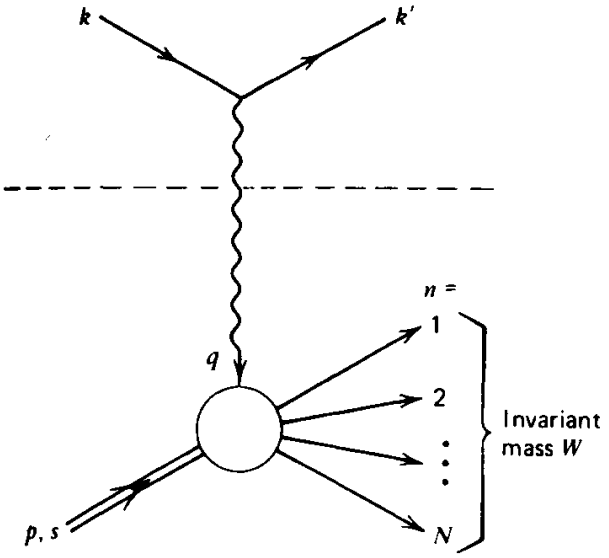


Figure 2.2: Lowest order Feynman diagram of a deep inelastic electron proton scattering. [22]

The internal structure of hadrons was discovered at HERA [23] in deep inelastic scattering (DIS) processes. Hereby an electron is scattering with a proton $ep \to eX$ where the target proton does not stay intact. In figure 2.2 the Feynman diagram in the lowest order of this process is shown. The electron with the initial momentum k and the final momentum k' exchanges a virtual photon with the proton of the invariant mass M. The energy which is transferred by the electron can be expressed by the Lorentz invariant variable

$$\nu = E - E' = \frac{p \cdot q}{M}, \tag{2.6}$$

where q is the transferred momentum and p the initial proton momentum. A quantity for the inelasticity of the scattering is the dimensionless Bjorken scaling

$$x = \frac{-q^2}{2p \cdot q} = \frac{-q^2}{2M\nu} = \frac{Q^2}{2M\nu}. \tag{2.7}$$

With the above variables the resulting invariant mass W can be then expressed by

$$W^2 = (p+q)^2 = p^2 + 2p \cdot q + q^2 = M^2 + 2M\nu + q^2 = M^2 + \frac{Q^2}{x} - Q^2. \tag{2.8}$$

One can see that in the case of an elastic scattering $W^2 = M^2$ the Bjorken scaling becomes $x = 1$ and $x < 1$ in inelastic scattering processes. In the parton model the scattering process is then considered as the interaction between the electron and the parton of the nucleon. If the mass and the transverse momentum of the partons are neglected, then the Bjorken scaling x represents the fraction of nucleon momentum carried by the parton at the moment of interaction.

In cases where the $Q^2 \gg M^2$, the cross section of DIS can be derived by

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}\Omega\,\mathrm{d}E'} = \left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{\text{Mott}} \times \left[\frac{1}{\nu}F_2(Q^2,\nu) + \frac{2}{M}F_1(Q^2,\nu)\tan^2\frac{\theta}{2}\right], \tag{2.9}$$

where the Mott cross section

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{\text{Mott}} = \frac{4\alpha^2 E'^2}{q^4}\cos^2\frac{\theta}{2} \tag{2.10}$$

is the differential cross section for the relativistic Coulomb scattering of the electron, also taking into account its spin [18]. Thereby E' and θ respectively are the energy and polar angle of the final state electron. Here α is the structure constant characterizing the electromagnetic coupling and not the strong coupling constant. The final differential cross section of the DIS is therefore the electromagnetic one extended by a dependency of the structure functions $F_1(Q^2,\nu)$ and $F_2(Q^2,\nu)$. These functions encode the inner structure of the nucleon and their interaction with the photon. Due to the fact that a proton consists of spin 1/2 particles F_1 can be express by F_2

$$2xF_1(Q^2,\nu) = F_2(Q^2,\nu), \tag{2.11}$$

which is known as Callan-Gross-Relation [24]. If the nucleon consisted only of three quasi-free Dirac particles, the valence quarks, the structure function would have no dependence on Q. In terms of the already mentioned particle distribution functions the structure function can be rewritten as the sum over the partons

$$F_2(x) = \sum_i e_i^2 f_i(x) \tag{2.12}$$

with e_i as the electric charge of the parton i. In figure 2.3 the measured cross section depending on Q^2 at HERA in deep inelastic proton electron scattering is shown for different values of x. The dependency of Q or x in the cross section 2.9 is directly correlated to the structure function. Since the gluons can interact with quarks and with themselves there is a small residual dependence on the momentum transfer Q which is also visible in figure 2.3.

The PDF itself cannot be calculated by perturbative theory but is directly measured by several experiments, e.g. at HERA. In figure 2.4 the measured PDF from HERA results for a certain Q^2 are shown. At high values of x one can see a peak of the valence up and down quarks. This corresponds to the picture of a low energy confined proton from two up and one down quark. Figure 2.4 shows also that for small x the density of so-called see quarks (xS) and gluons (xg) becomes dominant. In the cross section calculation higher orders of QCD perturbative corrections are absorbed by renormalization of the PDFs.

The concept of the factorization of the cross section is an approximation. One of the main problems is that any correlations between partons are neglected. Also, the transverse momentum of the partons is not included in the most common approach. Furthermore, to be consistent the PDF and partonic cross section both depend on the factorization scale. The exact structure of the PDF cannot be calculated from the theory.

However, the evolution of PDFs can be approximated with the Dokishitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) [26–28] or Balitsky-Fadin-Kuraev-Lipatov (BFKL) [29–31] evolution equations. They describe the evolution of the PDF with the momentum fraction x and momentum transfer Q. If the parton density is known at a certain value, the full PDF can be predicted by the DGLAP equations. From the DIS point of view this includes also higher order couplings between the photon and gluon interaction. This results also in an explicit dependency of the structure function F_2 on α_s and therefore also on Q^2 as shown in figure 2.3. In the region of small x and low values of Q^2 the BFKL equations become more valid than DGLAP equations. Since there are different approaches to describe the PDF, there are also different PDF versions determined by different groups.

In several experiments the measured data were used to determine the structure of the PDF. The most precise measurements for PDFs are the DIS data. But also Drell-Yan production and in particular the formation of Z^0 and $W^{\pm}$ are important probes of the quark content. Even the Higgs boson production has been proposed as a probe of gluons in the proton [32]. And, finally, the formation of QCD jets is of course a messenger from the internal parton structure.

One common PDF set combines data from the experiments H1 [33] and ZEUS [34] at the HERA collider and uses it for the calculation of the HERAPDF1.5 [35, 36], which is also shown in figure 2.4. The ABM collaboration also uses a combination of its previous results and HERA-I data [25] to provide ABM11 PDF [37]. Based especially on D0 Run-2 data [38, 39] at Tevatron the CT collaboration provides the CT10 NLO PDF set [40] and CT10 NNLO PDF set [41, 42] with next-to-leading-order (NLO) and next-to-next-to-leading-order (NNLO) calculations. A set of PDFs using also LHC data from the ATLAS [43, 44], CMS [45] and LHCb [46] experiments is NNPDF3.0 [47]. Another common PDF set which is also based on NNLO perturbative calculations and uses data from HERA, Tevatron and LHC is MSTW2008 [48] and its successor MMHT14 [49].

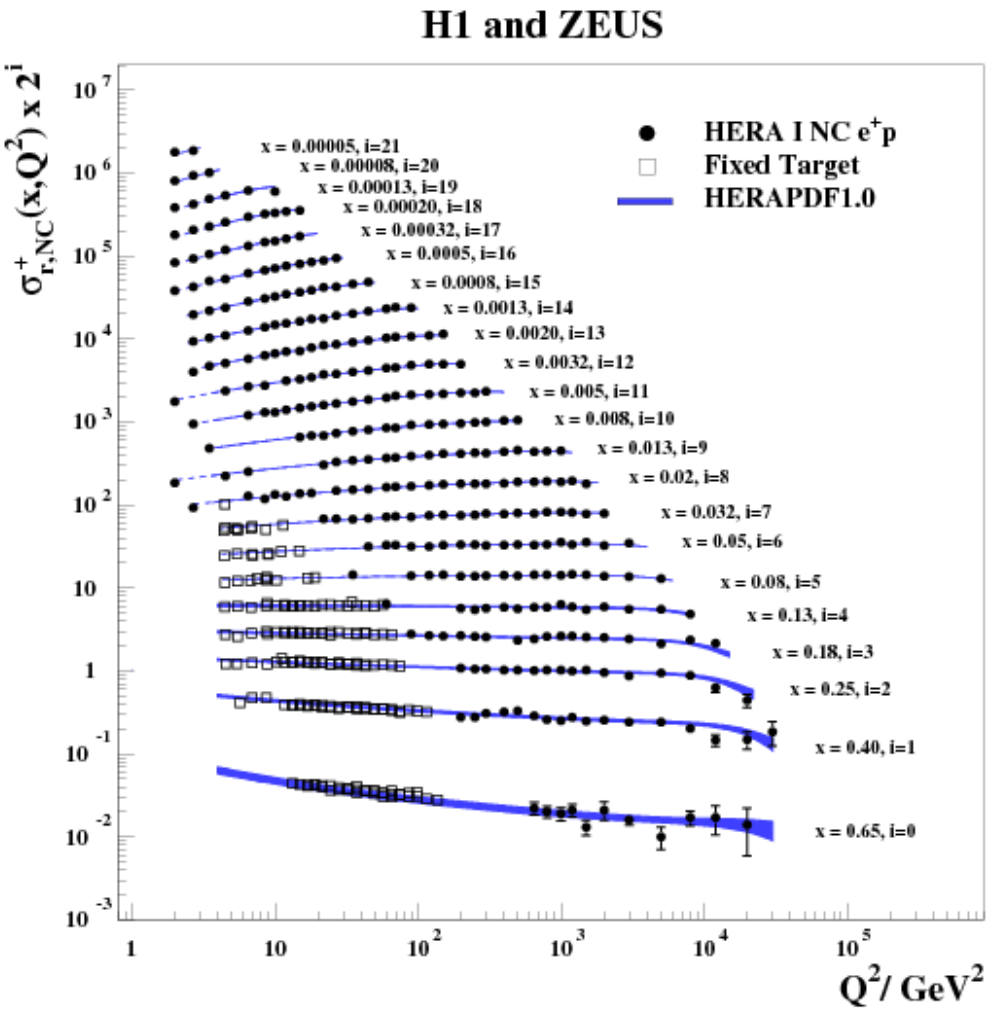


Figure 2.3: The measured cross sections directly depending on the structure function $F_2(x, Q^2)$ as function of Q^2 for different values of x in deep inelastic scattering at the HERA accelerator. [25]

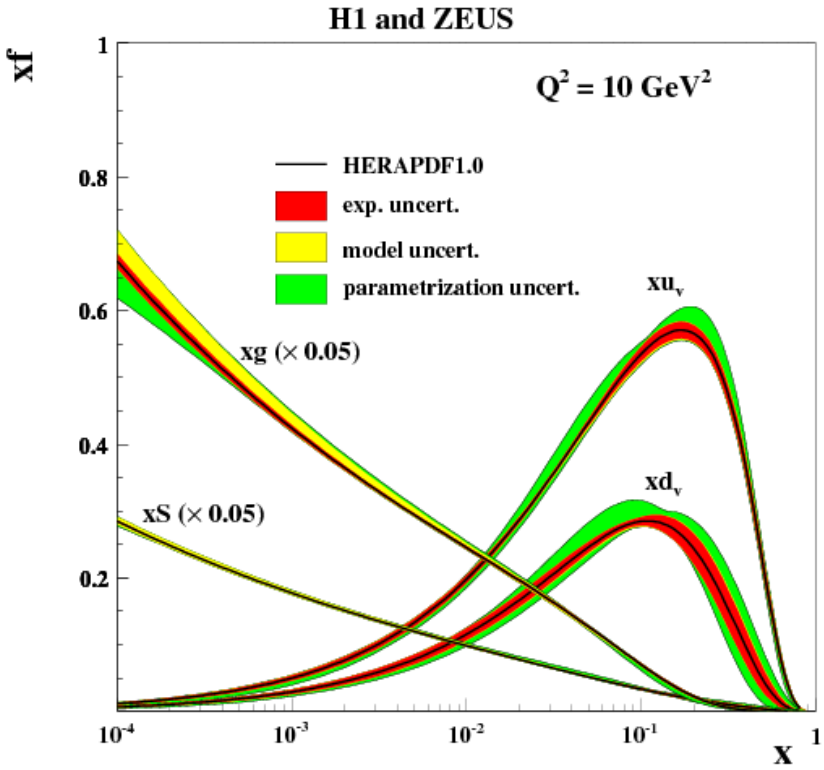


Figure 2.4: The parton distribution function $f_i(x)$ as function of x at a momentum transfer $Q^2 = 10\,\text{GeV}$. The contribution of the see quarks (xS) and gluons (xg) is scaled down by a factor of 20. [25]

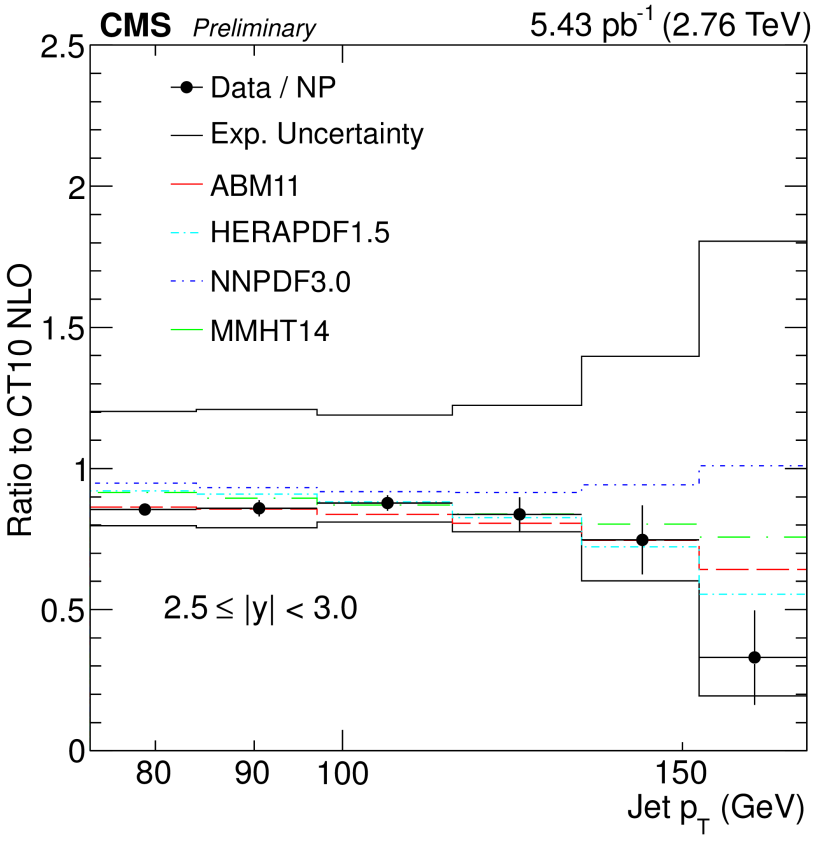


Figure 2.5: The ratio between the observed inclusive jet cross section in p_T and its theoretical prediction derived with the CT10 NLO PDF. The experimental systematic uncertainty is indicated by the solid black band. In colored dashed and dotted lines the ratios of theoretical predictions derived with other PDFs to the one derived with the CT10 NLO PDF are shown. [50]

2.2 Hadronization

At a larger distance and therefore smaller values of Q the strong coupling α_s is increasing as shown in figure 2.1. The consequences are that the partons cannot exist independently and form color neutral hadrons. This process is called hadronization. A phenomenological model to describe hadronisation is the concept of string fragmentation [51]. Thereby the partons are mathematically connected by hypothetical strings whose strength is directly correlated to the color field strength of the connected partons. From QCD the potential

$$V = -\frac{4\alpha_s}{3r} + kr \tag{2.13}$$

can be understood as a combination of a Coulomb like part and a linear increasing part with the distance r [52]. Imagine two quarks would move away from each other, due to the linear part in the potential the stored energy in the color field between the quarks is growing. At a certain point the stored energy is high enough to create a new quark anti-quark pair and the string breaks into two new strings. When the energy of this new state is high enough the breakup will go on further. This continues until all energy is used and only color neutral states remain which are identified as hadrons.

In such a mechanism a series of mesons but also baryons can be produced. This concept is also responsible for the dissipation of flavor and momentum, too. Therefore a scattered parton "hadronizes" in a collimated stream of hadrons which are typically called jets. These jets carry the momentum of the parton and are ideal to measure processes in hadron-hadron collisions. Hence the inclusive jet cross section is correlated to the parton cross section and thus also correlated to the shape of the PDF used in the calculations. This behaviour can be seen in figure 2.5 where the ratio of theoretical prediction of the inclusive jet cross section for different PDF to the predictions by the CT10 PDF [40] is shown. Also the ratio between the CT10 predicted jet spectrum to the one measured at CMS is illustrated by the black dots.

2.3 The underlying event in hadron collision

The previous description of the factorization theorem for hadron-hadron scattering only regards a single parton-parton interaction. But in general the final state of a hadron-hadron collision is influenced by many additional effects in a very important way and called underlying event [1].

One contribution appears at high energies where the momentum fraction x of the involved partons get smaller, leading to a huge increase in the parton densities. This fact causes a rise of the probability of parton-parton interactions. These interactions occur in addition to the hard parton-parton collision derived from the factorization theorem and are typically known as Multiple Parton Interactions (MPI). Another effect is that partons can

also radiate gluons and the gluons can introduce further parton showers. Depending on when this happens it is called Initial State Radiation (ISR), occurring before the main hard parton-parton interaction, or Final State Radiation (FSR), occurring afterwards. Also the remaining beam remnants, which are not color neutral and must hadronize, add contributions in the very forward direction. It would be possible to calculate the exact perturbative cross section taking into account all higher order corrections, the effects of the underlying event would be already included except for the beam remnants.

2.4 Different models and tunes

In Monte Carlo generators there are two different methods to simulate the hard scattering processes with following hadronization. Generators as Pythia8 [53] are based indeed on the factorization theorem to describe hadron-hadron collisions. They use DGLAP evolution equations to calculate the parton density function and take into account MPI with ISR/FSR. They also handle the beam remnants and perform hadronization with the string fragmentation method. For Pythia8 several tunes on experimental data exist such as 4C and Monash13. A tune is thereby a set of configuration parameters inside the generator in which normally the parameters are tuned to experimental results.

One of the basic Pythia8 tunes is 4C [54] which is obtained by using LHC 7 TeV data. Therefore, experimental studies of minimum bias and underlying event data in ATLAS [55, 56] and ALICE [57] were used as charged particle multiplicity and track p_{T} distribution. The name of the tune 4C reads as the fourth version of the tune using the CTEQ6L1 PDF set. Previous versions do not use LHC data and are based on Tevatron data. These previous tunes were also the starting point to fit LHC data while especially MPI and color reconnection parameters needed to be changed to get the final parameter set for tune 4C. This tune is often used when Pythia8 simulations are done together with MBR (Minimum Bias Rockefeller) simulation [58]. Here in the simulation process single, double and central diffraction-dissociation processes are included explicitly and contribute to the total proton-proton cross section.

Another Pythia8 parameter set which is not only obtained by proton-proton data from LHC and Tevatron but also by electron-electron data from LEP is the Monash13 tune [59]. Here, too, experimental results from minimum bias and underlying event data were used together with the NNPDF2.3 PDF set [60, 61] to fit the Monash13 parameters. Based on the parameter values of tune 4C is a CMS tune on underlying event data called CUETP8S1 [3]. Here the word CUET stands for "CMS Underlying Event Tune" and P8S1 for Pythia8 "Set 1". As the name emphasizes this tune is determined by looking into UE observables measured in CMS using also as 4C the CTEQ6L1 PDF set. The same tune on the data is repeated with the HERAPDF1.5 PDF set which results in the CUETP8S1 HERAPDF tune. The tune also exists using as starting tuning parameters from the Monash13 tune. In this case is is called CUETP8M1 and is based as Monash13 on the NNPDF2.3 PDF set.

A fundamentally different approach is taken by the cosmic ray generator as EPOS [62] and QGSJetII [63], which are based on the Gribov Regge field theory [64]. These generators were originally developed to describe extensive air showers induced by cosmic ray particles and therefore are tuned on hadron-nucleon or nucleon-nucleon interactions. They descibe the non perturbative part in terms of exchanging virtual particle states, the pomeron. But at high energy scales also these generators describe the perturbative part with QCD and DGLAP equations. The description of MPI effects is achieved here by multiple exchanges of the pomeron.

In QGSJetII the elementary hadron interaction is described by a parton cascade. In particular, the process is divided by a cutoff value Q_0 into a soft and semi-hard parton dynamics. The soft cascades are described by a phenomenological "soft" pomeron description while the "semi-hard" scattering process is composed of a DGLAP QCD parton ladder which is sandwiched between two soft pomerons. At higher energies non linear parton effects become important when the individual parton cascades start to overlap in the corresponding phase space. These non linear effects are taken into account by including pomeron-pomeron interactions to all orders. The inclusion of pomeron-loop and pomeron-tree contributions leads to a fundamental modeling of all types of diffractive final states.

Also in EPOS a modelling of soft and semi-hard pomeron phenomenology is used. But in addition, EPOS also accounts for energy momentum correlations between multiple re scatterings at partonlevel. The treatment of non-linear parton effects is based on the description of lowest order pomeron-pomeron interactions. The physics is reflected in the term "EPOS" which stands for Energy conserving quantum mechanical multiple scattering, based on Partons, parton ladders and strings, Off-shell remnants and Splitting of parton ladders. EPOS does not only take "open" parton ladders into account but also "closed" parton ladders which represent, elastic scatterings. These don't contribute to the final state but are crucial for the final cross section calculation. EPOS also includes phenomenological "high-density" effects, where a specific energy density per phase-space by a hydro-dynamic flow of the parton-fluid is simulated. Such a fluid phase was first identified in proton-proton collisions by the CMS collaboration [65].

3 Experimental setup

The results in this thesis are based on proton and heavy ion collision data provided by the Large Hadron Collider (LHC) and recorded with the Compact Muon Solenoid (CMS) experiment.

3.1 The LHC

In Geneva near the France-Switzerland border, the European Organization for Nuclear Research (CERN) operates and maintains the LHC collider.

With a diameter of 8.5 km and a circumference of 26.7 km the LHC is the largest hadron accelerator in the world. It is installed in the existing tunnel system of the CERN LEP machine and located between 45 m and 170 m below the surface. There are four major experiments placed at the interaction points (IP) where the beam particles collide. At LHC point 1 (IP1) and LHC point 2 (IP2) ATLAS and ALICE are installed while LHC-B is at IP8. On the opposite side of ATLAS at LHC point 5 (IP5) near the French village of Cessy the CMS experiment is performed whose data is used in this thesis [67].

The accelerator is based on two rings and an overall number of around 2500 superconducting magnets to accelerate and focus the beam particle packages. It is designed to accelerate protons up to kinetic energies of 7 TeV. The stored energy of 362 MJ in the circulating beams and 600 MJ stored in the superconducting magnets amounts to a total energy stored in the ring of around 1 GJ. A "Eurofighter" jet ($m \approx 11$ t) travelling with a speed of Mach 1.5 at 10 km altitude where 1.5 Ma = 450 m/s has the same kinetic energy. The magnets are cooled down to 1.7 K by superfluid helium. During nominal runtime the dipole magnets to accelerate the particles have a 8.3 T field and around 11.3 kA current. To squeeze the beam and focus it the quadrupoles have a 6.8 T field and work with the same current as the dipole magnets. The luminosity loss of the beams is very small and originates mainly from the collisions over a long period of time. Hence, inside the beam pipe an ultra high vacuum is established which is stabilized due to leakage at a pressure of 10^{-6} mbar. Along the ring structure, not near the interaction points of the experiments, the inner tube has a diameter of 50 mm. A cross section through the beam pipe with a dipole magnet and the supporting structures can be seen in figure 3.2. [67]

During operation CERN stated that its total consumption of electrical energy is around 1 TWh over a year, which is roughly a fifth of the consumption by the whole canton of

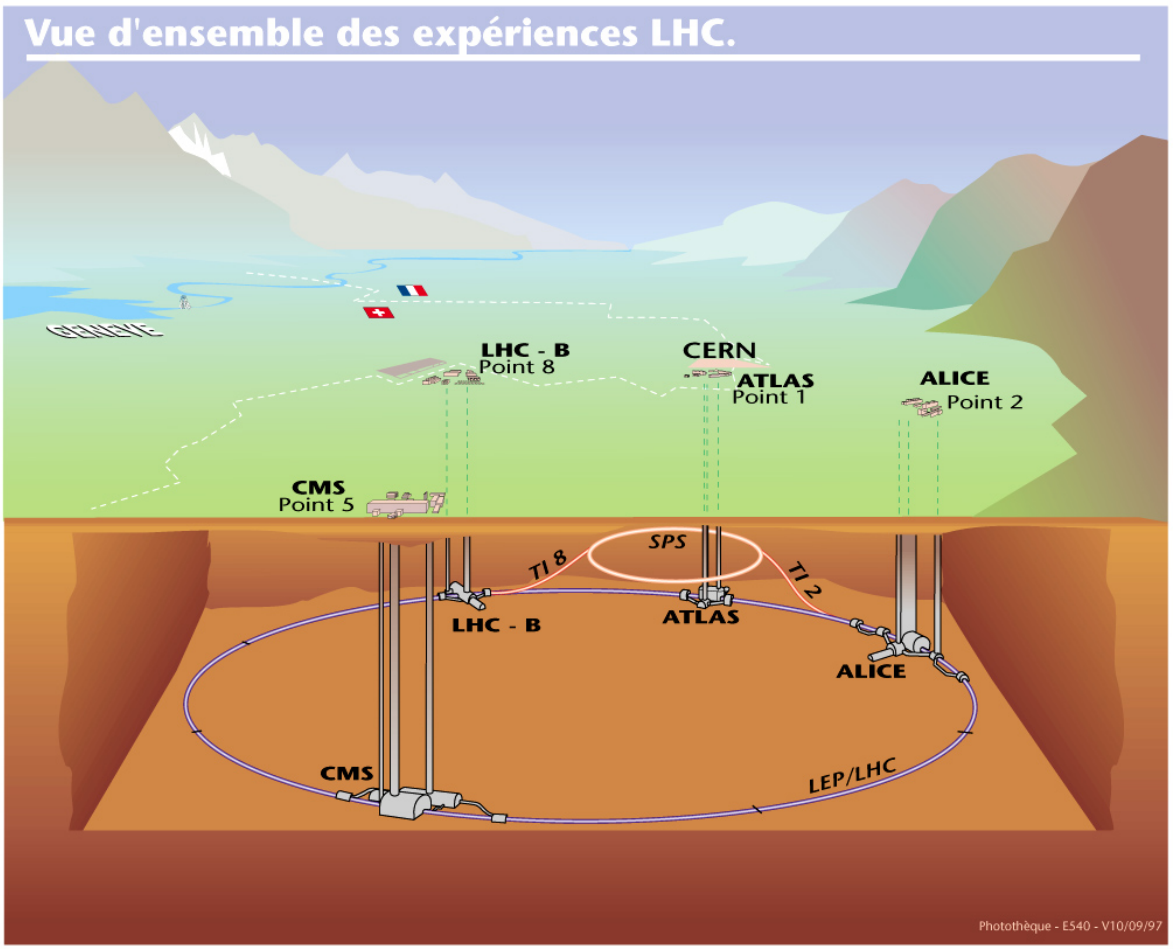


Figure 3.1: Overall view of LHC experiments. [66]

Geneva. With up to 45 % of that the LHC accelerator has the largest contribution to this consumption. For the cryogenics and the cooling alone around 200 GWh is needed. An additional rate of 10 to 12 % is consumed for the experiments at the LHC. [69]

The beam particles do not start from zero energy to be accelerated in the LHC ring. They will be pre-accelerated in smaller rings and then induced finally into the LHC ring. For protons the journey starts with an hydrogen bottle where the hydrogen will be first stored in a chamber. By align an electric field the electrons and protons are forced to separate. After separation the linear accelerator LINAC2 speeds up the protons and they are induced into the Proton Synchrotron Booster (BOOSTER) (see figure 3.3) where the energy is increased to 1.4 GeV per particle. On the way to the LHC these proton bunches go through the Proton Synchrotron (PS) where the bunches are also grouped into a train of bunches with the LHC spacing of 25 ns. From the PS the bunches leave with 25 GeV and travel through the Super Proton Synchrotron (SPS) where they finally reach an energy of 450 GeV and can be transferred into the LHC ring. This procedure is repeated until up

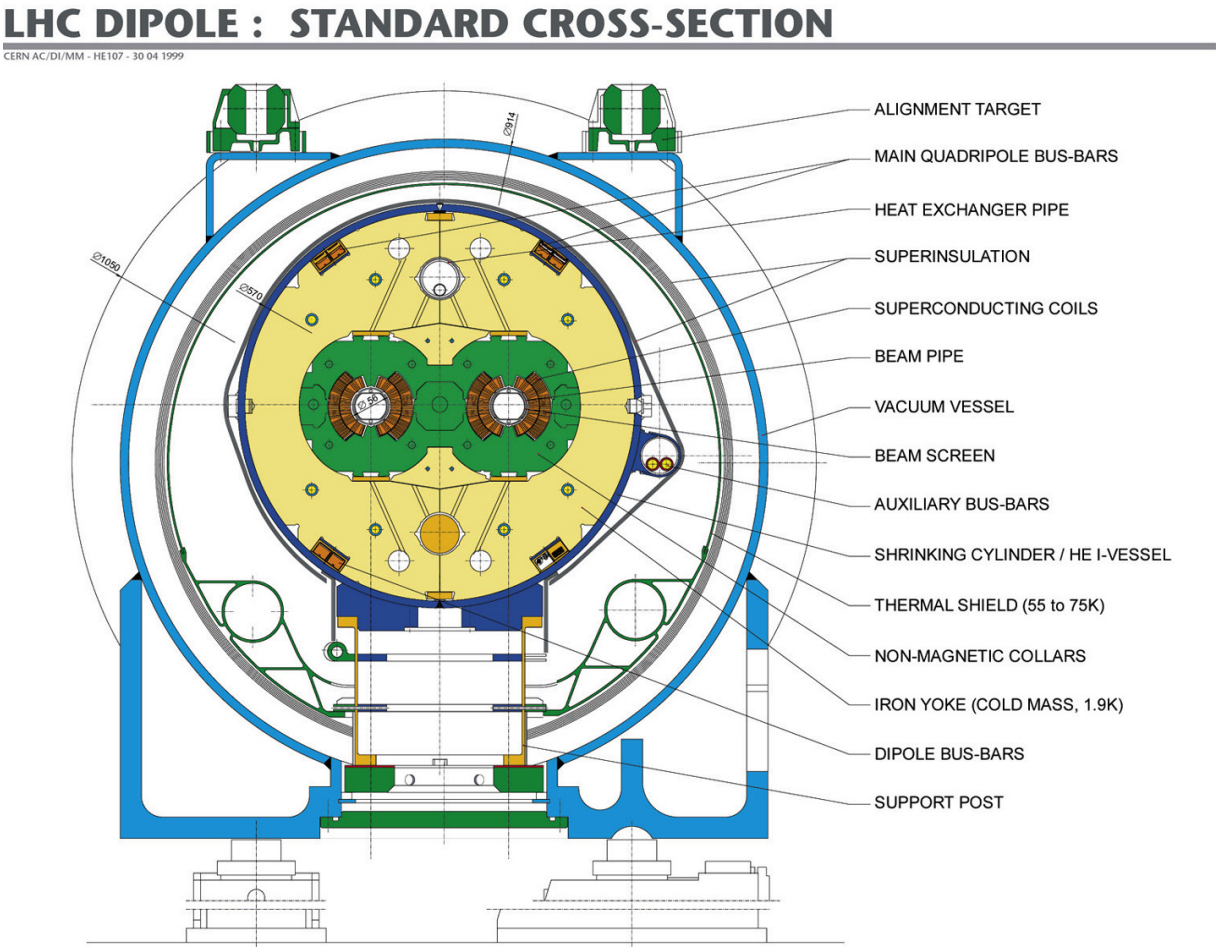


Figure 3.2: Diagram showing the cross-section of an LHC dipole magnet with cold mass and vacuum chamber. All lengths are in mm. [67, 68]

to 2808 bunches are circulating in both directions in the LHC ring. For heavy ions the injection chain looks similar except that here the particles start at LINAC3 and are then speed up by the Low Energy Ion Ring (LEIR) before they are injected into the PS. In the so-called ramp up period the beam is accelerated in the LHC to the distinct TeV energies. Then the bunches are focussed by the magnets to increase the instantaneous luminosity and start to record the collision. In the following especially the data recorded by the CMS experiment is discussed.

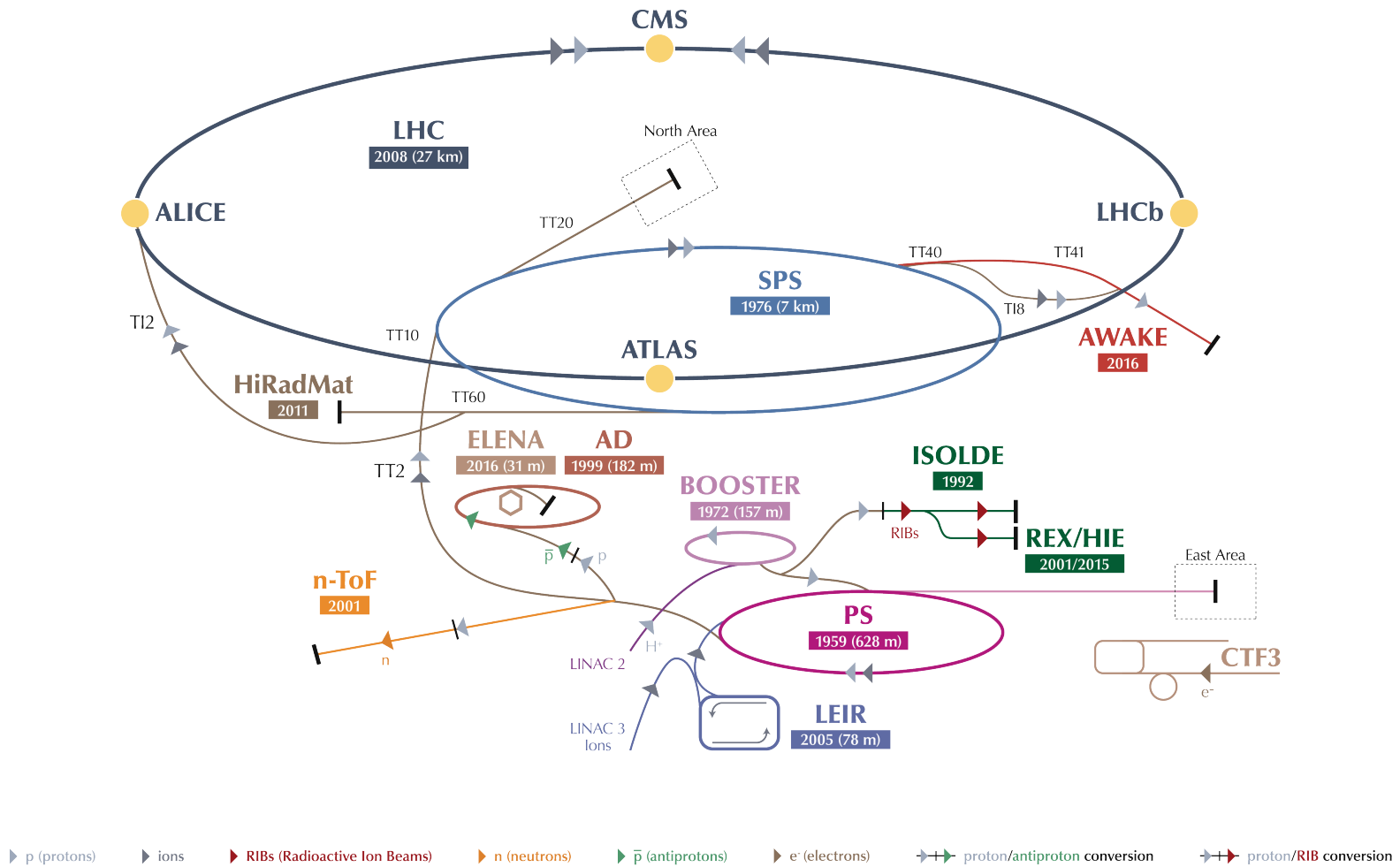


Figure 3.3: The CERN accelerator complex. [70]

3.2 The CMS experiment

The Compact Muon Solenoid (CMS) is one of the four major experiments using the data of LHC. CMS is installed in a cavern 100 m below the surface at IP5. With the CMS detector various aspects of elementary particle physics can be studied. For this purpose, the CMS detector measures high energy collisions of protons or heavy ions with up to 100 collisions in one single beam crossing. The detector consists of different layers of subdetectors to achieve a good performance and reconstruction of the outgoing products of a collision. Overall, the detector has a length of 21.6 m, a diameter of 14.6 m and a total weight of 12.5 kt. Since the beams can cross up to every 25 ns the detector materials and electronics have to withstand the radiation damage caused by this high particle flux. Additionally, the electronics and computing power have to handle a huge amount of information in short time for readout and preselecting events. [71]

Furthermore, to fulfil their physics tasks the subdetector components need to provide the following information:

- A good muon identification, momentum resolution and charge of the muons, which is achieved by a combination of the tracker with the muon chambers.
- In general a good charged particle tracking and reconstruction efficiency is required.
- Also a good electromagnetic energy resolution combined with a good isolation of electro magnetic particles even at a high particle flux.
- A wide geometrical acceptance including the hadron calorimeter is needed to measure jets and missing E_T (MET).

The coordinate system of CMS is fixed to the nominal interaction point. The y-axis heads upwards to the surface while the x-axis heads to the center of the LHC ring. Hence the z-axis points along the beam axis to the Jura mountains seen from point 5. Transverse energy and momentum describes the energy and momentum components in the xy-plane. Corresponding to the z-axis the angle θ and pseudorapidity $\eta = -\ln\tan(\theta/2)$ are defined.

In figure 3.4 a layout of the CMS detector can be seen. The most distinguishable part of CMS is the superconducting magnet which is capable to establish a 4 T magnet field. At the inner part is the tracker system composed of pixel and silicon detectors to reconstruct tracks and vertices with high precision. Outside the tracker the electromagnetic calorimeter (ECAL) is placed, followed by the hadron calorimeter (HCAL). Together ECAL and HCAL can distinguish particles by their electromagnetic and hadronic characteristics. Both calorimeters cover a region of $|\eta| < 3$ and are, like the tracking detectors, still inside the magnet tube. Around the solenoid the muon chambers are placed to measure muon momentum and charge. For the muon reconstruction the information from the muon chambers and the tracker is merged. By the forward calorimeters (HF) the η coverage of CMS is extended to 5.2. An even a higher acceptance of CMS is achieved by the CASTOR detector

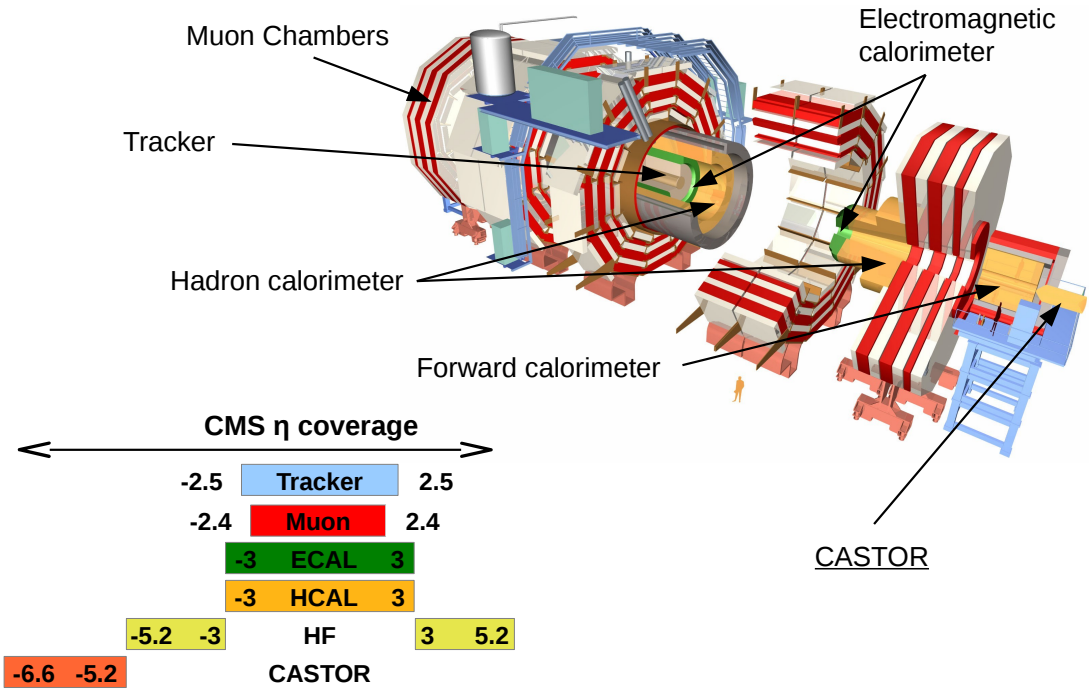


Figure 3.4: Scheme of the CMS detector. [72]

which will be described later. The forward region is additionally covered by the ZDC and FSC detectors.

The positioning of the subdetector systems follows a sophisticated design. To get a good resolution for vertices and secondary vertices the tracker should be located as close as possible to the beam axis. Any material between tracker and the actual collision point reduces the sensitivity of position and momentum by secondary and multiple scattering. Furthermore, the magnet field is most homogeneous in this area. As visualized in figure 3.5 particles like photons and electrons can be identified by a clear signal in the ECAL but no induced shower in the HCAL. The signals of electrons can be distinguished from those of photons by looking for a track pointing towards the shower position. Particles like pions and protons can preshower in the ECAL but will show clearly an energy deposit in the HCAL. Muons, however, will show essentially no measurable signal in the calorimeters but are visible in the tracker and the muon chambers. Also, cosmic muons are rejected since even when they produce a track it is very unlikely to be related to the vertex.

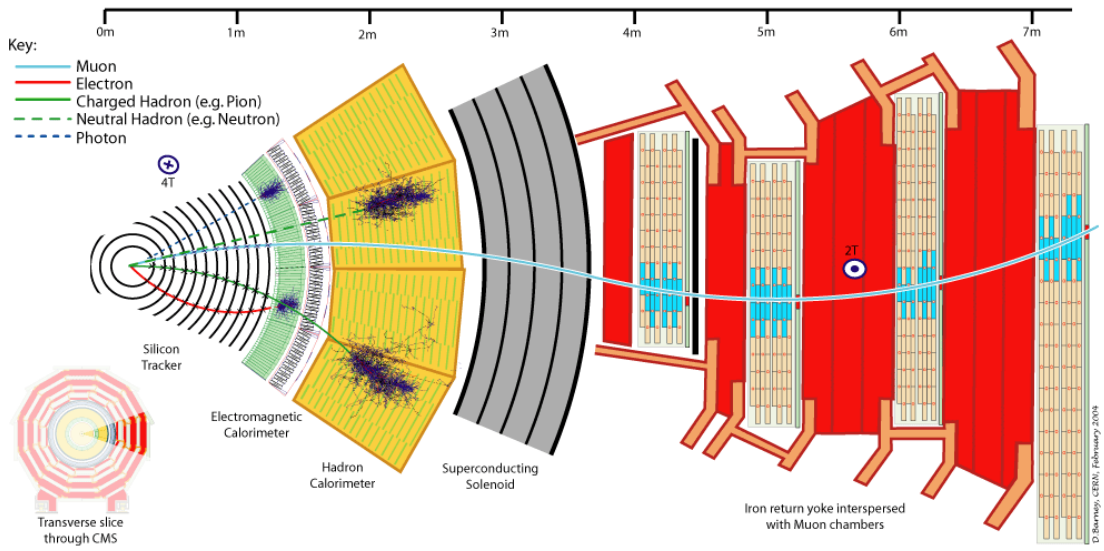


Figure 3.5: Cross section of the CMS detector showing how five different types of particles interact with the detector components. [73]

3.2.1 The central parts of detector

CMS Magnet

The heart of CMS is its superconducting magnet with a diameter of 6 m and a length of 12.5 m. In figure 3.6 one can see an illustration of the CMS solenoid and its support structure. The magnet is capable of establishing a 3.8 T field in the inner part and a bending power of 12 Tm which also covers the muon chambers. Because of this strong magnetic field high momentum particles also have a good resolution in the tracker system through bending.

The 4-layers winding cold mass of the magnet has a weight of 220 t and is made of NbTi as conductor. During nominal operation at 3.8 T the current is 19.14 kA and a 2.6 GJ energy is stored in the magnetic field. On the surface a helium refrigeration plant pumps fluid helium in the underground cavern to cool down the magnet to 4.5 K. Cooling down or heating the magnet takes several days and is therefore avoided during collision time periods.

Due to this high field in CMS the mechanical structures are put under strong mechanical stress. This can cause a slight movement of the subdetectors and their supporting structure. For example the CASTOR calorimeter moves while the magnet is ramped up and down since the whole "forward table" is sucked towards the main CMS body and also the massive radiation shielding moves. Even when it is behind a magnetic shield CASTOR and other forward detectors like TOTEM T2 see some remaining fraction of the field.

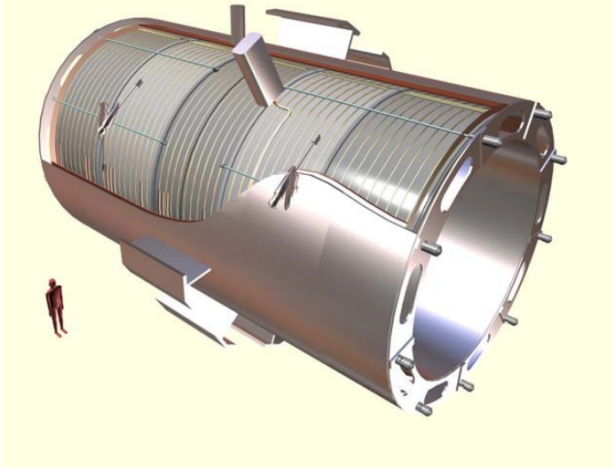


Figure 3.6: Artist view of the CMS magnet with supporting systems. [71]

Tracking system

The tracking system of CMS is designed to measure charged particle trajectories emerging from a collision at the interaction point with great precision and efficiency in the reconstruction of particles above 1 GeV/c in transverse momentum. Also, a good reconstruction of primary and secondary vertices is mandatory to identify decays of long lived particles and distinguish multiple independent collisions in a high pile-up scenario. In total it has a length of 5.8 m and a diameter of 2.5 m and covers a range in pseudorapidity of $|\eta| < 2.5$. The CMS solenoid covers the full volume of the tracker system into a homogeneous field.

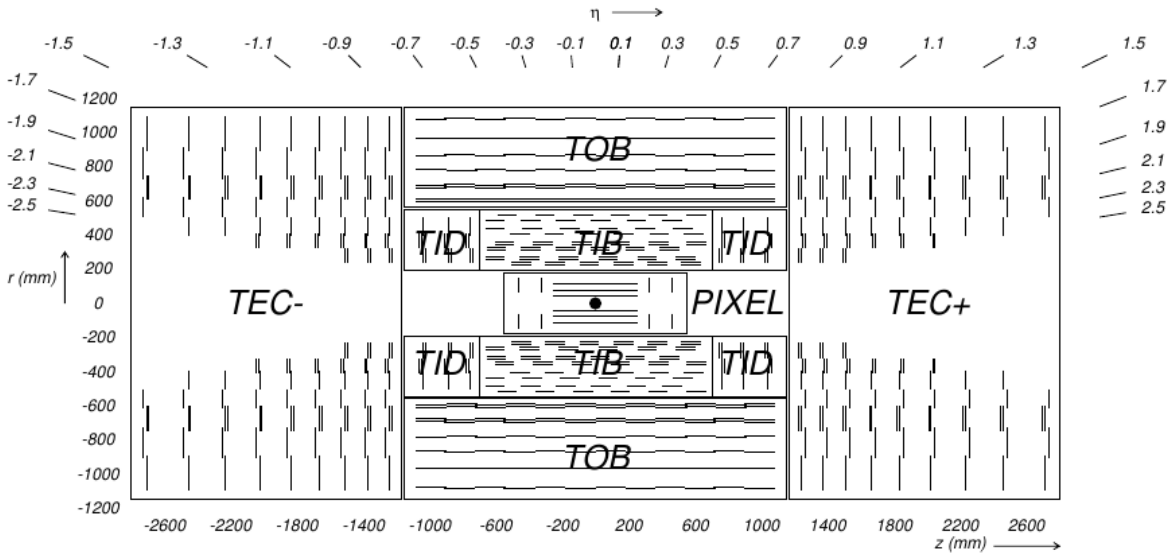


Figure 3.7: Schematic cross section through the CMS tracker. Each line represents a detector module. [71]

The design of the CMS tracking system as outlined in figure 3.7 is composed of several layers of different tracking detectors. The innermost part at radii between 4.4 cm and 10.2 cm is made of three barrel layers of pixel detectors (PIXEL). It is complemented by 2 pixel detector layers at each side of the barrel, called endcap, to cover the range up to $|\eta| = 2.5$. For a precise tracking a pixel size of $100 \times 150\,\mu\text{m}^2$ has been chosen for the detector layers. In sum the detector contains 66 million pixel which are distributed in modules over a total area of $1\,\text{m}^2$. This design provides up to 3 very precise trajectory points in the 3D space.

The radial size of the tracking system is extended up to 1.1 m by silicon strip trackers. At radii between 20 cm and 55 cm silicon micro-strip detectors (TIB/TID) with a cell size of $10\,\text{cm} \times 80\,\mu\text{m}$ are used and lead to a point resolution of $23 - 35\,\mu\text{m}$. In the outer region of the tracker ($55\,\text{cm} < r < 110\,\text{cm}$) the size of the strips (TOB/TEC) is increased to $25\,\text{cm} \times 180\,\mu\text{m}$ with a resolution between $35 - 53\,\mu\text{m}$. Since the noise by electronics is linearly dependent on the length and thickness of the strips the silicon sensors are placed in the outer regions to get a good signal over noise ratio. Therefore the sensors at the outer tracker have a thickness of $500\,\mu\text{m}$ while the sensors at the inner side are $320\,\mu\text{m}$ thick.

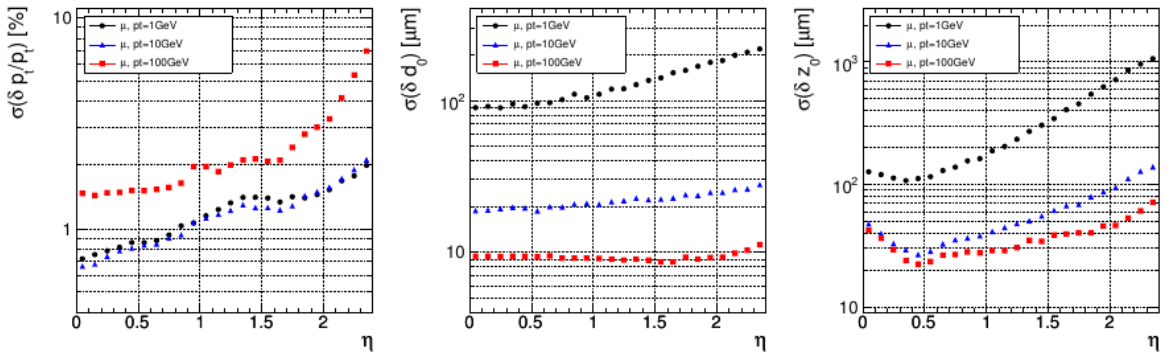


Figure 3.8: Resolution for single muons with a transverse momentum between 1 and 100 GeV. Transverse momentum resolution (left), transverse impact parameter (middle), and longitudinal impact parameter (right). [71]

The tracking design is based on the challenging environment provided by the LHC which provides a luminosity of up to $10^{34}\,\text{cm}^{-2}\text{s}^{-1}$. Therefore, the tracker volume is crossed on average by 1000 particles every 25 ns corresponding to a hit rate of 1 MHz per square meter at a radius of 4 cm from the interaction point. Hence, dense and fast read-out electronics is needed which requires additional cooling. On the other hand one wants to minimize the amount of material in the tracker since secondary interactions of the particles in the tracker will reduce momentum and vertex resolution. Due to the high particle flux, the tracker needs to be very radiation hard.

The performance of the tracker is shown in figure 3.8 and figure 3.9. In figure 3.8 the resolution of a single muon in transverse momentum and position in the longitudinal and

transverse direction of the beam can be seen. In the barrel region $|\eta| < 1.5$ the momentum resolution up to 100 GeV is better than 2 % while in the endcap region the resolution declines to 5 %. Because of the bending the vertex resolution is worse for lower momentum particles but still in a region of only several μm. Especially in z a good resolution is needed to separate vertices at high pile-up conditions.

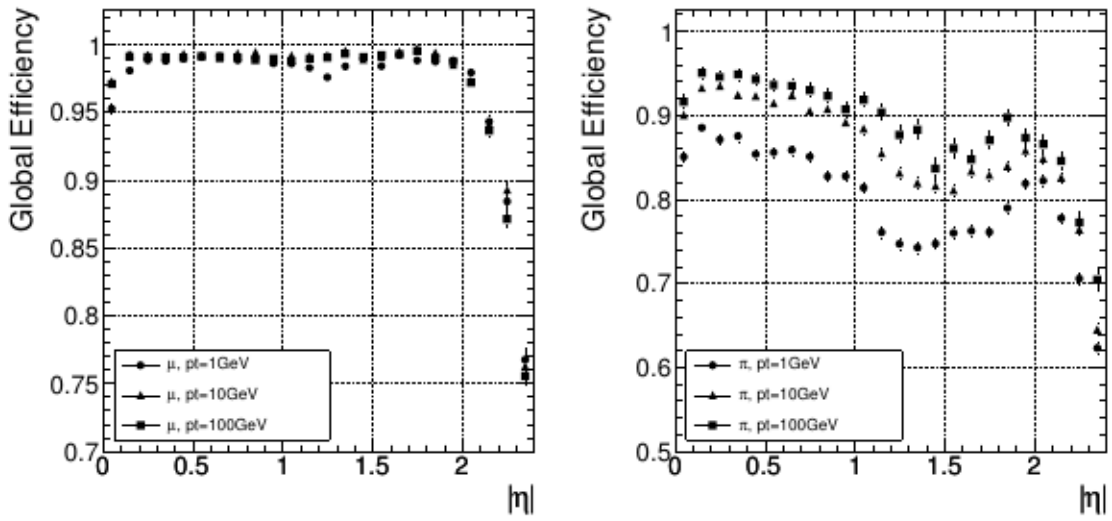


Figure 3.9: Track reconstruction efficiency for muons (left) and pions (right) depending on pseudorapidity for different transverse momenta. [71]

The reconstruction efficiency of the whole tracker system is shown in figure 3.9. For muons the overall efficiency is very high at 99 % except for the forward region $|\eta| > 2$ because of less coverage by the forward pixel detectors. The reduced efficiency of pions and hadrons is caused by the increased cross section of particle interactions with the tracker material. After particles pass through the tracker they will hit the calorimeters, first the electromagnetic calorimeter and then the hadron calorimeter.

The ECAL

The electromagnetic calorimeter in CMS is a hermetic homogeneous calorimeter made of 61200 lead tungstate ($PbWO_4$) crystals in the central barrel and closed by 7324 crystals on each side of the two endcaps. The requirements of a fast and radiation hard detector with a high granularity leads to the use of these high density crystals. The material is very appropriate for LHC collisions due to its short radiation length (8.9 mm) and small Molière radius (2.2 cm) which results in a compact and high resolutive calorimeter.

A picture of one scintillating crystal of the ECAL endcap is shown in figure 3.10. With 80 % of the light emitted in 25 ns the scintillation decay time is of the same order of magnitude as the LHC bunch crossing time. The maximum of the emitted blue-green scintillation light

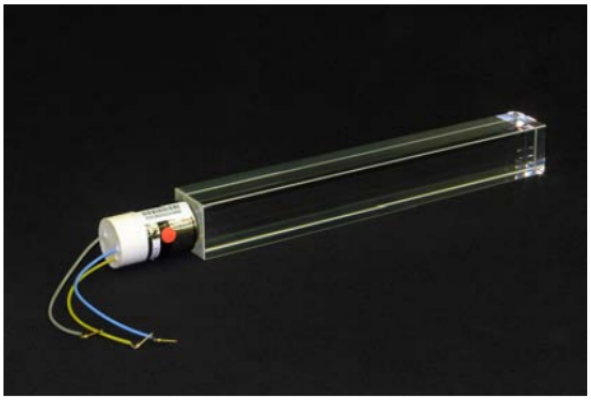


Figure 3.10: An ECAL endcap crystal with a PMT attached and a total length of 230 mm. [71]

is around the wavelengths of 420-430 nm while 4.5 photoelectrons are collected per MeV at a temperature of 18 °C. This temperature needs to be maintained at a high precision of ±0.05 °C since the number of emitted photons per MeV and also the amplification of the PMTs are very temperature dependent.

In the ECAL barrel (EB) a total of 61200 crystals are installed with a 360-fold granularity in ϕ and (2×85)-fold in η. They are mounted in a quasi-projective geometry with respect to the vertex. The length of the crystal is 230 mm ($25.8\,X_0$) with a front cross section of $22 \times 22\,mm^2$. In total the ECAL barrel has a volume of $8.14\,m^3$, weight 67.4 t and covers a range in pseudorapidity of $|\eta| < 1.479$. [71]

The ECAL endcaps cover a pseudorapidity range $1.479 < |\eta| < 3.0$. They are divided into two halves with 3662 crystals each. The crystal front area is $28.62 \times 28.62\,mm^2$ and they have a length of 220 mm which corresponds to $24.7\,X_0$. Together the crystals of the endcap have a volume of $2.90\,m^3$ and a weight of 24.0 t. Additionally, a preshower detector in the range of $1.653 < |\eta| < 2.6$ is installed in front of the endcaps to reject signals of neutral pions by their double photon ($\pi^0 \to \gamma\gamma$) decay. It also improves the determination between electrons and minimum ionizing particles.

In figure 3.11 the energy resolution of the ECAL detector depending on the initial electron is shown. The measurement was performed during beam tests before the final installation. The resolution was measured by shooting an electron beam with a well-known energy on the central crystal of an array of 3×3 crystals. This is repeated for several different electron beam energies. By fitting the measured points in figure 3.11 to the function

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \tag{3.1}$$

the given stochastic (S), noise (N), and constant (C) terms are determined. As shown the energy resolution is between 0.9 % at low energies around 20 GeV and even less than 0.4 % at higher energies above 250 GeV.

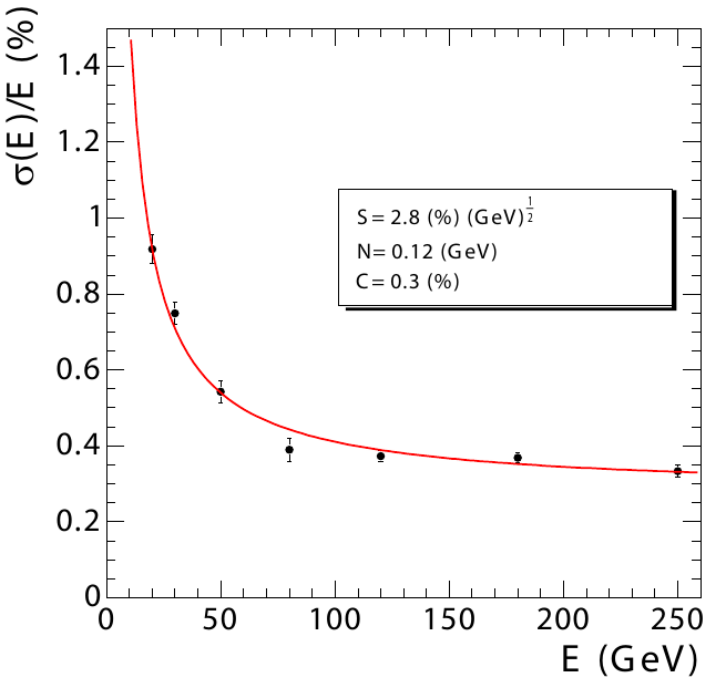


Figure 3.11: ECAL energy resolution as a function of electron energy as measured from a beam test. The stochastic (S), noise (N), and constant (C) terms of the fit are given. [71]

HCAL

Particles which pass through the ECAL will be absorbed by the hadron calorimeter (HCAL). The purpose of the HCAL is to measure jets and reconstruct neutrinos or other exotic particles via missing transverse energy. Hence it is important that the whole energy of the other particles is reconstructed. Since in the region for the HCAL barrel (HB) the area is limited by the outer radius of the ECAL ($R = 1.77\,\mathrm{m}$) and the inner radius of the magnet coil ($R = 2.95\,\mathrm{m}$) the amount of material is restricted. Therefore, as shown in figure 3.12, outside of the solenoid an outer hadron calorimeter (HO) is placed additionally to catch the remaining shower particles.

The HCAL barrel (HB) covers a range of $|\eta| < 1.3$ and is divided into two halves along the beam axis. In ϕ the barrel is divided in 36 wedges which themselves are divided again into for azimuthal (ϕ) sectors. Designed as sampling calorimeters the wedges consist of alternating layers of brass plates as absorber and scintillator plates. The layers are aligned in parallel to the beam axis while the scintillator is segmented into 16 sectors in pseudorapidity. This results in a granularity of $(\Delta\eta, \Delta\phi) = (0.087, 0.087)$. Starting with a 40 mm thick steel plate at the front the absorber consists furthermore of 8 brass plates of 50.5 mm and 6 brass plates of 56.5 mm thickness. The outermost layer consist again of a 75 mm thick steel plate as absorber for structural strength. Since the plates aligned along

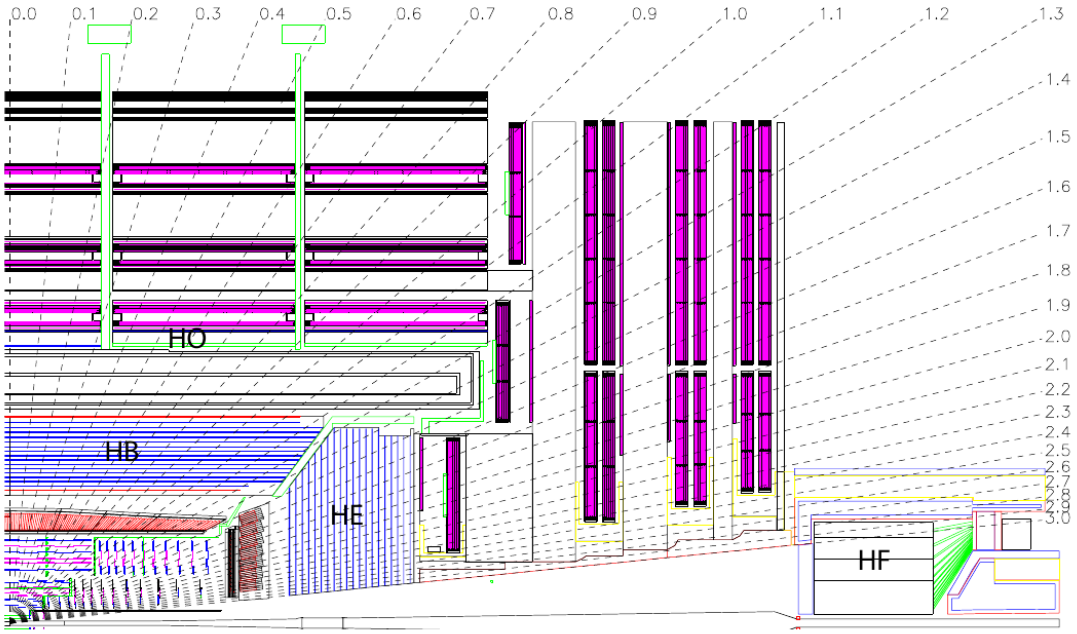


Figure 3.12: Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters. The black dashed lines give the angle in η. [71]

the beam direction the total absorber thickness at $\eta = 0$ with 5.82 interaction lengths (λ_I) is smaller than at the border ($|\eta| = 1.3$) with $10.6\,\lambda_\mathrm{I}$. The material of the ECAL adds additionally $1.1\,\lambda_\mathrm{I}$.

As mentioned, to contain the full hadron shower in the central region ($|\eta| < 1.3$) the HCAL is extended by the outer calorimeter (HO). Thereby the outer hadron calorimeter takes advantage of the solenoid coil as additional absorber material. At 90° polar angle the overall depth of the hadron calorimeter system is at least $11.8\,\lambda_\mathrm{I}$ and rising with $1/\sin\theta$. The segmentation of the outer hadron calorimeter roughly maps the hadron barrel calorimeter and has the same granularity in η and ϕ. It was shown by simulations that the energy deposit in the calorimeter system can be increased with the HO up to 4.3 % of the impact energy for 300 GeV pions.

The azimuthal angles closer to the beam pipe ($1.3 < |\eta| < 3.0$) are covered by the hadron endcaps (HE). They use C26000 cartridge brass plates as absorber to maximize the number of interaction lengths and to be non magnetic since the endcaps are mounted at the end of the magnet tube. Due to the overall design of CMS the HE is mounted to the endcap of the muon system and subsequently the ECAL endcap to the HE. With brass plates of 74 mm thickness and 9 mm gaps which contain the scintillator, the endcaps have a total depth of $10\,\lambda_\mathrm{I}$ including the material of the ECAL.

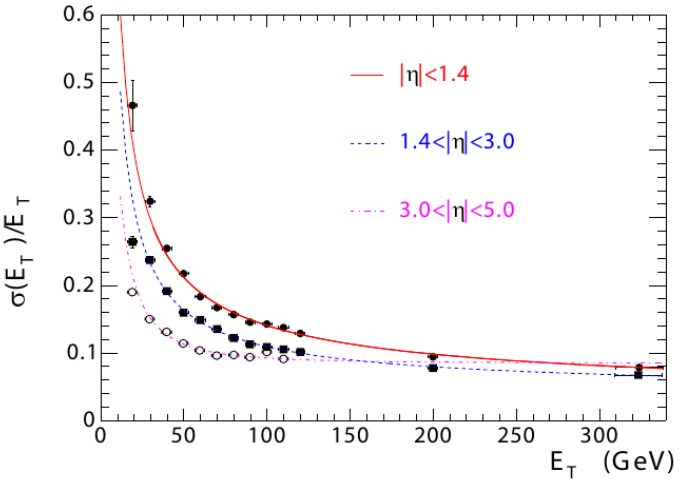


Figure 3.13: The jet transverse-energy resolution as a function of the jet transverse energy for barrel jets ($|\eta| < 1.4$), endcap jets ($1.4 < |\eta| < 3.0$), and very forward jets ($3.0 < |\eta| < 5.0$). The jets are reconstructed with an iterative cone algorithm (cone radius = 0.5). [71]

In figure 3.13 the transverse energy resolution for jets of the whole calorimeter system is illustrated. The resolution is shown separately for jets in the barrel ($|\eta| < 1.4$) and the endcap region ($1.4 < |\eta| < 3.0$). To reconstruct the jets an iterative cone algorithm with $R = 0.5$ was used. At lower energies the resolution is around 30 % to 40 % but improves to 10 % at higher energies.

Muon Chambers

For CMS to investigate QCD physics and fulfil its physics program the muon detection system is an essential part. Muons are ideal for analyzing different processes since apart from being easy to detect, in lepton decays for example compared to electrons, they are less affected by radiation loss in the tracker system.

Their detection system must be capable of identifying muons, measuring their momentum and triggering on muons. By the high magnetic field and its direction turn outside of the solenoid a good muon resolution and trigger functionality is provided. Together with the calorimeter the magnet also serves as a hadron absorber to identify muons. The material budget a muon has to traverse at different stations can be seen in figure 3.14.

The muon system in CMS uses 3 different types of gaseous particle detectors. It consists of a cylindrical barrel region and two planar endcap regions with about 25000 m^2 of detection plane. In the barrel region with a low rate of muons and background by neutrons drift tube chambers (DT) with rectangular drift cells are used. They cover the region $|\eta| < 1.2$ and are separated into 4 stations. By the first 3 each containing 8 chambers the muon

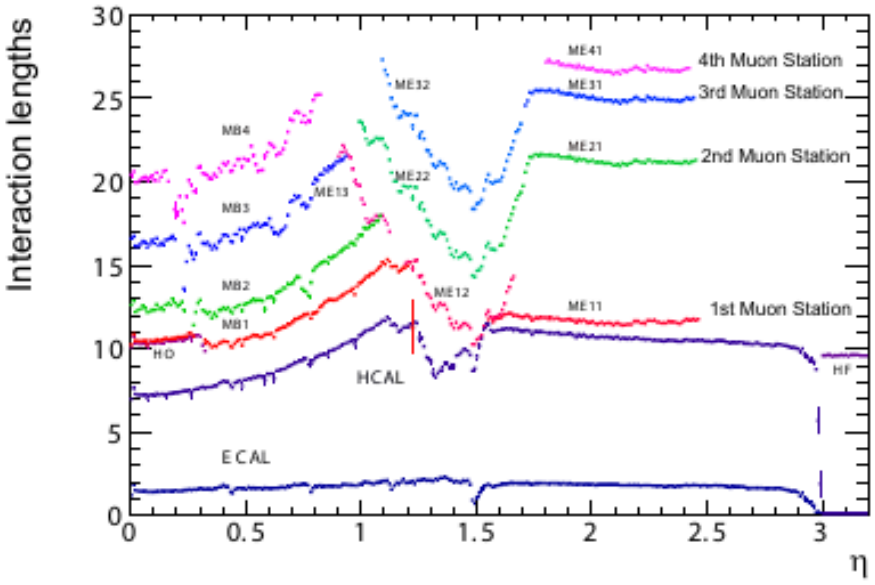


Figure 3.14: Material thickness in interaction lengths at various depths, as a function of pseudorapidity. [71]

momentum in the whole 3 dimensional space is measured. The last station also containing 8 chambers only measures the muon momentum according to the r-ϕ plane.

The muon endcap, where the muon flux and the background is higher, operates with cathode strip chambers (CSC) and covers the pseudorapidity region of $0.9 < |\eta| < 2.4$. The four CSC stations in each endcap are installed perpendicular to the beam axis and deliver in the r-ϕ plane a precise measurement of the muon momentum. For the muon trigger it is very important that the different sub systems DT and CSC trigger on the muon p_{T} efficiently and very fast. Therefore, a complementary dedicated trigger system made of resistive plate chambers (RPC) was added. They provide over the range $|\eta| < 1.6|$ a fast and independent trigger with a good time transverse momentum resolution. Their position resolution on the other hand is much worse than the one of the DT and CSC detectors.

Due to all its components the muon system covers the range $|\eta| < 2.4$ without any gaps in acceptance. In figure 3.15 the resolution of the transverse momentum (p_{T}) is shown once again as performance criteria of the detector system, on the left for the region $|\eta| < 0.8$ and on the right for $1.2 < |\eta| < 2.4$. In each plot the resolution is determined for the three cases of using in the reconstruction exclusively the inner tracker, exclusively the muon system or taking advantage of both systems together. For small η values the resolution in all cases is better than 10 %. At higher energies only the combination of tracking and muon detectors is able to reach this. In the higher pseudorapidity regions the muon chambers alone have a resolution from 20 % to 40 % up to 1 GeV in transverse momentum. By including the tracker also in this region the resolution improves by an order of a magnitude.

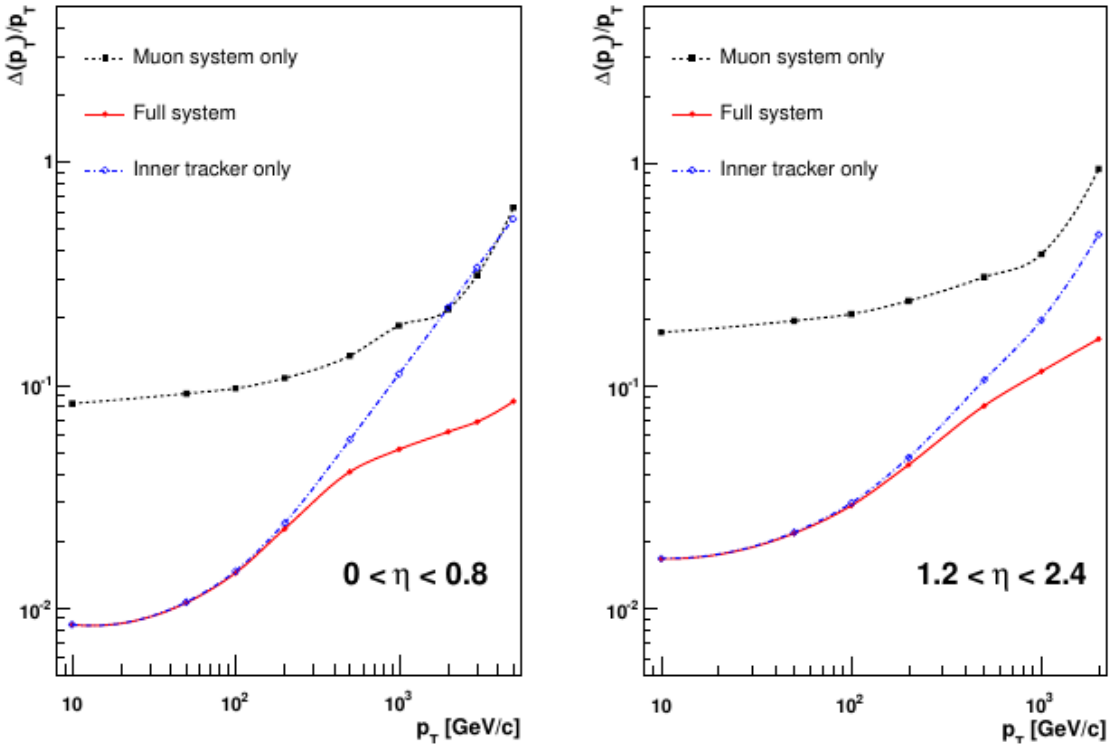


Figure 3.15: The transverse momentum resolution of muons as a function of the transverse momentum (p_{T}) using the muon system only, the inner tracking only, or both. On the left for $|\eta| < 0.8$ and on the right for $1.2 < |\eta| < 2.4$. [71]

3.2.2 The forward parts of the detector

Forward Calorimeters

In the very forward region after the muon endcap the hadron forward calorimeter (HF) is installed. With HF the acceptance range of the CMS calorimetric is extended up to $|\eta| < 5.2$. In this pseudorapidity region the flux of produced particles will be very high. Hence HF was designed to survive this high radiated environment by using a steel absorber with embedded quartz fibres as active medium. The fibres are made of a fused silica quartz core with $600 \pm 10\,\mu$m as diameter and polymer hard cladding. Inside the fibres the light is produced by charged particles via the Cherenkov mechanism and collected through air-core light guides by the photomultipliers at the end. In figure 3.16 one can see a cross section of HF and its components housed in the shielding.

The steel absorber is composed of 5 mm thick groove plates in which the fibres are inserted. Thereby half of the fibres run over the full depth of the absorber ($165\,\mathrm{cm} \approx 10\,\lambda_{\mathrm{I}}$) while the other half starts at a depth of 22 cm away from the detector front. The read out of the two different sort of fibres referred to simply as long fibres and short fibres is done separately.

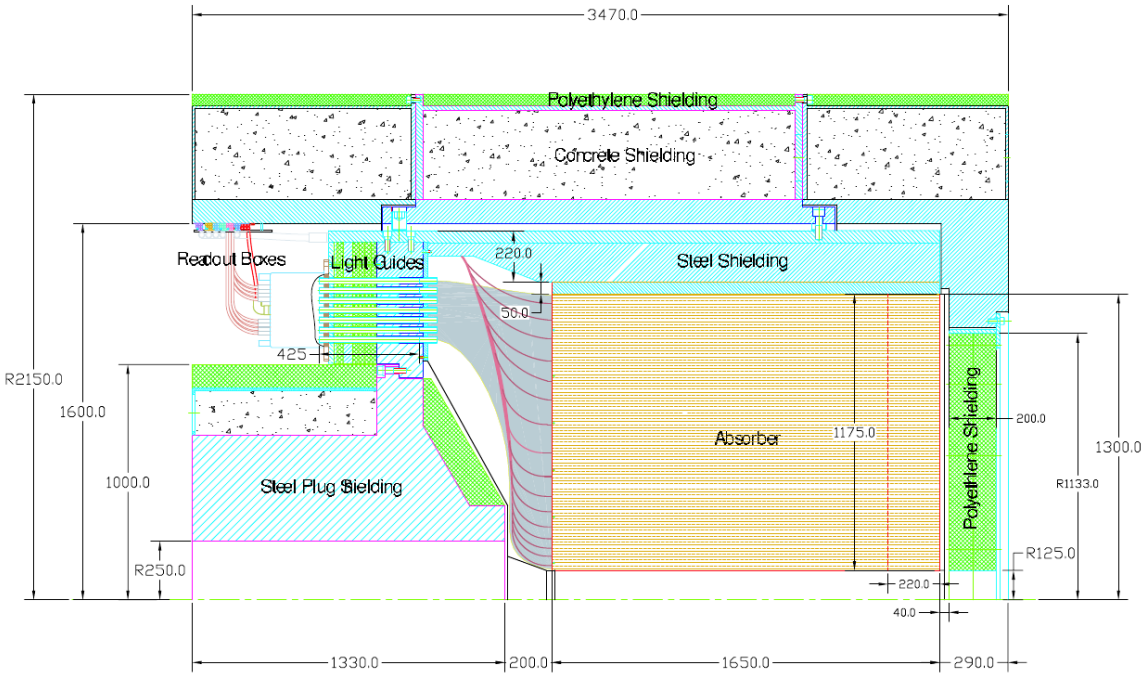


Figure 3.16: The cross sectional view of the HF calorimeter. The absorber in the beam direction measures 1650 mm. Bundled fibres (shaded area) are routed from the back of the calorimeter to air-core light guides which penetrate through a steel-lead-polyethylene shielding matrix. Light is detected by PMTs housed in the read-out boxes. Stainless steel radioactive source tubes (red lines) are installed for each tower and are accessible from outside the detector for source calibration. The interaction point is at 11.2 m from the front of the calorimeter to the right. All dimensions are in mm. [71]

By this construction the calorimeter is functionally divided into two longitudinal segments to separate showers caused by hadrons from electromagnetic showers induced by photons and electrons.

Essentially the calorimeter is a cylindrical steel structure surrounding the beam pipe with an outer radius of 130 cm and an inner radius of 12.5 cm. At both sides of CMS the front face of the detector is 11.2 m away from the interaction point. The fibres are arranged parallel to the beam axis and bundled to form towers of the size $\Delta\eta \times \Delta\phi = 0.175 \times 0.175$ except for the edges of the calorimeter where $\Delta\eta$ is 0.111 at $|\eta| = 2.9$ and 0.3 at $|\eta| = 5.2$. The calorimeter is shielded with layers of 40 cm thick steel, 40 cm of concrete and 5 cm polyethylene to protect its electronics and PMTs from radiation and at this position not anymore uniform magnetic field flux from the solenoid.

Referring back to figure 3.13 one can see the jet transverse energy resolution achieved by HF indicated by $3.0 < |\eta| < 5.0$. The forward calorimeter performs equally compared to the HCAL barrel and endcap at higher energies while at energies below 100 GeV the jet

E_T resolution gets better. This results in a transverse energy resolution of jets which is over a far region below 20 % and even better than 10 % for energy above 100 GeV.

TOTEM

The design purpose of the TOTEM experiment is to measure the total proton-proton cross section. Therefore TOTEM needs an acceptance up to very low polar angles (θ). TOTEM's overall coverage in pseudorapidity is $3.1 < |\eta| < 6.5$ and is achieved by two telescopes for charged particles on each side of the interaction point. Additionally, Roman Pots (RP) detectors are installed about 147 m and 220 m away from the interaction point in special movable beam pipe insertions. Their purpose is to detect the leading protons only a few mm away from the transversal beam center. [74]

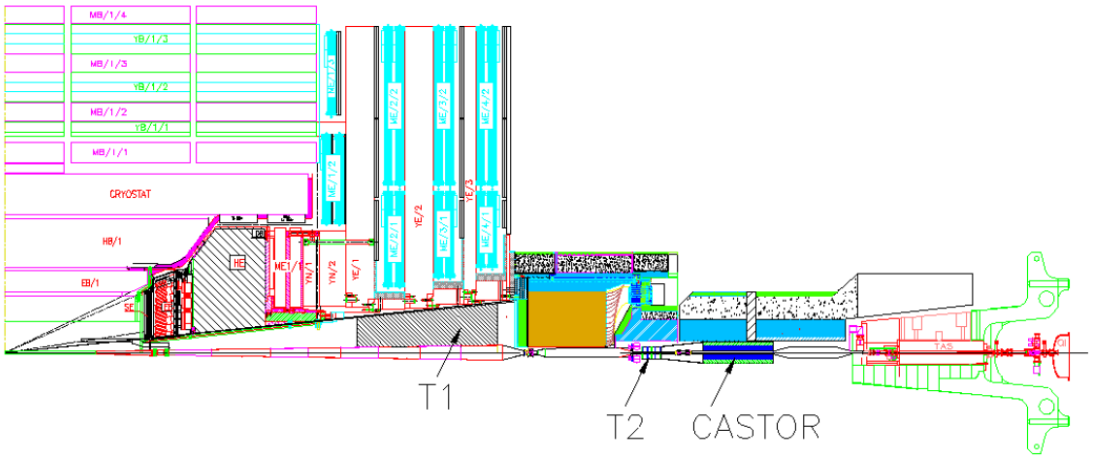


Figure 3.17: The TOTEM forward trackers T1 and T2 embedded in the CMS detector together with the planned CMS forward calorimeter CASTOR. [74]

To measure the total cross section the detectors have to fulfil several requirements:

- provide an inclusive trigger for minimum bias and diffractive events with minimal losses.
- the ability to reconstruct the primary vertex in order to reduce background by connecting tracks to the vertex.
- a left-right symmetry with respect to the interaction point to get a better control of the systematics.

These requirements influenced the choice of the detector technology and they are different from normal tracking detectors which achieve a guaranteed detection and reconstruction of each particle per event.

The telescope closest to the interaction point (T1) is at $z = 9\,\text{m}$ and placed between the beam pipe and the muon endcap directly in front of HF as shown in figure 3.17 and 4.1. On each side T1 consists of five planes of Cathode Strip Chambers (CSC) equally spaced in z and divided into two halves. Each detector plane is composed of six CSC wired chambers covering 60° in ϕ. These gas detectors are slow but with a gas gap of 10 mm still capable of handling the hit rate of TOTEM. T1 can detect charged particles in the range of $3.1 < |\eta| < 4.1$ and expect an inelastic interaction rate of 1 kHz.

Behind HF at $z = 13.5\,\text{m}$ on both sides of CMS the TOTEM telescope T2 is installed. Due to the fact that it is directly in front of the CASTOR detector (see figure 3.17 and 4.1) and has a similar acceptance in pseudorapidity ($5.3 < |\eta| < 6.5$), T2 data was of particular interest for this thesis to align CASTOR as described in section 4.4. On each side T2 is made of 20 consecutive semicircular half plates consisting of Gaseous Electron Multiplier (GEM) where on each detector layer 2 of these GEM half plates are slid together.

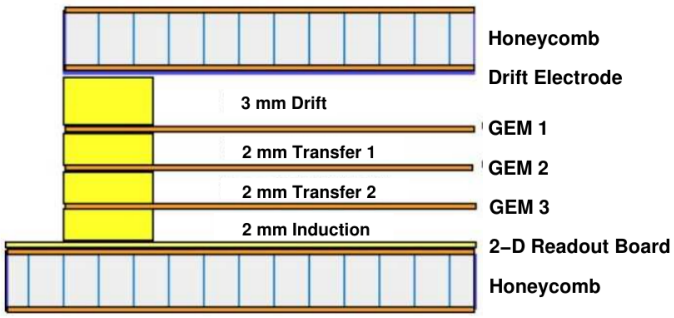


Figure 3.18: A side view of the T2 GEM detector structure with three Gaseous Electron Multiplier (GEM) amplification stages made of Cu-clad polyimide foils supported by honeycomb plates. A 3 mm drift space is followed by two 2 mm deep charge transfer regions and a 2 mm charge induction space. The large signal charges are collected in two dimensions, by a read-out board underneath of the induction layer. [74]

In figure 3.18 the structure of the GEM detectors is visualized. Each GEM detector is made of three foils with 3 mm and 2 mm distance from each other. The foils consist of 50 μm thick polyimide foil with a 5 mm copper cladding on both sides. Holes placed in the middle of the foils have a diameter of 65 -80 μm. At each foil the average amplification factor is 20. For the read out of each GEM detector 512 strips for radial and 1560 pads for azimuthal coordinates are used. This results in a grid of $2 \times 2\,\text{mm}^2$ at the inner side (closer to the beam) and $7 \times 7\,\text{mm}^2$ at the outer side of the detector. The residual resolution of a track hit in the two GEM detectors placed on one plate together is around 235 μm.

To detect the very forward protons the Roman Pots (RP) are used. By placing the detector inside a secondary vacuum vessel (called a pot) which is moved into the vacuum of the beam pipe the detector is still physically separated from the primary vacuum. Hence an

out gassing of the detector materials is prevented. Two Roman Pot stations are installed on each side of IP5. The closer one is placed 149.6 m away from the interaction point, the other one is 217.3 m away. Due to a dipole magnet between the two stations the caused dispersion difference helps to reconstruct the proton momentum.

Each station consists of three pots where two approach the beam vertically and one from the top but does not cross the beam. The pots consist of a stack of 10 layers of silicon strip detectors. Half of them are oriented +45° with respect to the horizon and the other half orthogonal to the first ones. They consist of 512 strips with a pitch of 66 μm and grouped into 128 channels. This results in a residual resolution of 19 μm for one Roman Pot.

With TOTEM most of the forward detectors at IP5 except for the CASTOR calorimeter are described. Since the results in this thesis depend mostly on data recorded by CASTOR the detector is described in the next chapter.

3.2.3 Data acquisition

At LHC two proton bunches can cross each other at the interaction points with a minimal distance of 25 ns. This results in a bunch crossing frequency of 40 MHz. Depending on the luminosity of the bunches the multiplicity of proton-proton collisions per bunch crossing can lead to a high occupancy inside the CMS detector. Since not all data is of physical interest and by the limitation of the CMS read out system a dedicated trigger system is needed to reduce the average amount of data per time.

In general is the trigger system is divided into two components. The first one is the hardware based Level-1 (L1) trigger which uses direct information from the tracking detector, calorimeters and muon chambers to reduce the rate to 100 kHz. The High-Level-Trigger (HLT) is the second part, which is software based and runs on generic CPU cores, using faster versions of the offline reconstruction to filter the events. Therefore it can include a complex Algorithm to search for particular event shapes like the Higgs decay. Together the two stages reduce the rate by a factor of 10^6 so that the data can be read out and stored with 100 Hz. In this thesis also L1 triggers based on the CASTOR calorimeter and additionally combined with triggers decisions from the TOTEM experiment are of special interest. Hence a more detailed description can be found in section 4.3.

An architectural layout of the trigger system included in the CMS Data Acquisition (DAQ) is shown in figure 3.19. The CMS DAQ is designed to collect the detector signals considering the trigger decisions and reconstructing the events. The data incoming at 40 MHz is stored temporally into the Detector Front-Ends before it is processed by the L1 trigger. If the synchronously arrived L1 decision by the Timing, Trigger and Control system (TTC) was positive, the data is pushed from the Front-End buffers into the DAQ system by the Front-End Drivers (FEDs). Hence the DAQ system must be able to handle the output rate of the L1 which corresponds to a data flow of ≈ 100 GB/s. Next, the event building network assembles the event fragments belonging to the same L1 and transmits the data to the

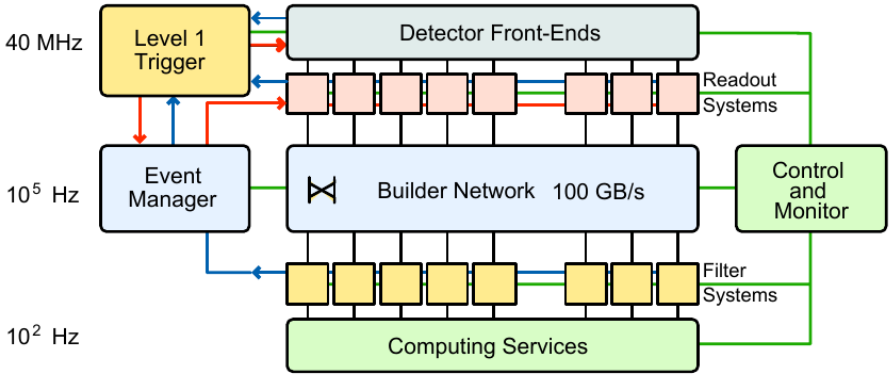


Figure 3.19: Architecture of the CMS DAQ system. [71]

HLT filter farm (Computing Services) for processing. Since the data arrives with a rate of several kHz this computing farm needs to be very powerful. On a 3 GHz Xeon CPU core the HLT algorithms will taken a average time of around 50 ms. Hence a DAQ system which runs at 50 kHz would need at least 2500 equivalent CPU cores. During collision time CMS is able by changing the trigger thresholds or prescales to optimize the data output for the DAQ and trigger structure.

4 The CASTOR forward calorimeter

The name of the CASTOR forward detector is an acronym of "Centauro And Strange Object Research". The detector is a sampling calorimeter made of tungsten and quartz plates and measures the Cherenkov radiation emitted by the particles passing through the detector. CASTOR is designed to measure the produced particles in heavy-ion and proton-proton collisions at very small angles.

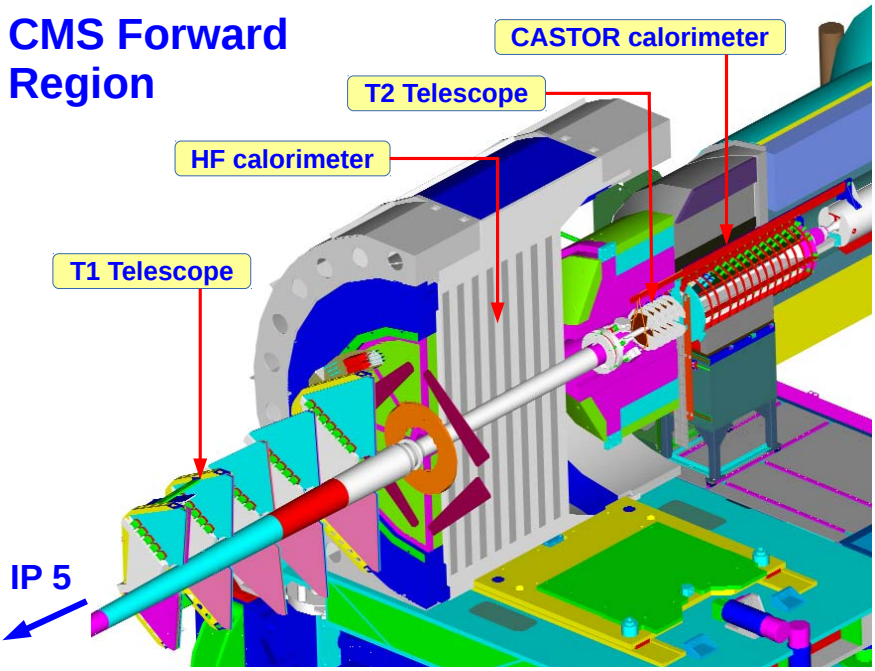


Figure 4.1: Here the CASTOR detector embedded in the forward environment of CMS is shown. The arrow on the left shows the position of the interaction point (IP) where the two beams are colliding. T1 and T2 tracking telescopes belong to the TOTEM collaboration. Between them the hadron forward (HF) calorimeter from CMS is positioned.

The calorimeter is positioned 14.38 m away from the nominal interaction point on the negative z side of the CMS detector covering the range in η from -6.6 to -5.2. As shown in figure 4.1 CASTOR is directly behind the TOTEM tracker T2. With a combined data taking of CMS together with TOTEM it is possible to combine reconstructed T2 tracks

with information about the energy from CASTOR. The hadronic forward calorimeter is in front of the T2 tracker. Hence HF is also a common source of background radiation escaping from its back for CASTOR and T2. Also the beam pipe has a conical form with an angle pointing exactly to the nominal interaction point. This shadows particles at a certain $\eta \approx -5.5$ completely but is better compared to a flat design where the shadowed area in η is much larger due to secondary interactions in the material of the beam pipe.

Because of the forward acceptance ($-6.6 < \eta < -5.2$) of the CASTOR calorimeter it is able to measure QCD physics at small-x. Furthermore, CASTOR is very sensitive to underlying event processes and MPI. Therefore the CASTOR detector is ideal for measuring variables such as energy or jets to test generators concerning the related physics. Together with the other detectors of CMS the measurable pseudorapidity acceptance is expanded to $-6.6 < \eta < -5$ including CASTOR. No other experiment is able to investigate physics over such a huge range.

4.1 Design and hardware

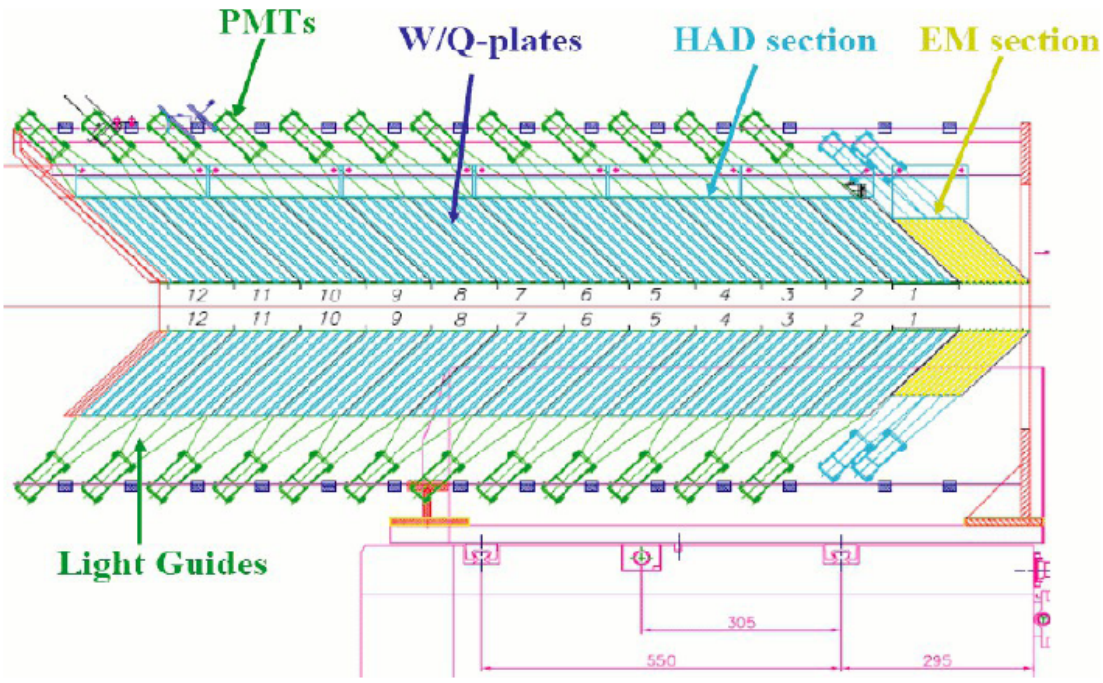


Figure 4.2: Scheme of the CASTOR calorimeter made of tungsten (W) and silica quartz (Q) plates shown with the collision point heading to the right side, the electro magnetic (EM) section is highlighted in yellow, the hadronic (HAD) section in cyan [75].

As already mentioned, CASTOR is a Cherenkov calorimeter built of alternating layers

of tungsten as absorber material with a density of $18.5\,\mathrm{g/cm^2}$ and silica quartz as active medium. These materials are favourable since they are radiation hard and non-magnetic which allows a compact design. The detector itself is split in two parts along the beam axis. The front part, where the tungsten plates have a thickness of 5 mm in beam direction and the quartz plates of 2 mm, is the electromagnetic (EM) part with a total depth of $20\,\mathrm{X_0}$. A particle which induces an electromagnetic shower (from the right side in figure 4.2) will deposit most of its energy in this section. Behind the EM section of CASTOR the hadronic (HAD) section follows, where the plates have twice the thickness in mm as in the electromagnetic part. With a deepness of $9.24\,\lambda_\mathrm{I}$ for the hadronic part the CASTOR calorimeter has, together with the EM section, a total depth of $10\,\lambda_\mathrm{I}$.

In ϕ the calorimeter is made of 8 symmetrically shaped octants each of which is divided into two sectors. Along the beam axis the electromagnetic part is also divided into two modules with $10\,\mathrm{X_0}$ each and the hadronic part into 12 modules with a depth of $0.77\,\lambda_\mathrm{I}$ for one module. Therefore CASTOR is separated into 224 readout units grouped in 16 sectors and 14 longitudinal modules.

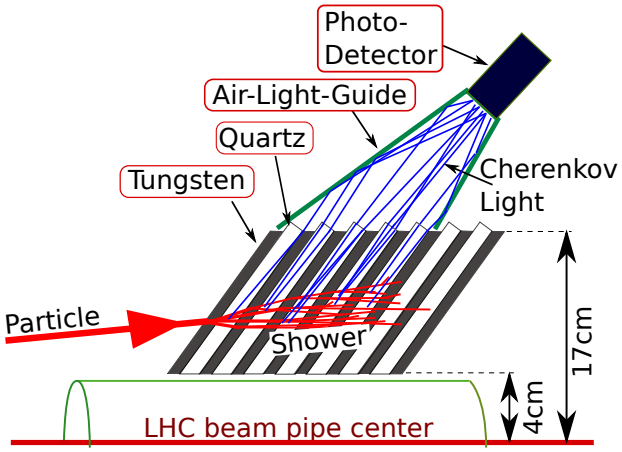


Figure 4.3: Scheme of CASTOR including the induced shower particles and the produced Cherenkov photons in the silica quartz plates.

To maximize the collected amount of Cherenkov photons the tungsten and quartz plates are tilted by 45° with respect to the beam axis (see figure 4.3). Since the refraction index of $n \approx 1.45$ for the silica quartz plates and assuming a velocity of $\beta \approx 1$ the emission angle θ of the Cherenkov light given by

$$\cos\theta = \frac{1}{n\beta} \tag{4.1}$$

is around 46°. Since the particles produced in a shower have different angles not parallel to the beam the produced Cherenkov light is also tilted to the plates. By total internal reflection at the border of the plates the light is guided outside.

The silica quartz plates of each of the 224 readout units are then connected to a PMT via an air-core light guide. The light guides collect the Cherenkov light produced in the quartz

plates of the corresponding readout unit and direct it to the photo multipliers. They are made of thin stainless steel sheets where the inside is filled with air and the surface is covered with a highly reflective foil. The photo multipliers are from Hamamatsu with fine meshed grid dynodes.

Detector calibration

All 224 channels have different efficiencies because of different high voltage settings and different light guides. The PMTs experience also various effects of the residual strength of the magnetic field by the CMS magnet. To compensate these effects an intercalibration of the channels is performed. Therefore events are recorded with a dedicated CASTOR trigger

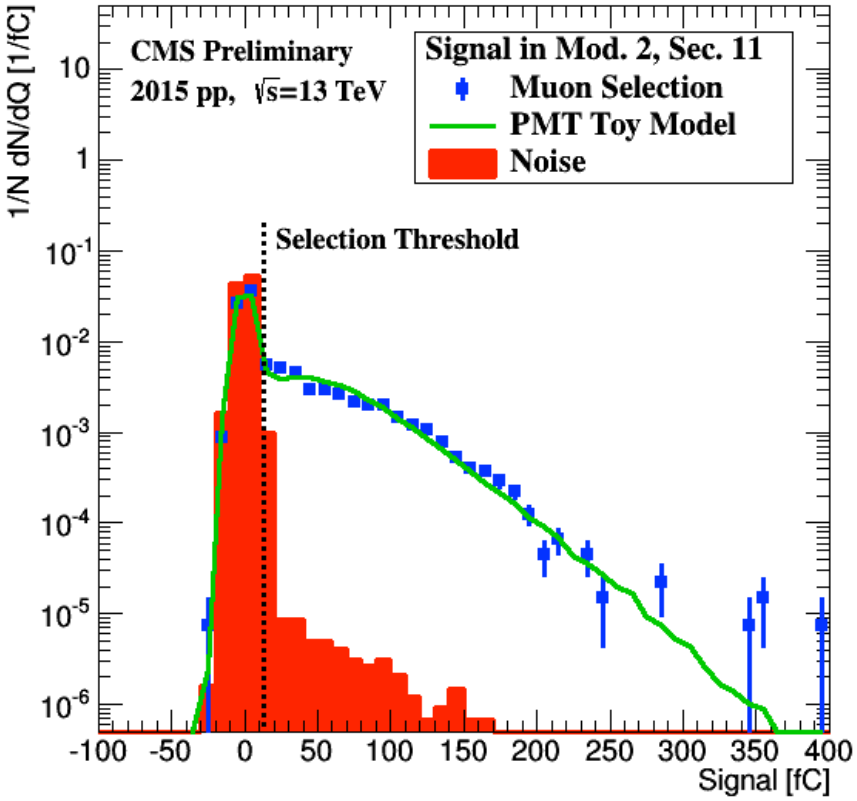


Figure 4.4: Typical energy spectrum of a CASTOR channel after an offline isolated muon event selection. In red the noise distribution of the corresponding channel is measured with non-colliding bunch data. The green line describes a mesh-type PMT with an average number of photoelectrons nPE of 0.5 and is obtained by a toy model of the fine-mesh PMT. The selection threshold is indicated by the vertical dashed line. [76]

to record muons from the LHC beam halo. These halo-muons are parallel to the beam axis and penetrate the detector over the entire length. The data are taken during the interfill periods when the beam is filled up with proton bunches and no collisions occur. With the recorded muons the response of the channels is equalized. It is assumed that muons act as minimum ionizing particles and deposit the same energy in each channel of the calorimeter. For the selection of muons in CASTOR it is required that in one sector at least one channel in the front, the middle and the back of the detector are above noise level, while the rest of the detector is silent. In figure 4.4 the muon response for one detector channel is shown. The PMT is described here by a toy model with an average production of 0.5 photoelectrons per muon. During different operation periods the condition of the detector such as high voltage, new PMTs and magnet field are changing. Hence the intercalibration with muons has to be repeated regularly. In addition, LED pulser data is taken to correct the changing conditions in combination with the muon data.

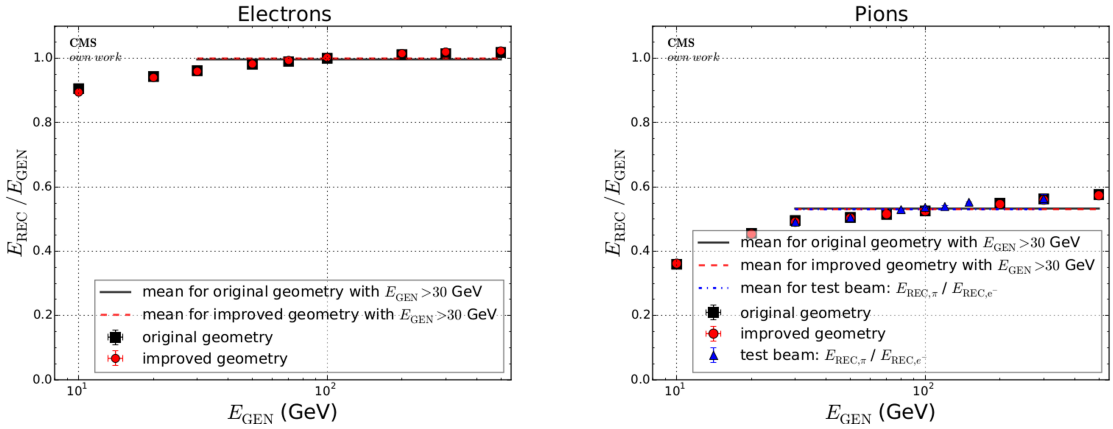


Figure 4.5: The response of reconstructed to generated energy E_{REC}/E_{GEN} of electrons (left) and pions (right) is shown for the CASTOR calorimeter. Test-beam data is compared to different simulations of the detector. For this thesis simulations with the improved geometry are used. [77]

For the calibration of the response also the noncompensating nature of the CASTOR detector needs to be taken into account. For this purpose in 2008 a test-beam study [78] was performed with a prototype of the calorimeter at the SPS beam line. As test-beam electrons, pions and muons were used and an energy resolution of 5 % to 20 % were found for pions and electrons respectively. The noncompensating effect is characterized by the ratio of pion to electron response (π/e) for electrons and pions with the same primary energy. It was found that the ratio of π/e is around 50 % in the detector, changing straightly with energy. Reproducing this test-beam measurement is the fundamental task of the radiation of the CASTOR detector simulation, see figure 4.5.

For the absolute energy calibration and the energy scale a data driven method is used by performing a cross calibration with 7 TeV data of the measured energy in the HF

calorimeter [79, 80]. The measured energy in HF ($-5 \leq \eta \leq -3$) is fully corrected to the hadron level with an uncertainty of 10 %. By using nine different models this energy is then extrapolated to the CASTOR acceptance ($-6.6 \leq \eta \leq -5.2$) within an uncertainty of 10 % as shown in figure 4.6. In CASTOR the extrapolated energy is then correlated to the hadron level response of the detector with an uncertainty of 5 % due to the noncompensation. The different components of uncertainties lead to a total energy scale of 15 %.

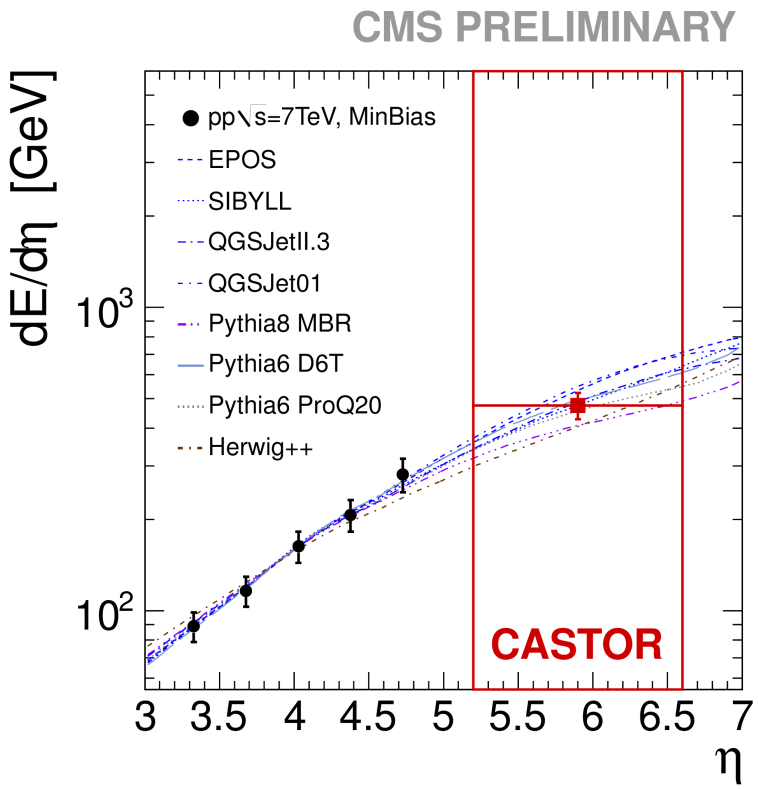


Figure 4.6: The energy prediction of different models fitted to the HF data from proton-proton collisions. The average extrapolation of the energy into CASTOR is indicated by the red dot. The two vertical red lines indicate the acceptance of CASTOR. [79]

4.2 Saturation correction

The charge integration and encoder (QIE) cards convert the analog signal from the CASTOR PMTs into a digital value. Thereby the charge from the PMT is integrated over time slices of 25 ns. For each time slice the integrated charge is then converted by the QIE into a 7-bit

digital number, so-called ADC-counts, in which a range of -3 to 26772 fC is covered. To get a good resolution for lower values of integrated fC and a very high dynamic range the conversion to ADC-counts is non linear as seen in table 4.1.

Range (Exponent)	Input Charge	ADC Codes	Gain
0	-3 fC — 36 fC	0 — 14	2.6 fC/bin
0	36 fC — 73 fC	15 — 21	5.2 fC/bin
0	73 fC — 104 fC	22 — 25	7.8 fC/bin
0	104 fC — 135 fC	26 — 28	10.4 fC/bin
0	135 fC — 174 fC	29 — 31	13 fC/bin
1	150 fC — 343 fC	0 — 14	13 fC/bin
1	343 fC — 525 fC	15 — 21	26 fC/bin
1	525 fC — 681 fC	22 — 25	39 fC/bin
1	681 fC — 837 fC	26 — 28	52 fC/bin
1	837 fC — 1032 fC	29 — 31	65 fC/bin
2	902 fC — 1877 fC	0 — 14	65 fC/bin
2	1877 fC — 2787 fC	15 — 21	130 fC/bin
2	2787 fC — 3567 fC	22 — 25	195 fC/bin
2	3567 fC — 4347 fC	26 — 28	260 fC/bin
2	4347 fC — 5322 fC	29 — 31	325 fC/bin
3	4672 fC — 9547 fC	0 — 14	325 fC/bin
3	9547 fC — 14097 fC	15 — 21	650 fC/bin
3	14097 fC — 17997 fC	22 — 25	975 fC/bin
3	17997 fC — 21897 fC	26 — 28	1300 fC/bin
3	21897 fC — 26772 fC	29 — 31	1625 fC/bin

Table 4.1: QIE conversion table of 7-bit ADC number to charge. The first column shows the value of the highest two bits in the 7-bit number, and in "ADC Codes" the last five bits are defined.

Only the ADC-counts for each time slice and channel of CASTOR are recorded. For later studies the ADC-counts are translated back to fC which are linear to the energy deposit inside the channel. In the years from 2010 to 2012 for each channel a recorded event of CASTOR had ten 25 ns time slices to sample the signal. In later years this was reduced to 6 time slices. How the signal pulse of the PMT is distributed over these time slices depends on the timing of the signal and the pulse shape. For the CASTOR detector the main part of the pulse will be integrated over the 4th time slice while the pulse tail is mostly contained in the following 5th time slice. A typical pulse shape from a CASTOR channel integrated over time slices can be seen in figure 4.7. It is obvious that a small fraction of the pulse extends well beyond the 5th time slice.

Later during reconstruction the ADC-counts for the two time slices 4th and 5th are converted to fC and summed up. If the integrated charge in one time slice is more than the

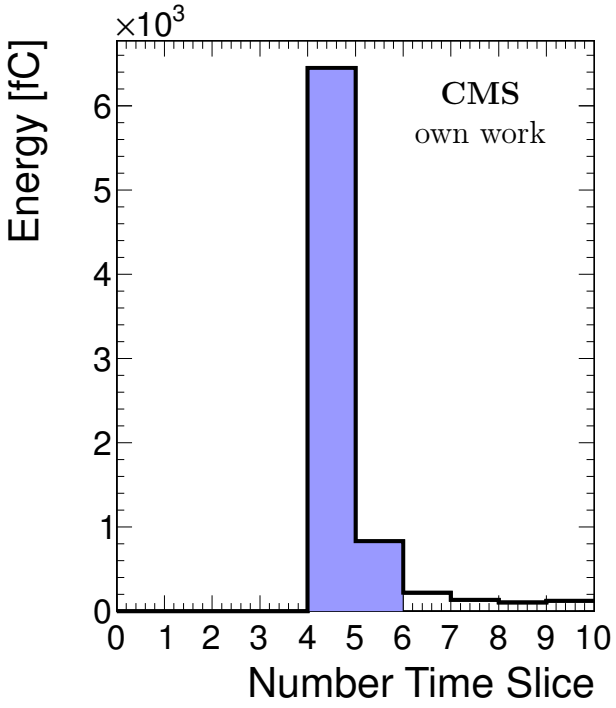


Figure 4.7: Averaged pulse shape of channel in sector 13 and module 1, shown are fC integrated over the time slices. The coloured area shows the two time slices which are later summed up during the reconstruction process to get the reconstructed channel energy in fC.

threshold of 26772 fC, the digital value for this time slice will be kept at the maximum QIE value of 127 ADC-counts and the original charge cannot be resolved anymore. In this case the channel becomes saturated. In 2010 for the MinimumBias data sample, 2.1 % of the events had at least one saturated channel while in a 2011 Commissioning sample just 0.5 % of the events had saturated channels. The ratio of events with saturated channels also depends strongly on the high-voltage settings of the PMTs of CASTOR. The effect of saturation in the CASTOR events is even more pronounced by selecting high energy events in CASTOR, in particular when selecting events with objects like high energy jets or electrons where the energy is clustered in a small number of channels in the CASTOR detector. This why it was of paramount importance to develop and implement a channel desaturation correction in the context of this thesis.

If a channel is saturated, then typically only the first time slice is saturated of the two summed up in the reconstruction due to the pulse shape. With growing energy and more severe saturation the value of the first time slice stays constant but the second time slice which covers the tail of the pulse will still grow in proportion to the signal. Since the

ratio of the integrated charge between the 4th and the 5th time slice is supposed to remain stable, then in the case the 4th time slice is being saturated the 5th time slice can be used to calculate the true charge of the pulse in fC. During the reconstruction of the CASTOR channels the original energy of a saturated channel can be recovered with the information of the second time slice by multiplying the energy in time slice 5 with a desaturation factor to get the energy of the first one. These desaturation factors are different for each channel of CASTOR because the different PMTs and cables for the signal transmission form a different pulse shape. Also the overall signal timing varies in the ns range from channel to channel. The factors also change with time.

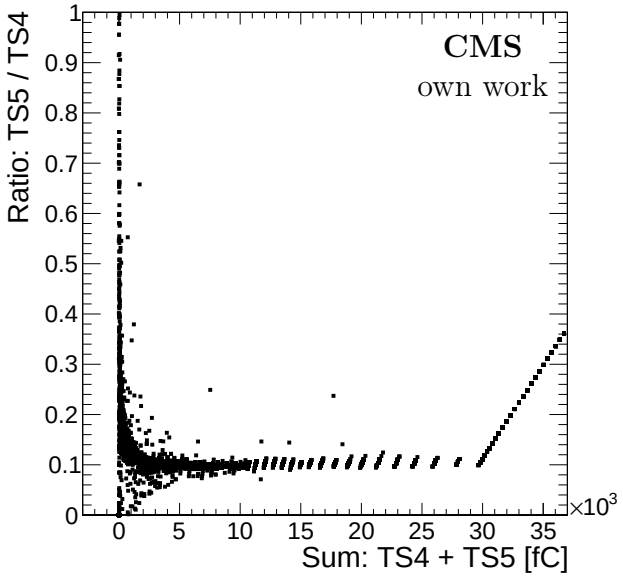


Figure 4.8: For an example channel (module 1, sector 15) in CASTOR on the y-axis the ratio between the integrated fC of the time slice with the main part of the PMT signal pulse (TS4) and its tail (TS5) is shown. This is shown versus the sum of the two time slices TS4 and TS5 in fC on the x-axis which is the reconstructed signal of the channel. For a higher signal the ratio is relatively stable. Because of the digitization process of the charge the ratio is not perfectly stable and jitters around the true value. The effect of saturation can be very well seen in the rising ratio for a very high signal in the time slice sum.

Figure 4.8 shows a typical measured ratio between the integrated charges of the two time slices used for reconstruction, depending on their sum in fC. In the region around 10 up to 25 thousand fC the pulse shape of the PMT and therefore also the ratio between the two time slices is stable. Here the remaining observed structure of data points is an artefact of increment using ADC-counts to calculate back the integrated fC. Starting from around 30×10^3 fC the first time slice (TS4) starts to saturate. Therefore with increasing energy the ratio also increases because only TS5 keeps rising. At the lowest energies the ratio

shows strong fluctuations according to the noise from the QIE and CASTOR. To estimate the stable value of the ratio the average is calculate in the range where the pulse shape is stable and no saturation effect is yet observed.

For this purpose special data is required that still contains the information about the time slices and ADC-counts. A list of datasets and runs is shown in table 4.2. No further

Table 4.2: List data used during commissioning periods to determine the desaturation factors.

dataset	runs	events	saturated events
/MinimumBias/Run2010A-v1/RAW	136066	66324	24
	136082	62487	25
	142928	58664	1325
	144089	66068	1380
/MinimumBias/Run2010B/RAW	146437	74453	1019
	148822	108233	2994
/Commissioning/Run2011A-v1/RAW	161439	8334	49
	178866	47457	1
/PAMinBiasUPC/HIRun2013-v1/RAW	210738	195614	477
	211390	126951	576
/MinimumBias/Commissioning2015-v1/RAW	245194	374031	0

selection of events is done.

As shown in figure 4.8 it is important to select signal ratios in the stable region. This is achieved by requiring that the number of ADC-counts in time slice 4 is within 110 and 126. Also there must be at least more than 5 data points inside this range to get a reasonable result. This allows easily to go very close to the point where the TS4 starts to saturate. As an example in figure 4.9 the ratio of the two time slices is shown depending on the sum of energy in GeV of the two time slices together. The points marked in red are used to get the final ratio for the channel, which is calculated as the average value in this region.

For 2010 the results are shown in figure 4.10 for all CASTOR channels. In some channels no value is given because of low statistics. Especially for the channels in the middle of CASTOR where the magnetic field of CMS suppresses the PMT gain resulting in low signals even with a reasonable energy deposit inside the channel. Also in the back part of CASTOR there are normally no high energy deposits because neither electro-magnetic nor hadronic showers penetrate into the tail of the detector. On the other hand, in those regions of the calorimeter saturation never occurs.

Because of minor changes in the timing or changes at the electronics of the CASTOR detector these ratios do change over time. Especially after re-installation of CASTOR inside the CMS experiment new factors have to be determined. How the ratios for one channel in CASTOR are changing between different runs can be seen in figure 4.11. Here

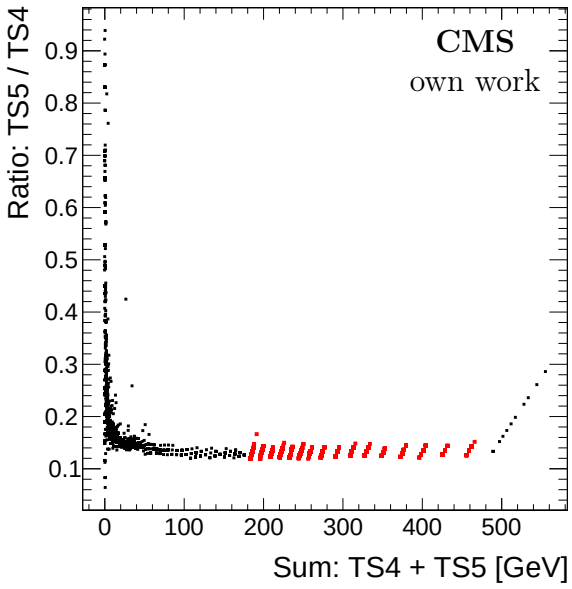


Figure 4.9: For the run 144089 the channel in the first module and sector 14 of CASTOR the ratio of the integrated charge of two time slices versus signal in GeV of the two time slices used in the reconstruction is shown. The points highlighted in red are used to determine the ratio $R_{\mathrm{TS5/TS4}}$. The black points are not used and only every 1000 event is shown.

a significant increase in the ratios between the run number 178866 from 2011 and 210738 from 2013 can be observed. An explanation is that in 2012 the CASTOR detector was upgraded before the detector was re-installed in 2013. One task during the upgrade was the exchange of many PMTs. These variations are typically on the order of 10 % but can also reach 40 % in extreme cases.

During the normal event reconstruction, the desaturation algorithm is used to measure the signal when the 4th time slice is saturated. Therefore, the values are written into the CMS conditions database to access them in the code via the "global tag". If it was not possible to calculate the factor for a particular channel, the mean value of all other channels inside the same module is written to the database. This is done as long as more than four channels in the module have a value. Otherwise the mean of the ratios of all channels in CASTOR is taken for the database. Thereby, all channels are associated with a value in the database.

Since software version `CMSSW_5_3_8` this desaturation procedure is used by default during event reconstruction. A CASTOR channel, also called rechit in the reconstruction software, uses these values from the database to recover the rechit energy if one time slice is saturated. The desaturation can be optionally switched off by a flag in the configuration. A rechit has additional information in the form of several flag bits beside the signal strength. One of these is the `HcalCaloFlagLabels::ADCSaturationBit` which is set true if either of the two time slices is saturated. This flag can be accessed by the user because after

reconstruction only the total rechit energy is saved. The desaturation software checks if the 4th time slice, with the main signal, is saturated and not the 5th one with the pulse tail. If the 5th time slice has the maximum value of ADC-counts, this is evidence for another problem. Therefore desaturation cannot be performed in this case. The new energy for the desaturated rechit is calculated via

$$E_{\text{new}} = E_{\text{TS5}}/R_{\text{TS5/TS4}} + E_{\text{TS5}} = E_{\text{TS5}}\left(1 + \frac{1}{R_{\text{TS5/TS4}}}\right) \tag{4.2}$$

where E_{TS5} is the energy of the 5th time slice and $R_{\text{TS5/TS4}}$ is the ratio value seen in figure 4.10 or figure 4.9 as taken from the database. In cases where the desaturated signal is smaller than the saturated one, which may occur when the 5th time slice is noisy, the rechit keeps the initial saturated value. If a saturated rechit was recovered, then the `HcalCaloFlagLabels::UserDefinedBit0` flag bit is set true and can be used during the analysis by the user.

In Monte Carlo detector simulation the simulated response of a channel is only calculated for the 4th time slice. In this situation a desaturation is not possible, but can be easily avoided by choosing reasonable gain values.

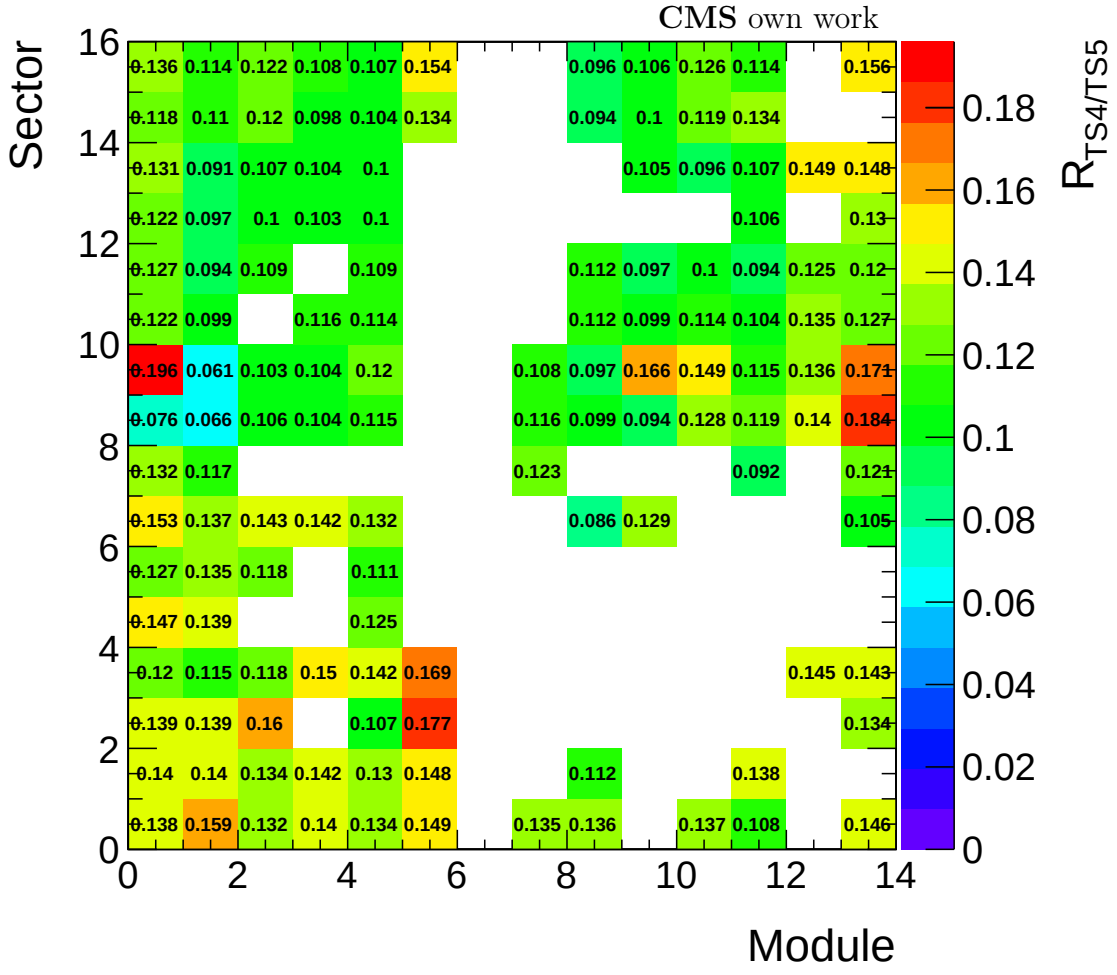


Figure 4.10: The final ratio $R_{\mathrm{TS5/TS4}}$ between the 4th and the 5th time slice of all channels in CASTOR for run 144089. On the x-axis the module number and on the y-axis the sector number are shown. The missing channels are missing due to a lack of statistics. These gaps are filled as described in the text.

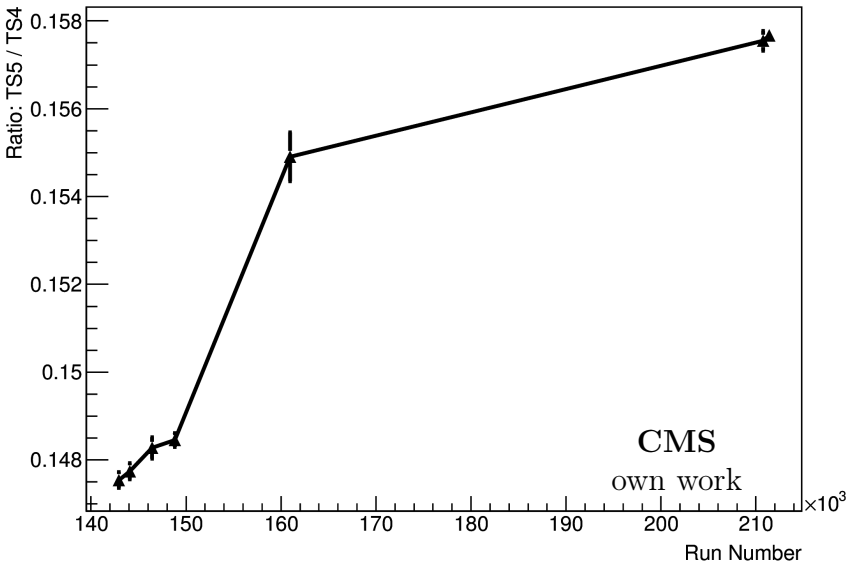


Figure 4.11: Depending on the CMS run number the final ratio of the 4th and 5th time slice for the channel in the first module and sector 5 of CASTOR is shown. The runs were taken in the years from 2010 to 2013. The error bar shows the RMS of the data points which are used to calculate the ratio of the run.

4.3 Trigger

As part of the CMS trigger system the CASTOR detector is able to trigger a CMS event. Because of the very forward acceptance of CASTOR the CMS experiment has the opportunity to trigger unique LHC physics. One possibility is to use the segmentation in the EM and hadronic part of the CASTOR detector to trigger very forward high energetic e^-/γ or pions and protons. Another one is to trigger high energy forward jets in which the prediction of different models, especially cosmic ray models, can be compared.

Since a few TOTEM triggers are connected to the CMS trigger system, TOTEM events can also be combined offline with CMS events. Especially combining information from the TOTEM T2 tracker with the CASTOR detector is an advantage because both have the same η-acceptance and see the same particles.

For the measurement of forward electrons in 2013 p–Pb data a dedicated electron trigger based on a combination of TOTEM T2 and CASTOR trigger bits was implemented. In order to measure the CASTOR jet spectrum with data taken in 2015 at 13 TeV also a jet trigger was developed as part of this thesis and implemented in connection with the work in this thesis.

4.3.1 Electronics

With a bunch spacing of 25 ns there is a theoretical collision rate of up to 40 MHz. This is far too much to be saved on the computer disk or for any data analysis. One needs to trigger events potentially interesting for physics in order to reduce the amount of CPU and memory usage. To do this there are two selecting steps in the CMS trigger system. At first, the events are selected by the level-1 (L1) trigger which is a hardware based trigger. In a second step the L1 triggered events are selected via a software based high-level trigger (HLT).

The L1 trigger decisions are created by the programmable electronics of the sub components in CMS, also for the CASTOR detector. Then these trigger bits are sent to the CMS global trigger system (GT) which is separated in two parts of trigger bits. One part are the L1 technical triggers which represent the trigger decisions directly coming from the CMS subdetectors. Up to 64 technical trigger bits can be received by the GT. The second part are the L1 algorithm triggers which are created by the global trigger logic. Here algorithm calculations like simple logical combinations of trigger bits or more complex operations like applying energy thresholds can be done. In sum there are 128 algorithm trigger bits which can be processed. Both the 64 technical trigger bits and the 128 algorithm trigger bits are sent to the final decision logic. In physics run conditions not all $128+64$ trigger bits trigger an event. A mask on the trigger bits is applied to decide which event will be triggered. The electronics was designed to handle an output rate of 100 kHz of the L1 trigger. Note that

for RUN II the GT was upgraded significantly [81] to relax the constrain on the number of algorithm trigger bits.

In CASTOR the L1 trigger bit generation starts with the signals from each channel sent from the QIE to the HCAL trigger and readout (HTR) cards. For each octant in CASTOR one HTR card combines the signals of 2 sectors into 4 octant trigger bits. One output connector of the QIE has space for the readout of three detector PMTs. Because one HTR card can be connected with eight of these QIE connectors, only the signals of the modules 1 to 12 can be processed in each sector. Therefore the last 2 modules, 13 and 14, are missing for the trigger bit generation. In total there are 24 channels per HTR card and 8 HTR cards for the whole CASTOR detector. Each HTR card has a 4-bit output. Inside the HTR card the ADC signals are first converted from ADC-counts to fC via a very simple lookup-table. This fC value from the lookup-table is not the same as the reconstructed one because the lookup-table is not as precise as the offline QIE conversion table due to limited memory space. Each HTR card takes the signals of two sectors in fC from the lookup-table to generate 4 octant trigger bits which are sent to the CASTOR technical trigger processor (TTP). The TTP combines the octant trigger to 4 CASTOR trigger bits. These four CASTOR trigger bits are sent as L1 technical triggers to the global trigger system of CMS and are part of the L1 trigger decision.

CASTOR is able to trigger eacg time slice (see sec. 4.2) synchronized to the bunch spacing of LHC. The HTR cards use for the calculation of the trigger decision the charge integrated signal of only one time slice even though it is known that the pulse shape of the signal is distributed over two time slices (see figure 4.7). If LHC were configured to have a bunch crossing every 25 ns, then the distribution of the pulse shape over two time slices would lead to mis-triggering because the pulse tail of the signal of the previous bunch crossing would add to the actual time slice. Therefore CASTOR is not included in such conditions. In any case when the signal is high enough the pulse tail in the second time slice could fire a trigger bit a second time directly after it was set during a collision. To avoid this, the HTR cards have a dead time of one time slice and therefore do not set any trigger bit if it was triggered already one time slice before.

After an L1 trigger decision has been made, the triggered event is sent to the software based filter system called high-level trigger (HLT). The HLT consists of several trigger paths which are based on specific L1 trigger bits or a logical combination of them. These paths run already parts of the reconstruction software to make more intelligent decisions on an event. To save CPU time only the HLT paths depending on the L1 triggers which were set in an event are considered. If the right L1 trigger is set and the event passes through the selection criteria of the HLT path, the event will be permanently saved. The simplest version of an HLT path is a pass-through of an L1 trigger decision. In a more complex scenario the HLT path runs software parts like the track or jet reconstruction and can perform a cut on jet-p_T or the track multiplicity. The different HLT trigger paths are clustered into data streams. If an event triggers a HLT path of a certain stream, it will be saved in the corresponding dataset. Thereby the HLT needs to handle the L1 input rate

of 100 kHz and reduce it to an output rate of around 1 kHz.

4.3.2 Forward electron trigger

During heavy ion runs there is a strong accompanying flux of photons, leading to a high cross section of ultra peripheral collisions (UPC) of photon-photon or photon-nucleus. This is caused by the strong electric field of the heavy ion nucleus. In 2013 this was the case during the p–Pb run period at LHC. Often in these ultra peripheral collisions there are two leptons/electrons produced which are directed into the forward direction. The CASTOR forward calorimeter can be used to detect such electrons/leptons. Therefore CASTOR provided an electron trigger which was the very first physics L1 trigger provided by CASTOR. In addition, this trigger was combined with TOTEM to use T2 tracking information at trigger level together with CMS. This also allows a combined reconstruction of T2 and CASTOR data to select electrons in the very forward direction.

Trigger functionality for electrons

The CASTOR trigger is generated octant wise in the HTR cards. Each octant produces an electromagnetic (EM) trigger and a hadronic trigger used as veto signal. The electromagnetic trigger operates separately on each sector and sums up the signals of the first two modules. If this sum in one sector is above the threshold of around 45 fC, the whole octant produces the electromagnetic trigger signal. The thresholds change slightly from sector to sector to compensate for different gains.

For the hadronic trigger the signals from the 4th to the 12th module are summed up. Because of possible electromagnetic shower leakage into the 3rd module of CASTOR this module is not taken into account. If in a sector the sum is above 110 fC, the octant sets the hadronic trigger bit. Here the thresholds differ for each CASTOR sector as well. As described above only the signals of the 4th time slice containing the main part of the signal are taken for the electromagnetic and the hadronic sums.

After processing the electromagnetic and the hadronic trigger bits of the CASTOR octants they are sent to the TTP to form a combined L1 technical trigger. The L1 technical trigger for CASTOR electrons selects events that have at least one octant with the electromagnetic trigger. But the event is vetoed if any octant triggered the hadronic part. Therefore only clean events with an energy deposit in the front part but not in the back part of the detector are taken, which resembles the characteristics of electrons/photons.

Additionally for the electron trigger a common CMS-TOTEM trigger, readout and event merging were developed by the CASTOR group from CMS side. One additional cable was installed to transmit 4 trigger bits from TOTEM to the CMS GT. A second additional cable was installed to transmit one single trigger (CMS L1-special-accept|L1SA) from the

CMS GT to TOTEM. The L1SA depends on a very small subsample of the GT trigger bits in the standard L1-accept (L1A). This is needed since the CMS L1 trigger has a rate of up to 100 kHz, but TOTEM can only read out with a rate lower than 1 kHz and the CMS HLT cannot be used to trigger TOTEM. The combined trigger rate of L1SA is selected to be < 1 kHz. This allows CMS to trigger events depending on the 4 TOTEM trigger bits and allows TOTEM via the L1SA to record events triggered with CMS. Offline these events can be merged with a very high efficiency from the two CMS and TOTEM datasets. Because TOTEM records every event which is triggered by the L1SA, CMS should also record every L1SA accepted event. Therefore CMS needs HLT paths with a pass through of the trigger bits included in the L1SA.

In 2013 the electron trigger was also part of the L1SA. It is a combined TOTEM-CASTOR algorithm trigger based on the described CASTOR technical trigger and a T2 low-multiplicity trigger from TOTEM. The T2 tracker consists of two identical tracking detectors. One is placed on the positive side of CMS and the other one on the negative side, directly in front of the CASTOR detector. Both the T2 tracker and CASTOR cover the same η-acceptance. TOTEM triggers in T2 the track multiplicity to be lower or equal than 4 at any of the two sides. There is no condition for the corresponding opposite tracker side relative to the triggered one. The track multiplicity trigger is not based on reconstructed TOTEM tracks but it is computed via a simple algorithm based on dedicated trigger-pad hits in T2.

Testing and validation of the trigger

For the first octant in CASTOR the functionality of the EM and hadronic trigger can be seen in figure 4.12. It shows the electromagnetic energy sum in fC of the first sector versus the 2nd CASTOR sector which is part of the same octant. On the left side the sum of the first two modules as EM energy is shown and on the right side the sum of modules 4 to 12 for the hadronic energy. The events with an EM or hadronic trigger in this octant are shown in blue, the others in red. For the EM and the hadronic trigger it can be seen clearly that if in one or in the other sector the energy is above the threshold, the octant will be triggered. This is the expected behaviour the triggers must have. This information cannot show this behaviour for each individual sector because the CASTOR trigger information is recorded on the octant level by the TTP. Therefore only the trigger information of the CASTOR octants and not the sectors separately are available for a validation of the trigger.

As part of my work as the CASTOR trigger contact person this trigger logic was also carefully investigated. And a problem which was found and fixed can be seen on the left side of figure 4.13. It shows for the third octant in CASTOR the energy sum of the first two modules of sector 5 versus sector 6. It indicates a misbehaviour of the EM trigger in the 5th sector. Even when the energy in the first two modules is very high, the octant does not trigger. This was identified to be related to a cable mismatch of the QIE to the HTR card.

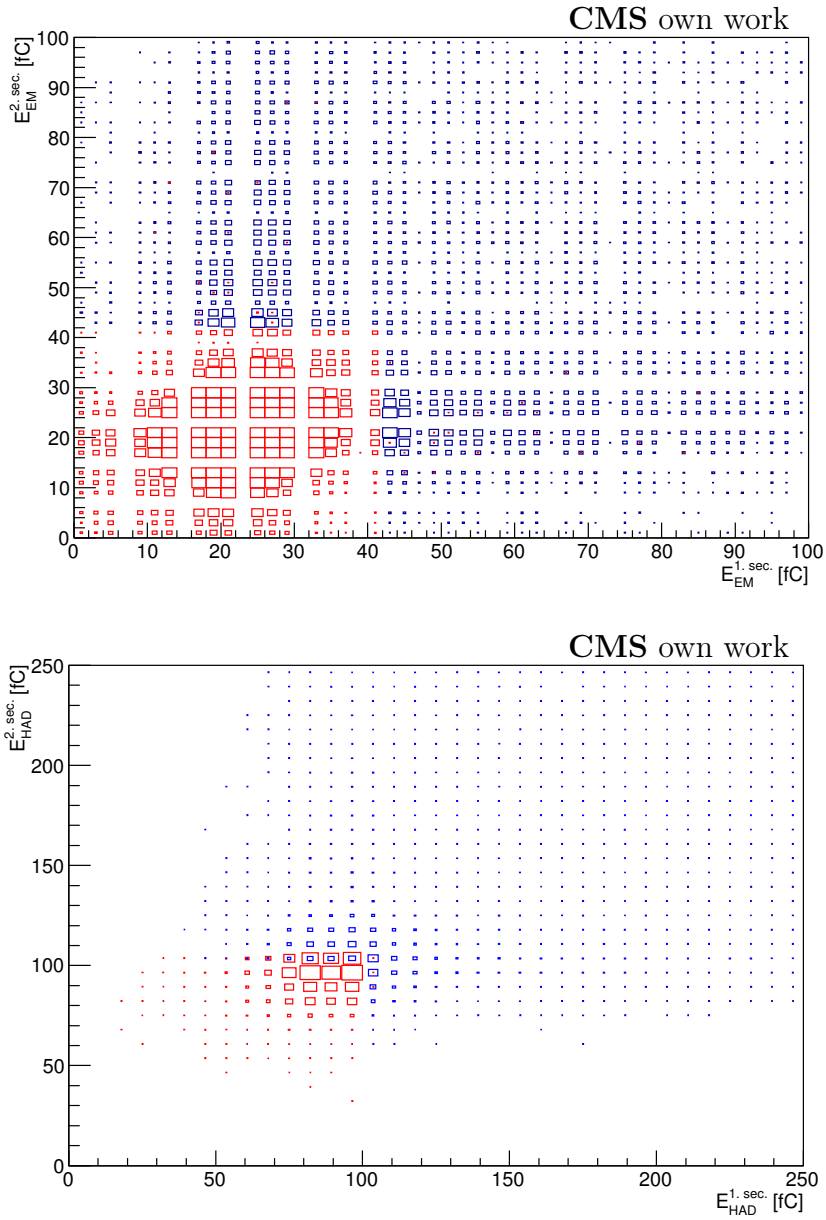


Figure 4.12: For the first octant in CASTOR the electromagnetic energy of the 1st versus the 2nd sector of the octant is shown on the top and the hadronic energy on the bottom side. If the electromagnetic trigger bit of the first octant has triggered, then on the top side the events are shown in blue, otherwise in red. On the bottom side the events are shown in blue if in the octant the hadronic trigger was set, otherwise in red. The size of the boxes are correlated to the number of events but the blue and the red boxes have a different scale.

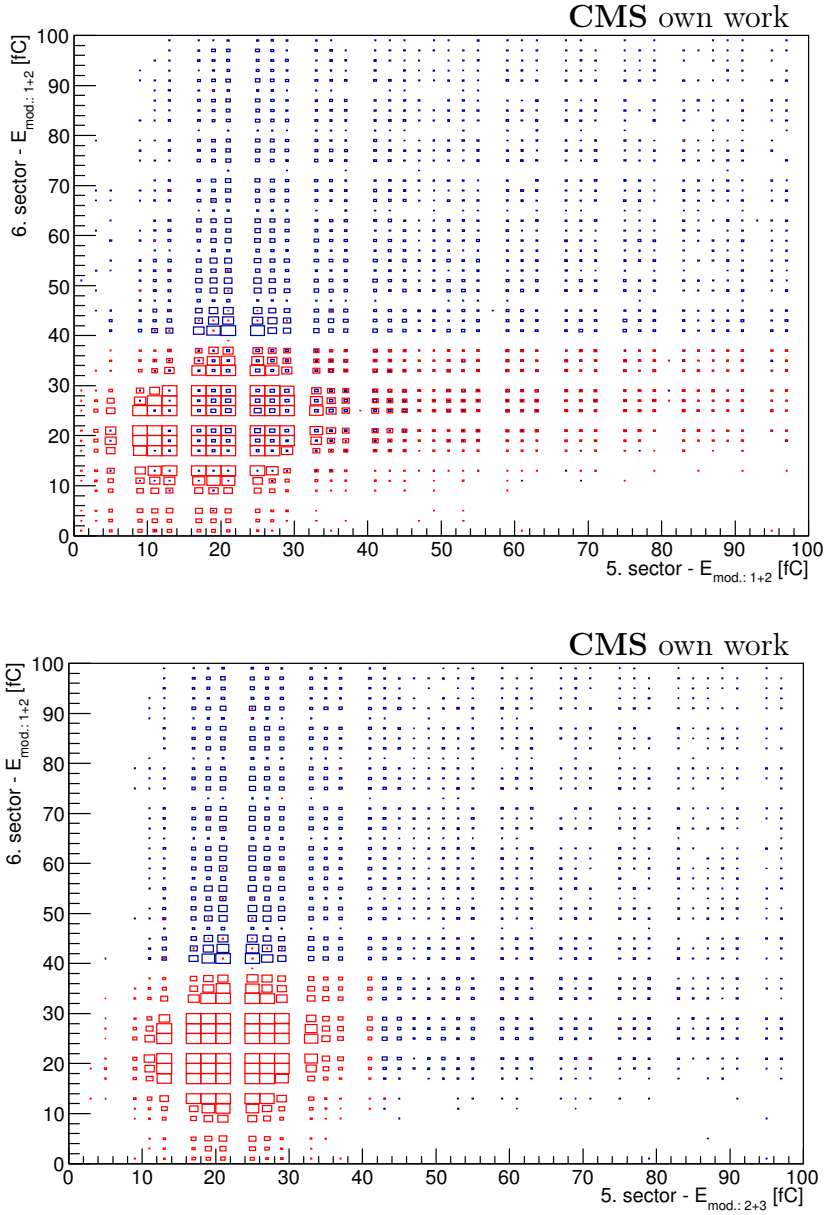


Figure 4.13: On the top side for the two sectors of the third octant the energy is shown as sum of the first two modules. The bottom side shows for sector 5 the energy sum of module 2 plus 3 versus the energy sum of module 1 plus 2 of the 6th. sector. In both figures the blue events indicate the electromagnetic triggering of the third octant and the red ones the opposite.

As described before the connection from the QIE to the HTR card always bundles three channels of a sector. The HTR cards have no information about the channel number and take the input order of the connectors to determine the channel assignment. For the

energy sum of the first two modules the HTR card takes the signals of the first two cable connections. Due to a mismatch in the cables for some sectors the HTR card summed up for the EM energy the signals of the 2nd and the 3rd module as it is the case of sector 5. This is considered in the right plot of figure 4.13. There is the sum of the 2nd and 3rd module taken as electromagnetic energy of the fifth sector accounting for this mismatch of the channels in the HTR card. The right side of figure 4.13 shows again the expected behaviour of the EM trigger in the 3rd octant. But still there is a small number of events without a trigger signal even when energy is well above threshold. This can be seen on the right side of figure 4.12. Most probably this happens because the octant EM trigger bit often was set already one time slice before and therefore cannot trigger again in this time slice.

The same problem occurs also in the hadronic trigger. But the hadronic trigger sums up all channels in a sector from module 4 to 12 where a permutation of the order would not lead to a different result and the trigger still works correctly. Due to a bug in the programmable hardware of the HTR card for some sectors the hadronic trigger summed up one channel twice. After the discovery of this problem it was also fixed.

The electron trigger was very successful in recording a large sample of T2+CASTOR electrons which were then used for the only physics data driven CASTOR alignment. And only the complete understanding and debugging of the EM and the hadronic trigger of 2013 allowed the development of a stable physics trigger in CASTOR like the high energy jet trigger in 2015.

4.3.3 CASTOR jet trigger in 2015

One objective in the 2015 data taking was the analysis of jets in the CASTOR detector. Because the detector is positioned very forward, around $\eta = -6$, such jets easily have an energy of more than a TeV. At proton-proton collisions of 13 TeV these high energy forward jets are very rare and a trigger is needed to gain sufficient statistics in this phase space region. For this purpose two jet triggers, a high- and a medium-energy jet trigger, were provided for CASTOR. Note that while CASTOR has no η segmentation, the reconstructed energy and the reconstructed p_{T} of a jet in CASTOR are strongly correlated and the energy can be estimated via $E^{\mathrm{jet}} = \cosh(6) p_{\mathrm{T}}^{\mathrm{jet}} \approx 200 p_{\mathrm{T}}^{\mathrm{jet}}$.

Like the EM and the hadronic trigger the medium jet trigger is generated for each octant in the HTR cards. For the medium-energy jet trigger in each sector separately the sum in fC of the signals in the modules from 1 to 12 is taken. If in one sector this sum is above the threshold of 15740 fC, which is to be around 850 GeV, then the medium jet trigger is triggered by the corresponding octant. If any octant has set the trigger bit, it will be set for the whole CASTOR detector and sent to the GT as an L1 technical trigger.

This is very similar to the generation of the high-energy jet trigger in CASTOR, except that the HTR card electronics sum up the signals of mod 4 to 12 in one sector. If this sum

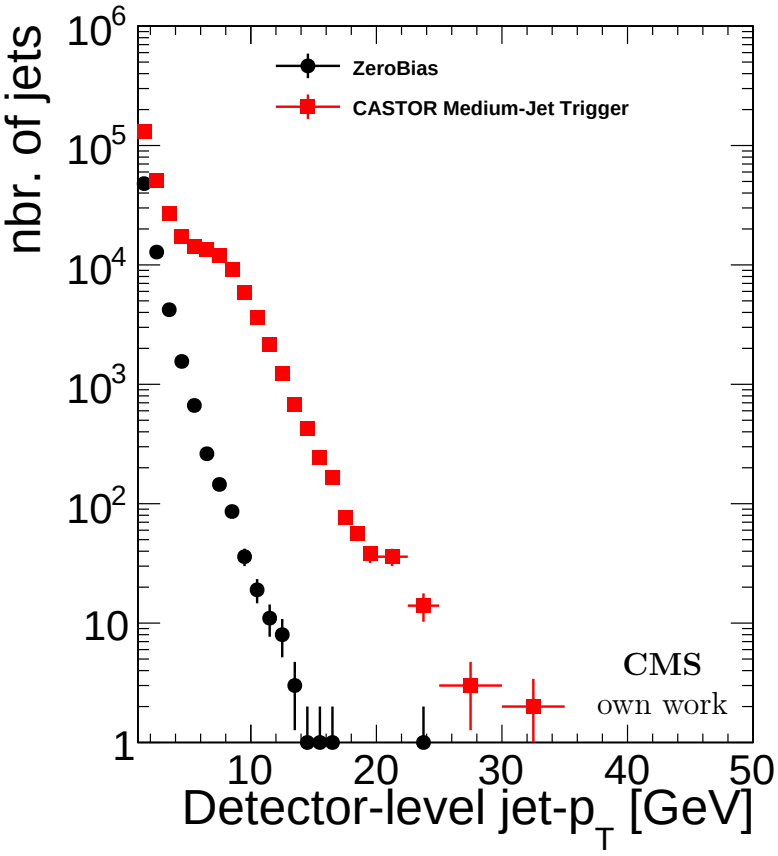


Figure 4.14: For different datasets the overall number of jets in 13 TeV proton-proton collisions of the run 274623 is shown. In black for ZeroBias and in red for events with the CASTOR medium-jet trigger. The amount of datataking was around $0.212\,\mathrm{nb}^{-1}$ in this run.

is above 13000 fC and additionally the signals in the first two modules of the sector are saturated, the whole octant triggers the high-energy jet bit. The condition of the saturated front modules was applied not to be biased by hadronic jets when using a very high total sector energy sum. To exceed a high threshold a strong activity on the whole length of CASTOR is needed due to the saturation effect. But during physics runs it was concluded that the trigger was not efficient and therefore not useful for analysis.

For a single run in figure 4.14 the total number of jets depending on the detector-level jet p_T is shown for ZeroBias and the medium-energy jet triggered events. In the medium-energy jet trigger sample there are many more jets, especially in the higher p_T region, than for ZeroBias. Because of a much lower rate of the CASTOR medium jet trigger compared to the ZeroBias trigger the medium jet trigger had only a prescale of one, while the ZeroBias trigger had an effective prescale of 205.5. For the jets with higher p_T where the jet trigger reaches the full efficiency (see figure 4.15) this effective prescale factor of 205.5 is exactly

the ratio between the medium jet trigger and the ZeroBias p_T spectrum in figure 4.14. The relatively large amount of jets with low p_T for the jet triggered events is produced by events with multiple CASTOR jets where a high p_T jet which triggers CASTOR is accompanied by a low p_T jet. Figure 4.14 shows clearly the benefit in statistics compared to ZeroBias by using the CASTOR jet trigger. The kinematic limit in p_T in this distribution is about 30 GeV.

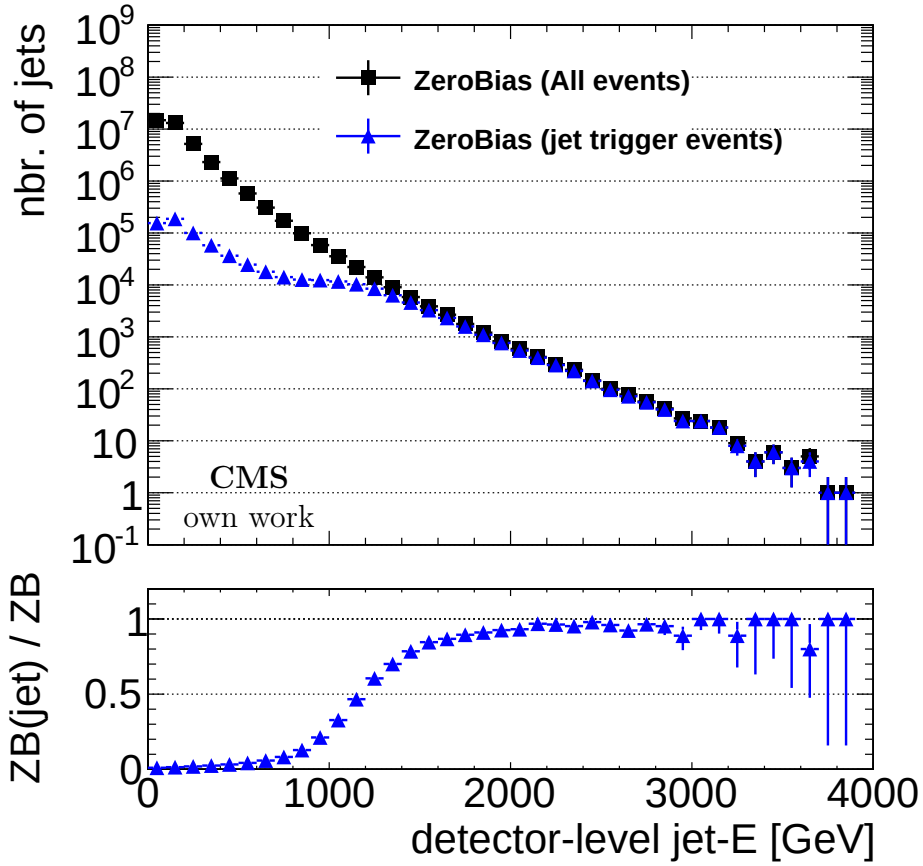


Figure 4.15: In the top plot the number of jets versus the detector level jet energy is shown, in black for ZeroBias events and in blue only for ZeroBias events with the CASTOR medium jet trigger. The bottom plot shows the ratio of the triggered events divided by all ZeroBias events and expresses the jet trigger efficiency depending on the jet energy.

The trigger efficiency was estimated by a ZeroBias data sample in which CASTOR medium jet triggered events were searched for. At the top of figure 4.15 the total number of jets depending on jet energy can be seen for ZeroBias events in black and in blue events with medium jet trigger which are a subsample of the ZeroBias events. The ratio of these two distributions is shown at the bottom of figure 4.15. It shows the trigger efficiency which reaches at around 2 TeV the plateau where the trigger is approximately 100% efficient.

Note that the error bars for the trigger efficiency were calculated via the "Clopper-Pearson interval" and not by Gaussian error propagation.

CASTOR jet trigger simulation

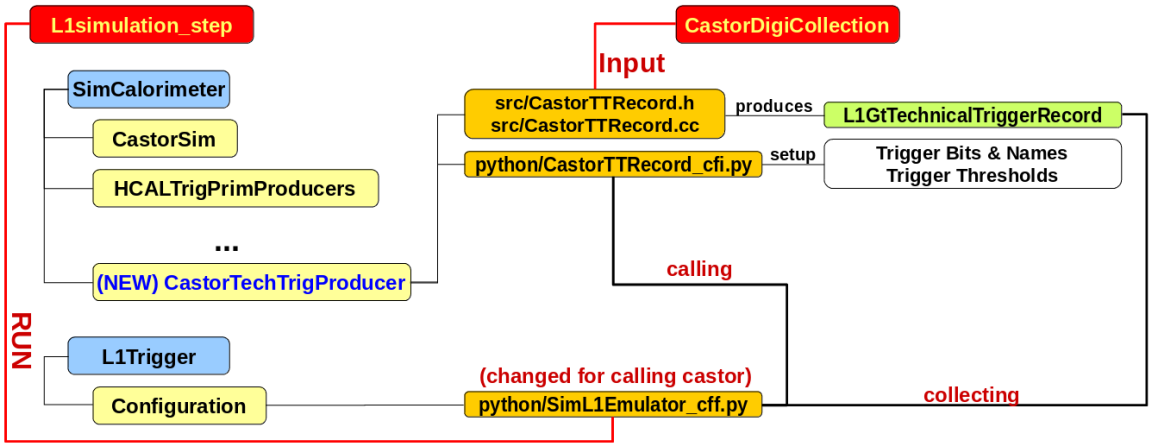


Figure 4.16: Scheme of the inclusion of the CastorTechTrigProducer module in the CMS simulation software.

For the CASTOR jet trigger for the first time in the CMS detector simulation software a module was added for a full simulation of the CASTOR triggers. A scheme of how the new package in the detector simulation framework is included can be seen in figure 4.16. Since version `CMSSW_7_5_0` the CASTOR jet trigger, muon trigger, gap trigger and the electron trigger from the previous chapter are included. But in the standard configuration the electron trigger is not simulated since it was not part of the 2015 CASTOR L1 trigger menu.

In preparation for the proton-proton collisions in 2015 it was necessary to estimate the CASTOR jet trigger rate for different thresholds to assure that the trigger rates do not reach the limits of the L1-accept or the HLT. Therefore simulation studies with a PYTHIA8 MBR sample were used. For different thresholds in GeV the black data points in figure 4.17 show the rate of the jet trigger for different thresholds of the jet trigger. The rate is calculated by $\sigma_{\text{process}} \times L_{\text{inst}}$ with the cross-section σ_{process} which can be seen on the right y-axis of figure 4.17 and the instantaneous luminosity $L_{\text{inst.}}$. To calculate the instantaneous luminosity the formula

$$L_{\text{inst}} = \frac{\text{bx}}{\text{bs} \times 3564} \times \frac{\lambda}{\sigma_{\text{inel}}} \left[\frac{\text{Hz}}{\text{b}}\right] \tag{4.3}$$

was used where the number of colliding bunches bx = 40, the bunch spacing bs = 25 ns, the inelastic cross-section $\sigma_{\text{inel}} = 80\,\text{mb}$ for pp collisions at 13 TeV and the pile-up (PU) $\lambda = 0.4$. The number 3564 refers to the total number of possible bunches per orbit. Therefore the

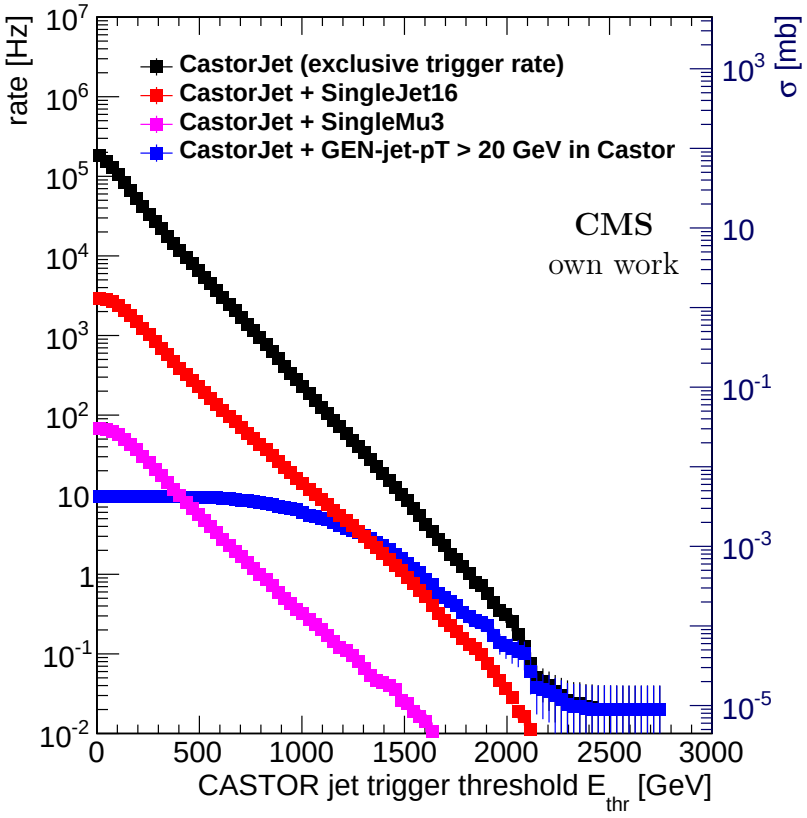


Figure 4.17: The estimated rate depending on the CASTOR jet trigger threshold is shown for *pp* collisions at 13 TeV with 40 colliding bunches, an interaction probability of 0.4 and a total inelastic cross section of 80 mb. The black data points show the rate of the jet trigger alone while the other data points show different combinations of the jet trigger with other CMS L1 triggers. The right y-axis shows the corresponding cross section in mb.

factor $1/(25\,\text{ns} \times 3564) \approx 11\,\text{kHz}$ corresponds to the orbit frequency of the LHC beam. In figure 4.17 the rate was estimated for the CASTOR jet trigger alone and for different scenarios where the jet trigger is combined with central CMS triggers/objects as it could be done on L1 algorithm or HLT level.

These simulations were used to select the optimal trigger working point for the data taking. A rate of 300 Hz without pre-scaling was aimed for. In the jet trigger simulation the trigger threshold on the sector energy was set to 65000 fC which is 975 GeV in simulation. This results in an estimated trigger rate of around 270 Hz. A comparison of the jet trigger efficiency depending on jet energy from MC and from ZeroBias data can be seen at the top of figure 4.18. The efficiency curve was produced in a similar way as for figure 4.15. One can see that in MC the turn-on of the trigger efficiency starts a little bit later in MC

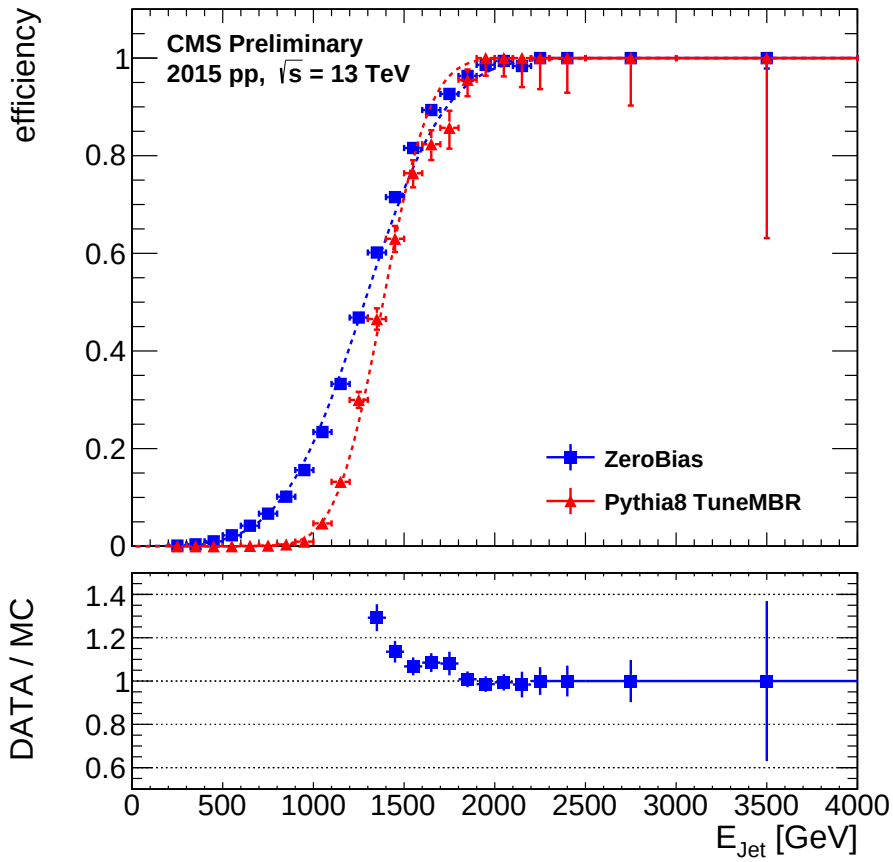


Figure 4.18: Shows at the top the CASTOR jet trigger efficiency depending on the jet energy in GeV for MC (red) and ZeroBias data (blue). The dashed line is a chi squared fit to the data points with the gaussian antiderivative. The ratio of the two efficiency distributions is shown at the bottom.

but is steeper than in data. The smaller steepness in data could be an effect of the inter-calibration of the CASTOR channels which is needed in the data reconstruction but not in MC. The errors at the top of figure 4.18 are again calculated via the "Clopper-Pearson interval". The dashed lines are a simple fit of the cumulative gauss function on the data points for a better illustration. At the bottom of figure 4.18 the ratio between the two efficiency curves is shown. At energies above 2 TeV both the data and the simulations reach full efficiency. No trigger-related correction factor is needed in this region.

4.4 CASTOR alignment with tracks from TOTEM T2

The goal is to align CASTOR by using events with single electron tracks in the T2 tracker which also hit the CASTOR detector where they result in an electromagnetic signal in the first two modules (first 20 X_0). Such an event with a single electron of 320 GeV is shown in figure 4.19. The idea is that an electron hitting CASTOR close to the border of two ϕ-segments will result in a splitting of the signal in both ϕ-sectors. With the high position resolution of the hits in the T2 tracker the structure of the ϕ-segmentation will become visible. Consequentially, the exact position of the CASTOR front face can be determined from the data and is only restricted by the resolution of T2.

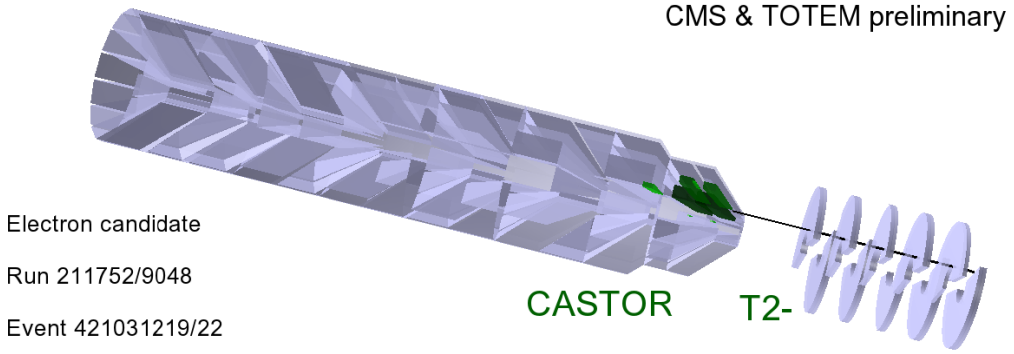


Figure 4.19: A graphical view of an event with a single electron passing the TOTEM T2 detector from the right and inducing an electromagnetic shower in the front part of CASTOR. The reconstructed energy of the electron was 320 GeV.

The T2 tracker from the TOTEM collaboration [74] is perfectly suited for this purpose because the η coverage of T2 ($-6.5 < \eta < -5.3$) overlaps almost perfectly with the acceptance of CASTOR ($-6.6 < \eta < -5.4$). Thus it is possible to find an isolated activated sector in CASTOR and one good associated track in T2.

4.4.1 Selection of events for alignment

To align CASTOR one good track on the CASTOR side (negative z) of the T2 tracker is needed and furthermore events where in the CASTOR detector the energy deposit is in one isolated sector so that the T2 track can be associated with this sector. How to determine "good" T2 tracks and find isolated CASTOR sectors is part of this section.

For this task a dedicated data sample was recorded in 2013 with a custom very forward electron trigger. Data were collected with CMS and TOTEM simultaneously using a special trigger hardware handshake. The data for this analysis was then merged offline to yield all available information content. This trigger as described in sec. 4.3.2 was a combined TOTEM-CASTOR algorithm trigger and also referred to as CMS algorithm trigger bit 99 for the position of the trigger bit in the trigger word.

Classifying T2 tracks

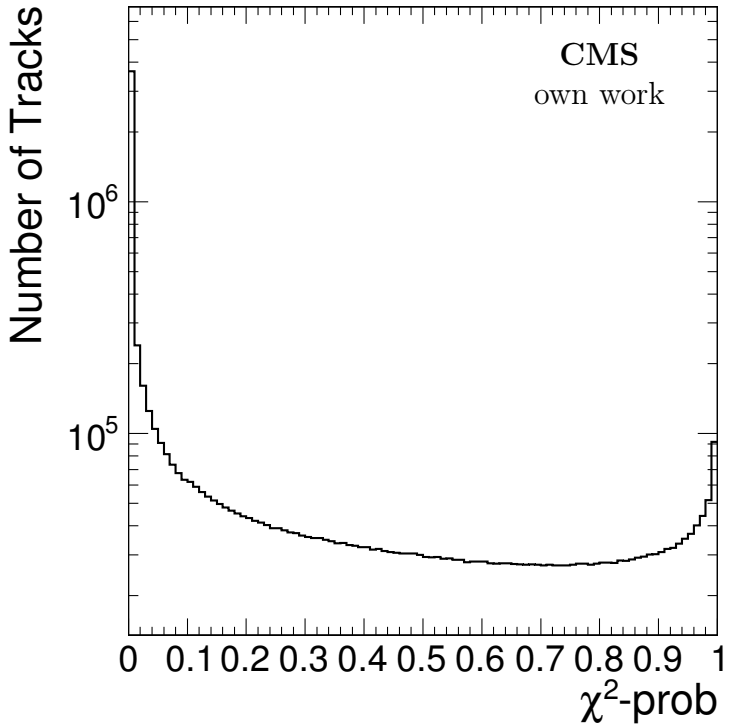


Figure 4.20: χ^2-Probability of all T2 tracks for events triggered by the CASTOR electromagnetic trigger and the TOTEM low multiplicity trigger.

The first step is to remove T2 tracks with a bad χ^2-fit with a standard selection cut on the χ^2-Probability of the fit. Only tracks with

$$\chi^2_{\text{Prob.}} >= 0.01 \tag{4.4}$$

are used further on. The distribution of the χ^2-Probability for all T2 tracks can be seen in figure 4.20.

Because of its geometrical position the tracker T2 sees a lot of particles coming from second interactions and radiation of the HF and endcap detectors of CMS, which are also reconstructed. These tracks are called secondary tracks whereas tracks from the interaction point (IP) are called primary tracks. To distinguish primary tracks from secondary tracks the z_{Impact}-parameter of the tracks is the strongest variable for separation (see figure 4.21). The z_{Impact}-parameter of a track is defined by the z value where the track intersects with the plane containing the nominal beam axis vector and which is defined by the normal vector pointing from the beam axis to the first hit of the T2 track (see figure 4.22). This definition of z_{Impact} was chosen by the TOTEM colleagues since it is numerically stable and not affected by systematic misalignments. The definition is very similar to the closest point

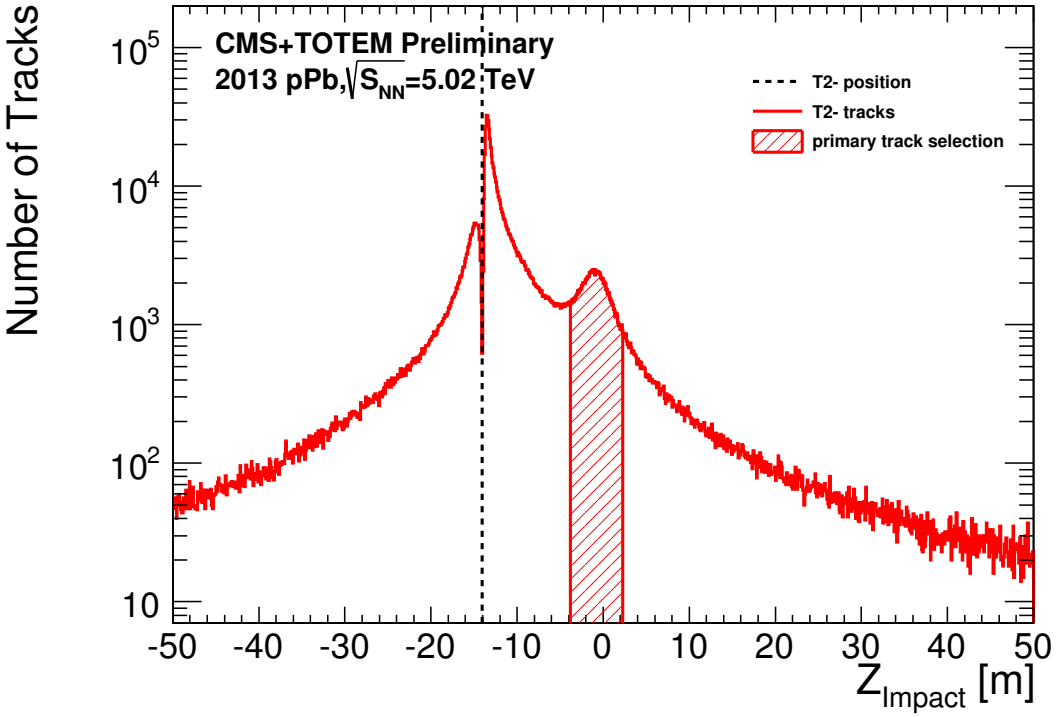


Figure 4.21: z_{Impact} distribution of T2 tracks from T2 tracker on the minus side of CMS. The shaded area shows the cut to select primary tracks. The dashed black line is the mean position of the T2 minus tracker in z.

of the T2 track to the beam axis but more stable to misalignments of the TOTEM tracker such as shifts and tilts. As seen in figure 4.21 the primary tracks show up in the peak close to 0 m while secondary tracks are much more spread with one strong peak around -14 m. This is also the mean position of the T2 tracker. Here the secondary tracks are almost perpendicular to the beam axis. These tracks are produced in the materials and surfaces surrounding the T2 detector. Since the tracker can only reconstruct tracks that hit at least four layers, the efficiency of reconstructing tracks drops as the tracks become more perpendicular to the z-axis. This shows up as the dip seen in the number of tracks for z_{Impact} resembling the T2 tracker position.

To select primary tracks the z_{Impact} distribution is fitted in the range of -10 m up to 10 m on both sides separately to estimate optimal cuts on z_{Impact}. For this purpose a sum of two Gaussian functions with the same mean value μ is used for the peak of primary tracks and a sum of two exponential functions to describe the background of secondary tracks. The two exponentials are needed to describe the different slopes left and right from the primary track peak. See also figure 4.23 for the fit of the z_{Impact} distribution of the T2 tracker on the negative side of CMS. The fit function has the following form:

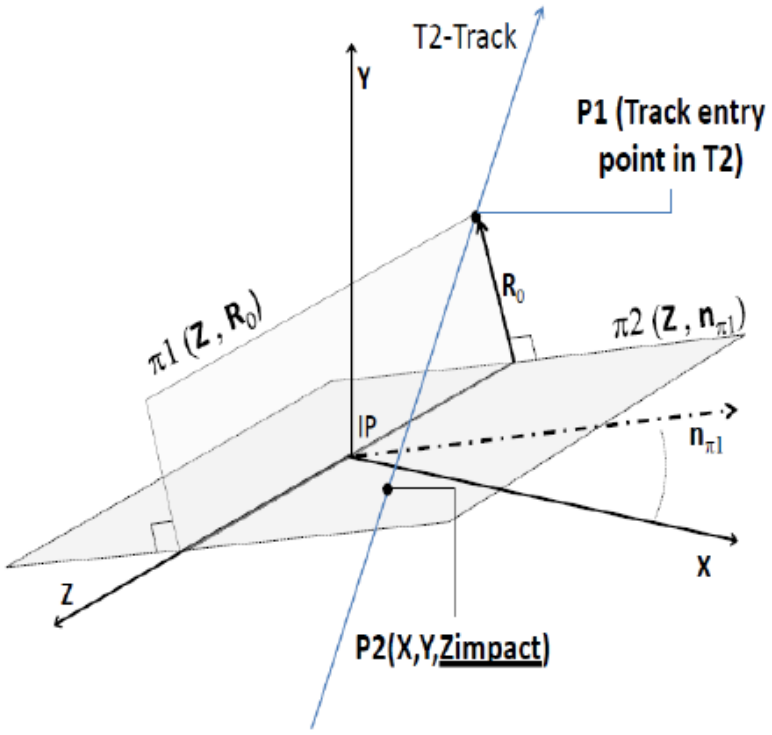


Figure 4.22: Scheme of the definition of the parameter z_{Impact} of T2 track [82].

$$f(z_{\text{Impact}}) = \frac{A_{\text{gauss},1}}{\sigma_1\sqrt{2\pi}} e^{-\frac{(z_{\text{Impact}}-\mu)^2}{2\sigma_1^2}} + \frac{A_{\text{gauss},2}}{\sigma_2\sqrt{2\pi}} e^{-\frac{(z_{\text{Impact}}-\mu)^2}{2\sigma_2^2}} + A_{\text{exp},1} e^{b_1 z_{\text{Impact}}} + A_{\text{exp},2} e^{b_2 z_{\text{Impact}}} \tag{4.5}$$

are used where σ_1 and $sigma_2$ as the width of the Guassian functions and b_1 as well as b_2 is the slope of the exponential function. The result of the fit can be seen in figure 4.23 and in table 4.3 where the numbers of primary tracks is given by sum $N = A_{\text{gauss},1} + A_{\text{gauss},2} = 8532$ due to the fact that the Gaussian part of the fitted function is normalized while to get the number of secondary tracks the exponential part has to be integrated over the correct range. A cut in z_{Impact} of

$$-3770\,\text{mm} \leq z_{\text{Impact}} \leq 2230\,\text{mm} \tag{4.6}$$

is chosen to select primary tracks which are around 2σ away from the fitted Gaussian peak position. With this cut the number of selected primary tracks is 6925 and 3926 selected secondary tracks.

Tracks after the selection steps (4.4) and (4.6) are considered primary T2 tracks. For the CASTOR alignment events which have exactly one primary track on the negative side of the T2 tracker are selected [82].

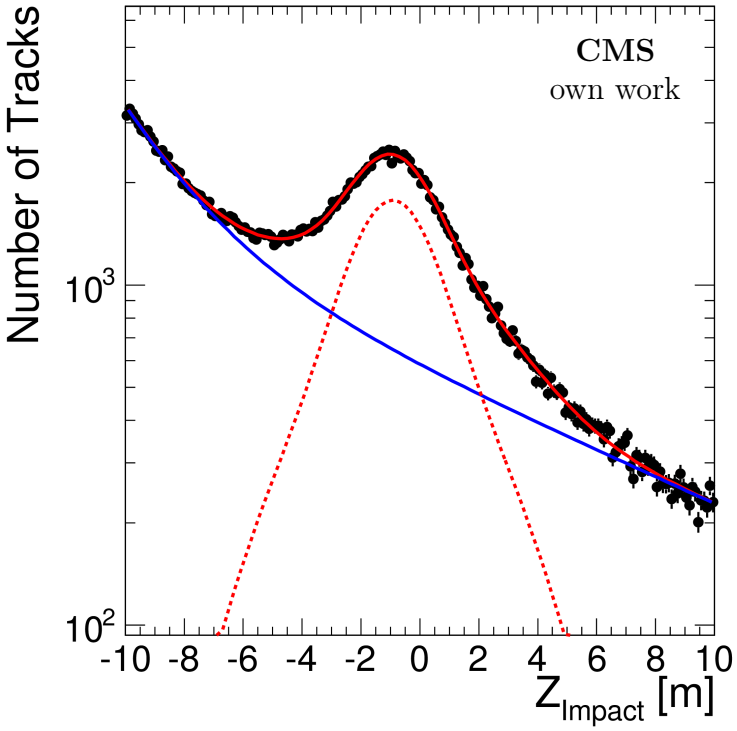


Figure 4.23: Result of the fit of the z_{Impact}-parameter for tracks from the negative side of T2. The solid red line shows the overall fit of the data. The black data points are for all tracks with a χ^2-probability ≥ 0.01 only including events with the common TOTEM-CASTOR trigger. The result for the pure signal of primary tracks estimated by the two Gaussian distributions is plotted as a dashed red line while the background of secondary tracks is plotted as a solid blue line.

Determination of an isolated sector in the CASTOR detector

Additionally, in CASTOR it is necessary to require energy deposit in one isolated electromagnetic sector. Since the data sample was collected with the common TOTEM-CASTOR trigger, it can be already assumed that there is no energy deposit in the hadronic part of CASTOR. Every signal in CASTOR is considered as electromagnetic response from electrons or gammas. To search for an isolated activated sector it is first checked that the hottest sector (with the highest energy) is above noise level, which is 5 GeV. Second, all other sectors except for one of the two neighbouring sectors to the hottest one have to be below noise level. The sector with the highest energy deposit is reconstructed as isolated sector from electromagnetic shower.

If in one event one single isolated sector is found in the CASTOR detector and also one single track in the T2 tracker, then the event will be classified as originally from one single electron hitting T2 and CASTOR. One consequences is that the position of the electron

$A_{\text{gauss},1}/10^3$	3.48 ± 0.18
$A_{\text{gauss},2}/10^3$	5.48 ± 0.20
μ/mm	-0.923 ± 0.013
σ_1/mm	1.28 ± 0.04
σ_2/mm	2.92 ± 0.09
$A_{\text{exp},1}$	26 ± 5
$A_{\text{exp},2}$	559 ± 6
$b_1/10^{-3}\,\text{mm}^{-1}$	-432 ± 13
$b_2/10^{-3}\,\text{mm}^{-1}$	-89.8 ± 2.7

Table 4.3: Result of fitting the z_{Impact} distribution with a Gaussian plus exponentiation term.

in the detectors is only sensitive to the geometry of the electromagnetic (front) part of the CASTOR detector. It is not directly sensitive to the alignment of the hadronic (back) part of the CASTOR detector.

Extrapolation of tracks to the calorimeter surface

To determine the position where the tracks hit the calorimeter the last track hit point in the T2 tracker is used as start to extrapolate the track geometry to the CASTOR front face, which is defined as $z_{\text{CASTOR}} = -14385\,\text{mm}$.

The ϕ direction of each track is taken from the χ^2-track-fit while the angle η is calculated directly as the mean of the η values from all hits in the tracker associated to this track. In TOTEM terminology this is also called η_2. For primary tracks from the IP this value is more precise than the fitted angle.

4.4.2 Alignment by fitting the detector surface

In the following, using events with an activated isolated sector in CASTOR and an associated single primary track from the T2 tracker on CASTOR side, the geometry of the front face of the CASTOR detector can be seen as shown in figure 4.24. In figure 4.24 every hit from a T2 track in the detector is coloured by one activated sector. Starting with the blue marked track hits for sector 1 ($0 \leq \phi \leq 360°/16$) the track hits are alternately coloured for all following sectors in red, black or orange.

Fit procedure

The idea is to find the optimal position of the front face of the calorimeter by moving it until the number of track hits inside the corresponding sector area is maximized. For the

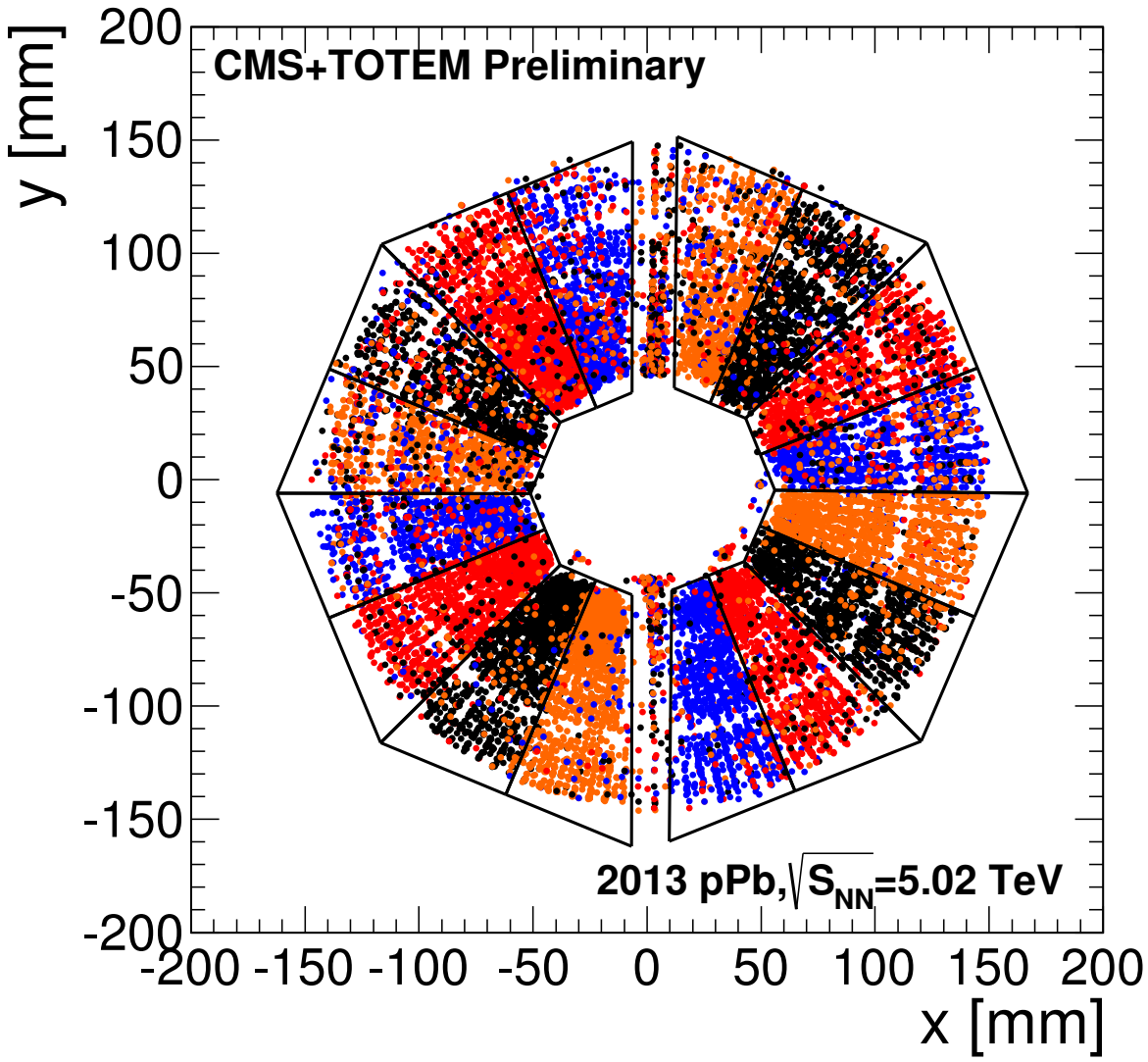


Figure 4.24: Hit position of T2 tracks extrapolated to CASTOR front Z position. Track hits are coloured by the isolated activated sector in CASTOR. Alternating colouring starting from sector one. The solid lines show the fitted position of the front shape (the EM part) of CASTOR. [83]

fit the minimizer package MINUIT [84] is used, which provides stable algorithms to find the minimum of a multi-dimensional function. MINUIT estimates the parameter errors by searching for the parameter values where the cost function (idially a χ^2-function) has a difference of 1 to the minimum.

Since MINUIT is always searching for the minimum, the number of track hits inside the sector area is inverted to define the fit-function. Hence the pseudo $\chi^2(\vec{\theta})$-function, which is minimized depending on the parameter vector $\vec{\theta}$ separately defining the position of the

two CASTOR halves, is

$$\chi^2(\vec{\theta}) = \sum_{i=1}^{\text{sec}} \frac{(1 - N^i_{in}(\vec{\theta})/N^i_{tot})^2}{\sigma\left(\frac{N^i_{in}(\vec{\theta})}{N^i_{tot}}\right)_{stat.}} \tag{4.7}$$

where $\sum_{i=1}^{\text{sec}}$ is the sum over all 16 CASTOR sectors in ϕ and $\sigma(N_{in}(\vec{\theta})/N_{tot})_{stat.}$ is the error of $N_{in}(\vec{\theta})/N_{tot}$ calculated by normal propagation of the uncertainties from $N_{in}(\vec{\theta})$ and N_{tot} which are Poisson distributed. It is called pseudo-χ^2 because the underlying distribution are net Gaussian. Not only are the statistical errors important, but also the uncertainties on the track parameters and consequentially also the hit position of the track in CASTOR. Additionally, even in events with one good track and one isolated sector, the track may not correspond to the sector. A more stable and solid estimation of the uncertainties of the fit results is described in the following section.

A second issue is the fact that this χ^2 is a stepwise function. When moving the position of the detector the hit of one track is inside the sector area or drops out. MINUIT works with the derivative of the functions, which it appoximates numerrically, so the stepping becomes a problem in the minimizer algorithm.

To avoid this N_{in} is not directly calculated by the number of hits inside the area. It is assumed that the hit position has a Gaussian smearing related to the track uncertainties in the TOTEM tracker of $\Delta\phi = 9.435\,\text{mrad}$ and $\Delta\eta = 0.0175$. Assuming now a hit as a normalized Gaussian function with mean of the hit position and a width of the tracker resolution of T2, N_{in} is now calculated as

$$N^{\text{sector}}_{in} = \sum_{\text{hits}} \int_{A_{\text{sector}}} g_{\text{hit}}(\vec{r}) d\vec{r} \tag{4.8}$$

as the sum of the integrals over the Gaussian function

$$g_{\text{hit}}(\vec{r}) = \frac{1}{\sqrt{2\pi}\sigma_{\text{hit}}} \exp\left(-\frac{(\vec{r} - \vec{r}_{\text{hit}})^2}{2\sigma^2_{\text{hit}}}\right) \tag{4.9}$$

inside the sector area. The integral is numerically not trivial because of the shape of the sectors, so it is estimated by using

$$\int_{A_{\text{sector}}} g_{\text{hit}}(\vec{r}) d\vec{r} \simeq \int_{r_{min}}^{r_{max}} g_{\text{hit}}(r) dr \times \int_{r\phi_{min}}^{r\phi_{max}} g_{\text{hit}}(r\phi) d(r\phi) \tag{4.10}$$

where the integral is factorized into two integrals of radius and phi which is quite near the correct value as long as the σ is small against the area size and the borders of the sectors are nearly orthogonal to each other. Both conditions are fulfilled in this case.

With eqs. (4.7), (4.8), (4.9) and (4.10) MINUIT can now work with a smooth function to find the optimal position where the number of track hits in each sector is maximized. For this purpose MINUIT is fitting the position parameters $\vec{\theta}$, which are an x- and a y-shift

as well as rotation in ϕ, for the two CASTOR halves separately (on the near side and far side according to the accelerator ring center). The final result of the detector position is shown in figure 4.24, where the left half is the near side and the right half the far side of the detector; it is also put into table 4.6 which is also corrected for the known shift of the tracker T2. The pure fit result without any correction (here TOTEM tracker alignment) and with uncertainties from MINUIT, which, for the previously mentioned reasons, are not really reliable, are shown in table 4.4.

	CASTOR near-side	CASTOR far-side
$x-$shift/mm	11.40 ± 0.30	-6.59 ± 0.26
$y-$shift/mm	-4.1 ± 0.6	-6.3 ± 0.5
$\delta\phi$/deg	-0.66 ± 0.30	-0.10 ± 0.26

Table 4.4: Result of the fit with errors quoted by MINUIT. A positive $\delta\phi$ is defined by a rotation in positive nominal ϕ direction.

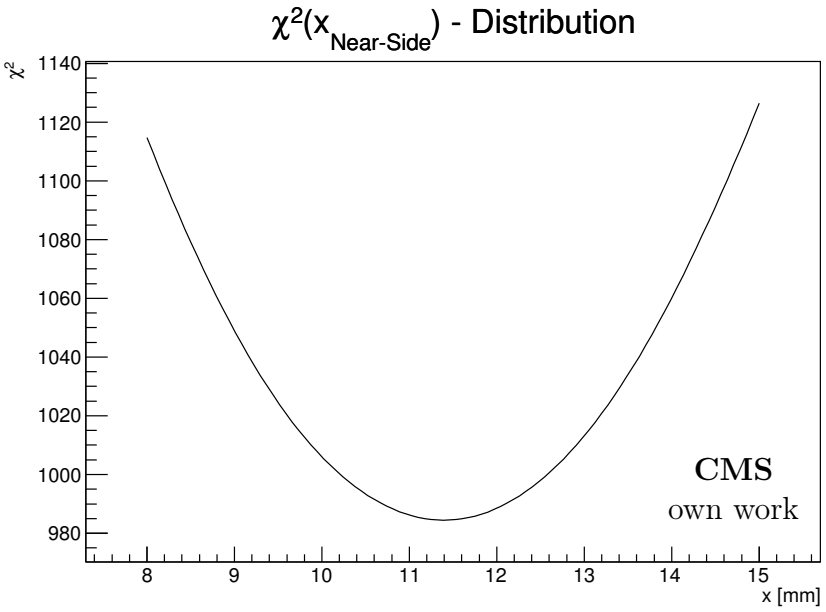


Figure 4.25: Scan of χ^2 function (4.7) depending on the x-shift parameter of the detector near side. All other parameters are set to the values where χ^2 has its minimum (see table 4.4).

As a cross check of the stability of the fit and to verify that the χ^2 function (4.7) shows no unexpected behaviour around the minimum, the χ^2 is plotted for the different parameters. In all cases it shows a smooth quadratically shape like in figure 4.25.

Uncertainty estimation via bootstrapping

As said before the minimized χ^2-function does not follow a real χ^2-distribution. Therefore we need another way of estimating the uncertainty of the fit result. This is done using a "bootstrap" method.

Bootstrapping is a method that randomly generates new data samples and repeats the fit procedure on these samples. The randomly generated samples should follow the true distribution of data. But this is unknown, so the bootstrap method resamples from real measured events, which by definition represents the true distribution.

The purpose is not to produce again exactly a same sample as in the data but to reproduce the most relevant features and distributions of the data. Due to the acceptance and symmetry of the T2 tracker and the CASTOR detector (as seen in figure 4.24) it is recommended to use cylindrical coordinates with a small correction. To get a good reproduction of the distribution, especially the ring of low multiplicity also seen in figure 4.24, which is originated by additional material from the beam pipe in front of the T2 tracker, we use ρ in the nominal CMS system at $(0, 0)$. The distribution of track hits in ρ can be seen in the top plot of figure 4.26.

Because the fit is very sensitive to borders of the sectors, thus to the change of the multiplicity of track hits in one sector when crossing the border of this sector, the nominal ϕ as variable is not advisable. Instead, the system is moved in the system of each of the CASTOR halves which are shifted and rotated by the optimal values found in the minimization process (values in table 4.4). Furthermore, only the distance in ϕ from the border of the activated sector is considered. Since the CASTOR surface is divided in 8 ϕ-symmetrical octants, and is not symmetric for all 16 sectors, it is also taken into account if the sector has a number which is odd or even. The distribution in ϕ corresponding to the sector coordinate system is shown on the lower side of figure 4.26 in blue for odd sectors and red for even ones.

In the next step for each sector a number of randomly positioned track hits is genrated with the same statistic as this sector has in the real data. The ρ and ϕ of each track hit are chosen randomly, weighted with the distributions in figure 4.26. An example of such a generated track hit distribution at CASTOR surface can be seen in figure 4.27.

	CASTOR near-side		CASTOR far-side	
	MINUIT	bootstrap	MINUIT	bootstrap
$\sigma_{x-\text{shift}}$/mm	0.30	0.14	0.26	0.13
$\sigma_{y-\text{shift}}$/mm	0.6	0.23	0.5	0.27
$\sigma_{\delta\phi}$/deg	0.30	0.13	0.26	0.13

Table 4.5: Comparison between the uncertainties given by MINUIT from and the uncertainties given by the variation of fit results in the bootstrap technique.

The technique of generating random track hits in CASTOR and fitting the detector surface as described before is repeated one thousand times. Each fit returns a new fit parameter results. The distribution of the fit parameters is shown in figure 4.28. The variance of the fit result is the excepted uncertainty of the complete data sample fit shown in figure 4.24. In table 4.5 the fit uncertainties from bootstrapping and MINUIT are compared. The mean values of the fits are slightly shifted with respect to the results from table 4.4. This is a typical feature and is explained by an imperfect description of the data by using the distributions in figure 4.26 to randomly generate new track hits in CASTOR.

Systematic uncertainties on the CASTOR position

With an estimation of the statistical uncertainties of the fit there is still the question of systematic effects. The first check is the dependence of the fit on the data selection criteria such as energy deposit or the allowed energy ratio between two neighbouring sectors. Secondly, the dependence on internal parameters such as the z-position of CASTOR in the experiment and the smearing of the track hits used in (4.9) is evaluated. And finally the uncertainty of the alignment of the T2 tracker is considered as well.

For the energy dependence only the energy deposit of the main activated sector is checked. The track hits are sorted by the energy of the corresponding sector and divided equally into five energy bins. Figure 4.29 shows for all events the energy deposit in this isolated sectors. The fit and the bootstrapping procedure is then done for every data sample of each energy bin. The resulting values and uncertainties are shown in the first column of figure 4.30 and figure 4.31, where the error bars represent the uncertainty from bootstrapping.

No significant systematic effects have been identified in this approach. The maximal variation found in figure 4.30 and figure 4.31 is added to the systematic uncertainties. The final set of systematic uncertainties is summarized in table 4.6. This is the first, and so far the only precise measurement of the alignment of CASTOR in CMS.

	CASTOR near-side	CASTOR far-side
$x-$ shift/mm	$13.4 \pm 0.1(\text{stat.}) \pm 0.8(\text{syst.})$	$-4.6 \pm 0.1(\text{stat.}) \pm 0.8(\text{syst.})$
$y-$ shift/mm	$-5.4 \pm 0.2(\text{stat.}) \pm 0.9(\text{syst.})$	$-7.6 \pm 0.3(\text{stat.}) \pm 0.9(\text{syst.})$
$\delta\phi$/deg	$-0.7 \pm 0.1(\text{stat.}) \pm 1.3(\text{syst.})$	$-0.1 \pm 0.1(\text{stat.}) \pm 1.3(\text{syst.})$

Table 4.6: Final values of the CASTOR position for the two detector halves where a positive $\delta\phi$ is defined by a rotation in positive nominal ϕ direction.

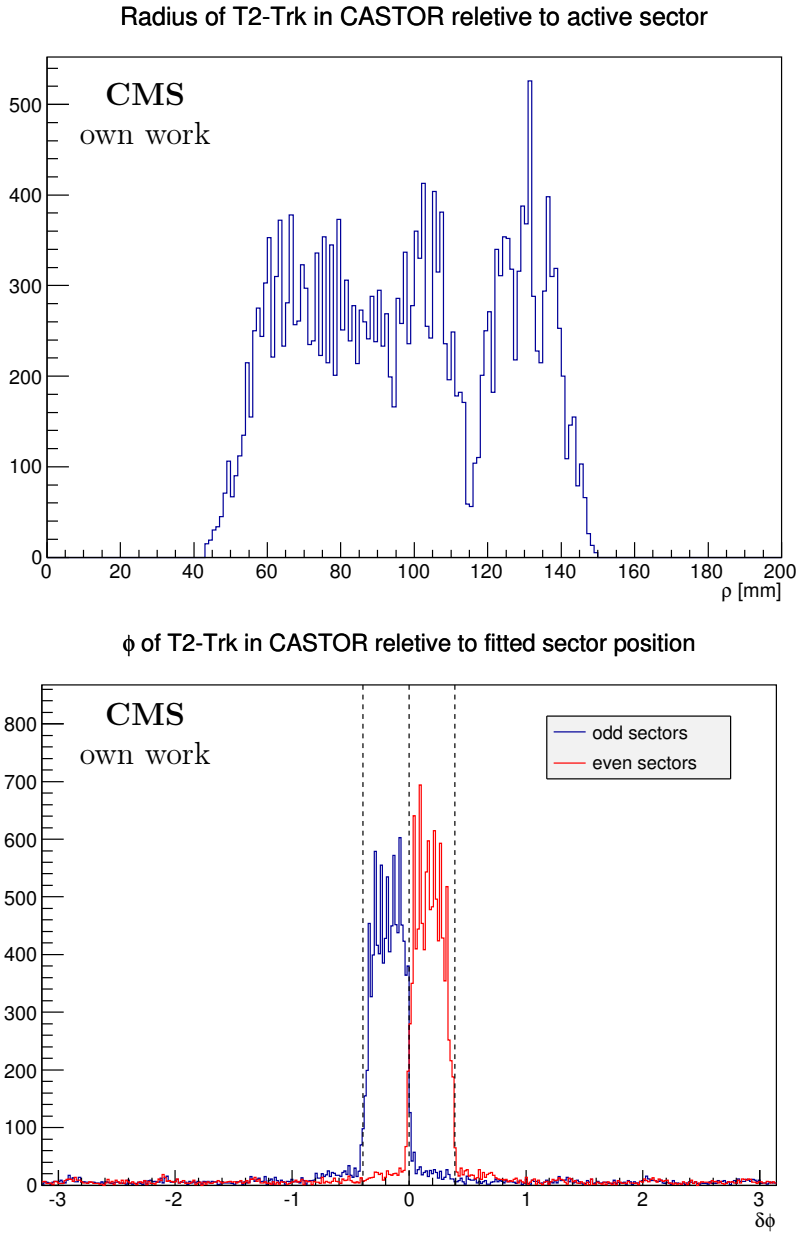


Figure 4.26: The top plot shows the distribution of track hits in absolute ρ. The lower plot shows it in ϕ relative to the center of the CASTOR octant, which is hit by the track.

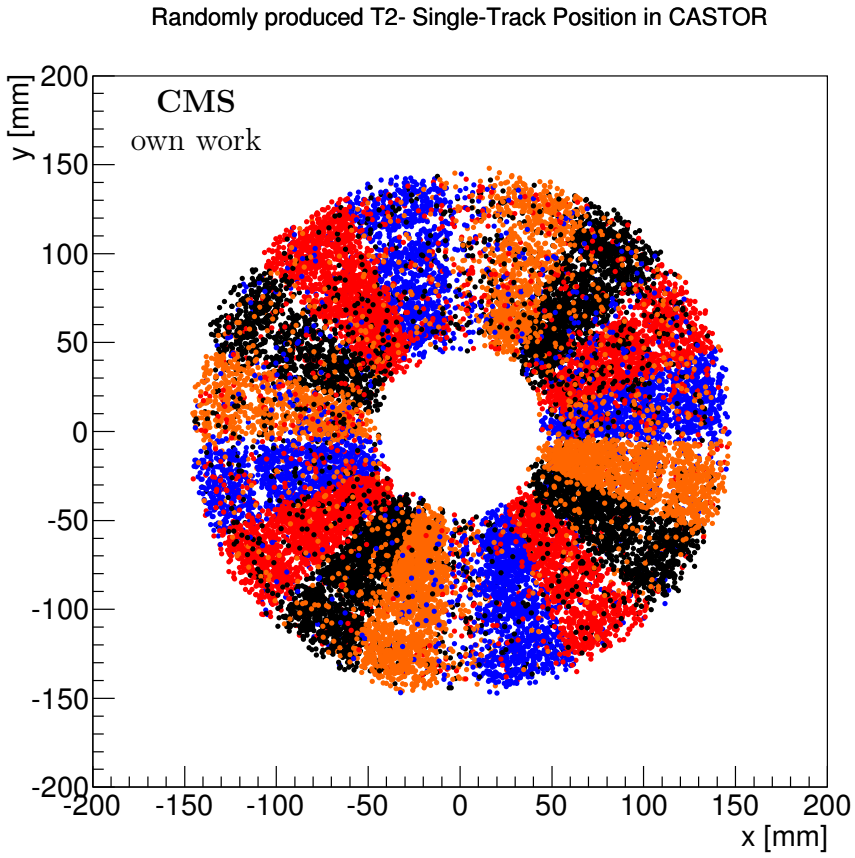


Figure 4.27: Randomly generated T2 track hit positions in CASTOR via ρ and ϕ distribution in figure 4.26

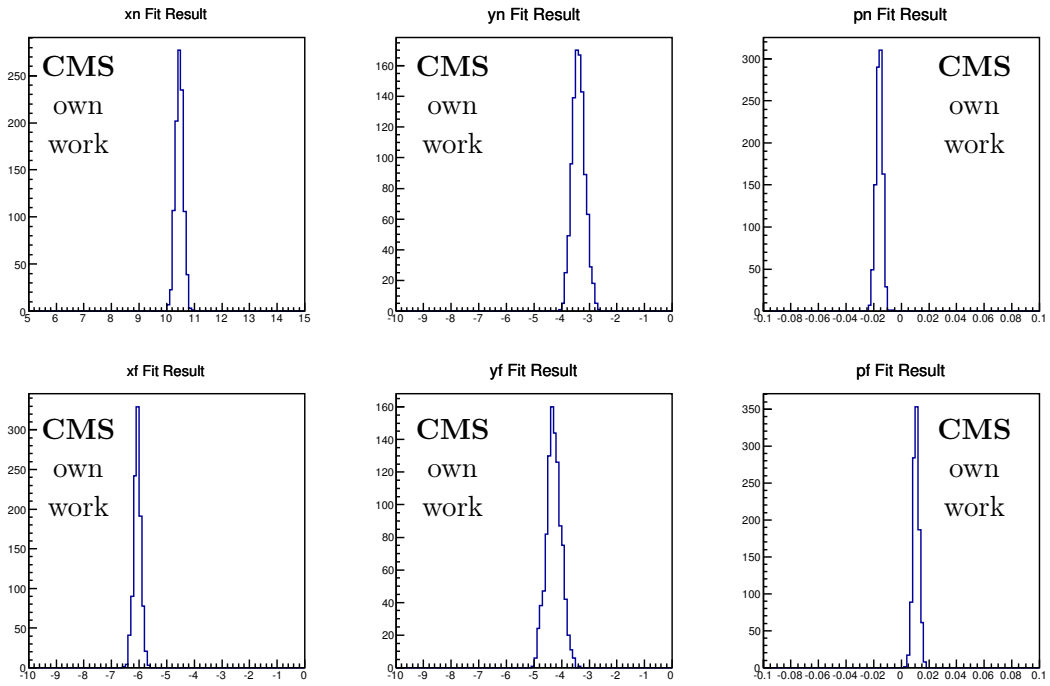


Figure 4.28: Distribution of fit results which originate from the bootstrapping procedure. In the top row the plots xn,yn & pn correspond to the x-shift,y-shift & ϕ rotation of the near side of CASTOR while in the bottom row the plots xf,yf & pf correspond to the x-shift,y-shift & ϕ rotation of the far side of CASTOR. On the y-axis the number of fits is counted. The units of x and y are mm and of ϕ the unit is rad.

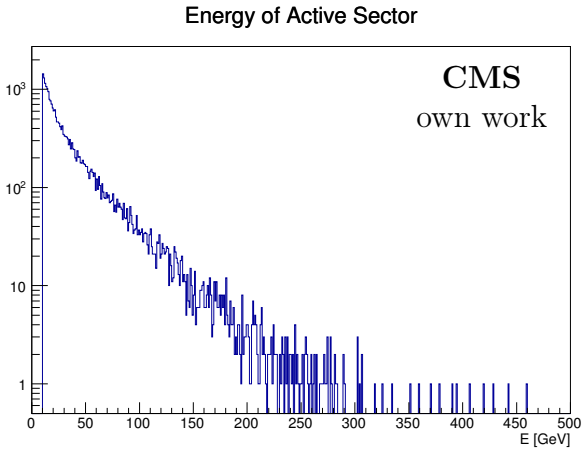


Figure 4.29: Energy distribution of isolated activated sectors in CASTOR with one corresponding TOTEM track.

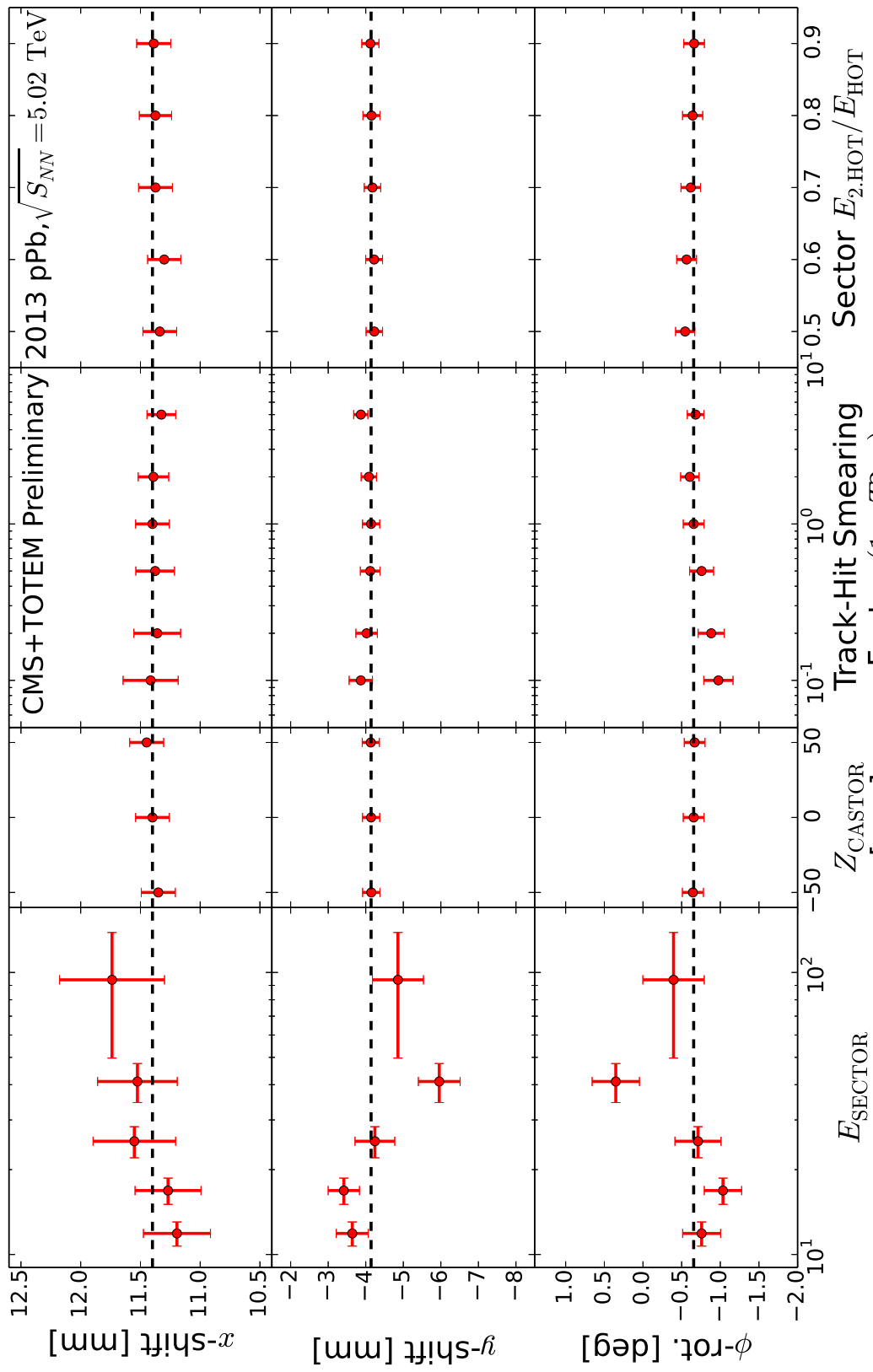


Figure 4.30: Figure shows results of the positioning fit of the CASTOR half closer to the LHC ring center for different data samples.

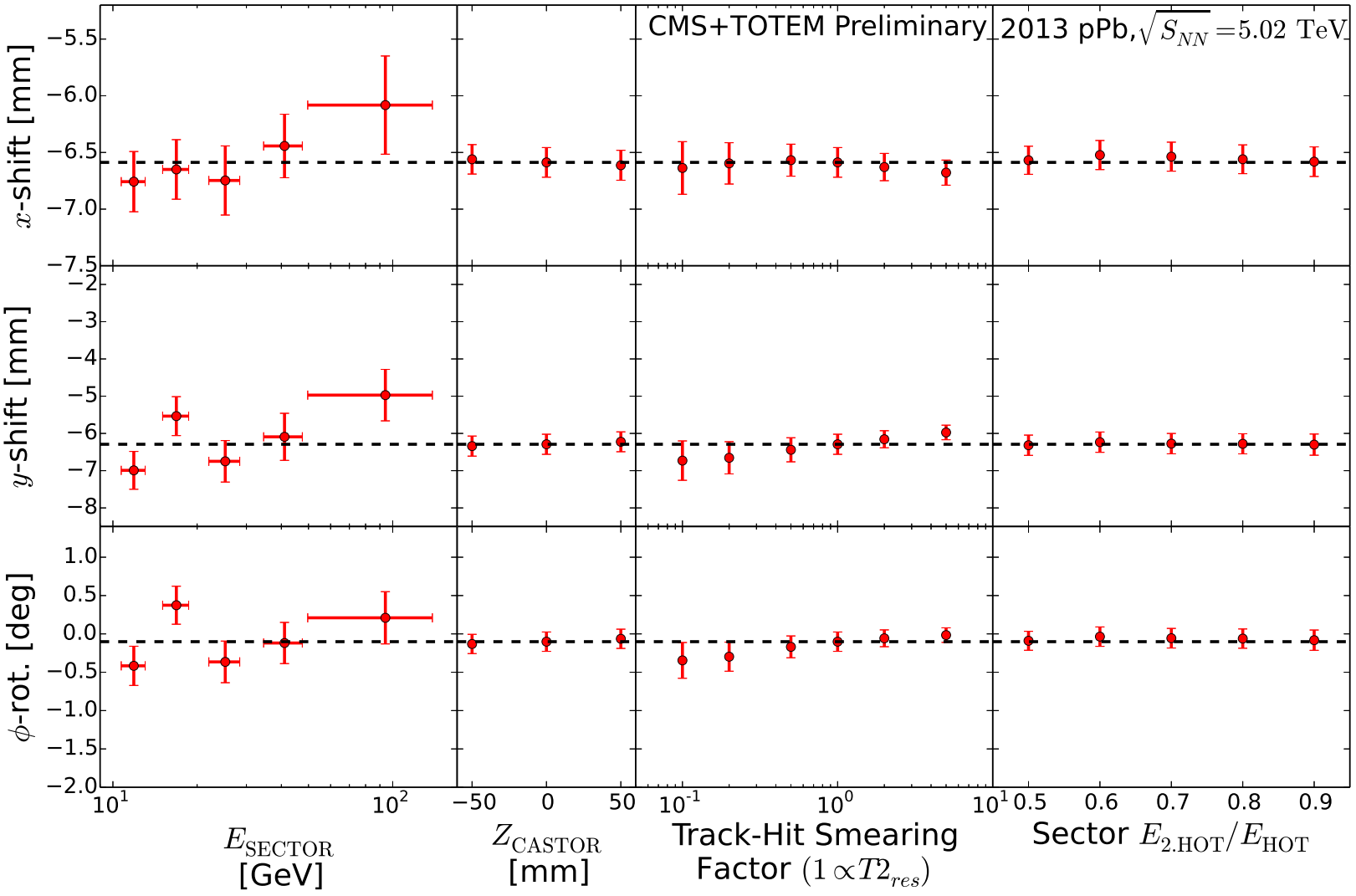


Figure 4.31: Figure shows results of the positioning fit of the CASTOR half farther from the LHC ring center for different data samples.

5 Forward jet measurement

5.1 Jet reconstruction

The description of LHC physics is for the most part based on QCD calculation in terms of quark and gluon scattering. However, quarks and gluons are not directly observable particles. After hadronization they produce a collimated bunch of hadrons, a so-called jet, which conserves the direction, energy and quantum numbers of the parton. These hadrons can then be observed by the experiments at LHC as well as CMS.

Therefore, a stable jet algorithm is needed to regain information about the involved partons. At generator level this algorithm clusters final state particles into a jet and at detector level it uses position and energy deposit of calorimeter cells or hits in the tracker. The jets, which are computed by the algorithm, should be stable in number and orientation against soft emissions of the parton (infrared-safe) and collinear splitting of the parton into two partons (collinear-safe) [85]. These requirements are very important for the jets to be insensitive to the underlying event and pileup.

There are two major kinds of jet algorithms: the iterative cone (e.g. [86]) and the sequential clustering algorithms [87–89]. Iterative cone algorithms build up cones around the constitutions with the highest p_{T} to create jets. Typically these algorithms are not collinear and not infrared safe except for the seedless infrared-safe cone (SIS) algorithm [90]. On the other hand clustering algorithms are colinear and infrared safe by construction. They are based on a distance measurement between the constitutions to cluster them until a breakup condition is fulfilled. In the LHC experiments the sequential clustering algorithms, in particular anti-k_t as well as SIS, are commonly used. In this very forward jet measurement the jets are defined by the anti-k_t algorithm.

5.1.1 The anti-k_t jet clustering algorithm

As a special case of the sequential cluster algorithms, the anti-k_t algorithm also fulfils the important requirement of being infrared and collinear-safe. In general the sequential cluster algorithms, which are characterized by the integer like parameter p, combine particles into jets by introducing the distance measure between the particles i and j as

$$d_{ij} = \min\left(k_{ti}^{2p}, k_{tj}^{2p}\right) \frac{R_{ij}^2}{R^2} \tag{5.1}$$

for a jet of cone-size R and with the radial distance betwean the particles

$$R_{ij} = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}, \tag{5.2}$$

where k_{ti}, η_j and ϕ_i are the transverse momentum, rapidity and azimuth of the particle i. The distance of the particles to the beam is then

$$d_{iB} = k_{ti}^{2p} \tag{5.3}$$

The cluster algorithm is looking for the minimal distance of all d_{ij} and d_{iB}. If d_{ij} is the minimum, then the particles i and j are merged into a single object i, which is then treated as new particle. Otherwise, if d_{iB} is the smallest, the object i is considered as jet and removed from the list of generated particles. This procedure is repeated on the remaining particles until no particle is left. Note that not necessarily particles in the strict sense are meant but every object on detector or hadron level where the main parameter as transverse momentum and orientation can be determined.

In eqs. (5.1) and (5.3) the size of the jets is defined by the radius parameter R while the parameter p governs the relative power of the distance versus the energy and determines different subclasses of clustering algorithms. For the parameter value $p = 1$ this is the k_t clustering algorithm [87] while for $p = 0$ the procedure corresponds to the Cambridge/Aachen algorithm [88]. In this analysis the anti-k_t algorithm is used as jet finding procedure, which corresponds to a negative value of $p = -1$.

One major difference between the anti-k_t and the other cluster algorithm is that the anti-k_t algorithm creates a regular cone like shape for the jet area while the others generate jets with very irregular shapes [91]. The regular cone like shape created by the anti-k_t algorithm is generally not influenced by the soft particles in the event. Especially the anti-k_t algorithm cannot create ghost jets by soft particles as it is possible in the k_t algorithm. This is reasonable since the k_t algorithm starts with the softest particles in (5.1) to cluster jets. The anti-k_t, on the other hand, starts with the highest p_T particles for the jet clustering. In terms of jet energy calibration or pile up subtraction a stable cone shape generation of the jet area is favourable. For this reasons as well, the anti-k_t algorithm is commonly used for jet reconstruction at CMS, but other jet finding algorithms are supported, too.

5.1.2 Hadron level jets

To compare data with MC simulations the anti-k_t algorithm with the same radius parameter is applied simultaneously at hadron and detector level. At hadron level all final state particles from the MC generator are clustered to jets except for particles beyond the standard model. Final state particles at generator level with a mean decay length of $c\tau > 1\,\mathrm{cm}$ are considered to be stable. These particles are also processed with the detector simulation software GEANT4 [92] to simulate their response in the CMS detector.

For the anti-k_t algorithm a radius parameter of $R = 0.5$ is chosen for the clustering of the very forward jets. Also in the central rapidity region of CMS this is typically the preferred parameter value [93]. However, in data taking at high luminosities smaller values of R can be an advantage since they reduce the impact of pileup. In the reconstruction software of jets the implementation of the anti-k_t algorithm in the software tool FASTJET [94, 95] is used. The energy and the direction of a resulting jet is the four-vector sum of all constituents that are clustered in the corresponding jet.

5.1.3 Detector level jets

For the reconstruction of jets in the CASTOR detector the CASTOR towers are used as fundamental entity for the anti-k_t algorithm, replacing the generator particles at hadron level. A tower is the sum of the inter/calibrated energy of all good channels in one sector. Channels which show a bad behaviour or are simply dead during data recoding are marked as bad in the database. These channels are not used during reconstruction at all. If the tower energy E_{tow} is below the noise level of $E_{\text{tow}} < \sqrt{N_{\text{ch}}} E_{\text{noise}}$ where N_{ch} is the number of active channels in the tower and $E_{\text{noise}} = 0.65\,\text{GeV}$ is the noise level of a single channel, then the tower is discarded.

Additionally to the energy the direction of the tower in ϕ is given by the geometric center of the tower, while the tower η is given by the energy weighted η-mean of the channels in the tower. But because of the nonexistent η segmentation of CASTOR detector the η values of the channels and therefore also of the towers are always close to $\eta \approx 6$. With this information the towers of CASTOR are clustered into jets. As in the previous section the anti-k_t algorithm as it is implemented in FASTJET is used with a radius parameter of $R = 0.5$. The computed jets are the four-vector sum of the clustered CASTOR towers.

The typical shape of a reconstructed detector level jet in CASTOR is shown in figure 5.1 compared to different generator samples. For this purpose events with a single reconstructed jet above $p_{\text{T}} > 3\,\text{GeV}$ in CASTOR are taken and the proportional tower energy of a jet depending on the distance to the jet center is shown. It demonstrates that most of the jet energy is inside the $R = 0.5$ cone marked by the two black dashed lines and thus, this radius parameter is a good choice for the reconstruction. Separated from the jet the data shows some structure around the noise level which is a relict of the segmented structure of the CASTOR detector, which is well produced by the MC samples. As seen in the ratio plot of figure 5.1 the MC energy shape is fluctuating around the measured one by $10\,\% - 20\,\%$. The different models describe different aspects of the data, while it seems that for larger ϕ-distances EPOS describes the data best.

Due to the calorimeter design, especially the segmentation in beam direction z can provide additional properties that can be used to characterize a CASTOR jet. In the first place, one can differentiate between the electromagnetic energy E_{EM} and the hadronic energy E_{HAD} in a jet where E_{EM} is the sum over all channels in the first two modules and E_{HAD}

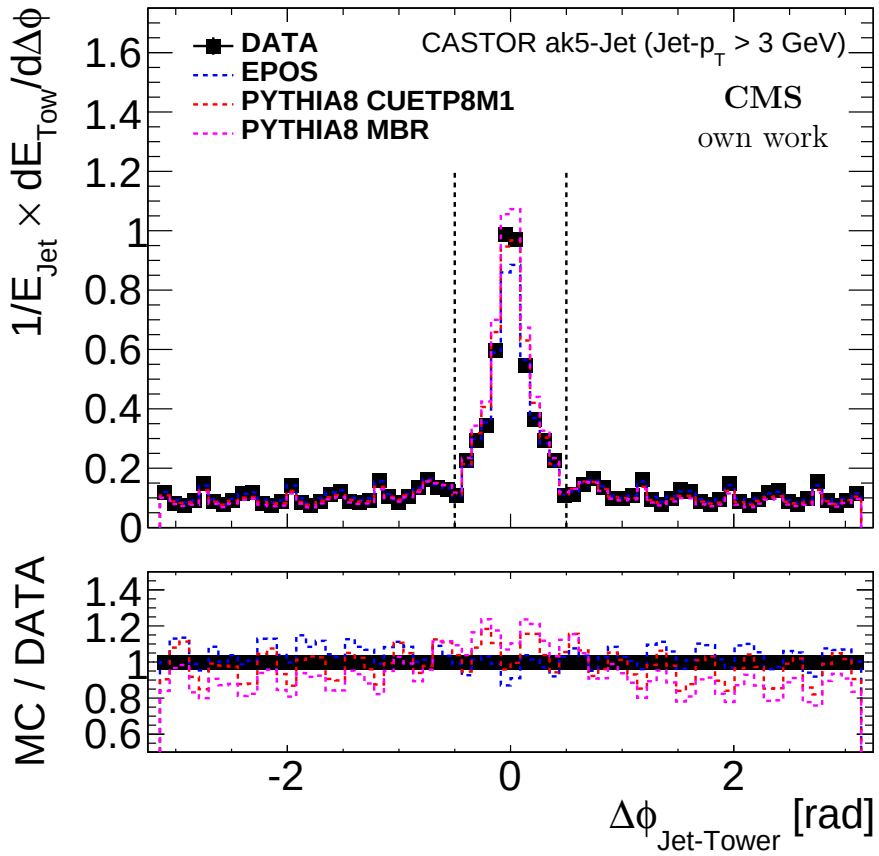


Figure 5.1: Average proportion of the CASTOR tower energy inside a jet depending on the $\Delta\Phi$ distance to the jet center. The two black dashed lines correspond to the jet radius $R = 0.5$.

is the sum over all channels in the modules 3 to 14, which are contributing to the jet. The ratio of the electromagnetic to the total energy E_{tot} in a jet is defined as

$$\text{fem} = E_{\text{EM}}/E_{\text{tot}} \tag{5.4}$$

and the ratio of the energy of the hottest channel E_{hot} clustered by the towers into a jet to the total jet energy is defined as

$$\text{fhot} = E_{\text{hot}}/E_{\text{tot}}. \tag{5.5}$$

Furthermore, using the z-position of a channel z_{ch} and the azimuth of a tower ϕ_{tower} the following additional quantities are derived

$$\sigma_\phi = \left[\frac{1}{E_{\text{tot}}} \sum_{\substack{\text{tower} \\ \in \text{jet}}} E_{\text{tower}} \left(\phi_{\text{tower}} - \phi_{\text{jet}}\right)^2 \right]^{\frac{1}{2}}, \tag{5.6}$$

which is the width in ϕ, the average longitudinal depth

$$\langle z \rangle = \frac{1}{E_{\text{tot}}} \sum_{\substack{\text{channel} \\ \in \text{jet}}} z_{\text{ch}} \tag{5.7}$$

and the longitudinal width

$$\sigma_z = \left[\frac{1}{E_{\text{tot}}} \sum_{\substack{\text{channel} \\ \in \text{jet}}} (z_{\text{ch}} - \langle z \rangle)^2 \right]^{\frac{1}{2}} \tag{5.8}$$

can also be defined for a jet in the CASTOR detector. These jet properties are later useful to compare the jet reconstruction between MC and normal data at detector level.

5.2 Data acquisition

For the jet analysis the recorded CMS data of proton-proton collisions at center of mass energy of $\sqrt{s} = 13\,\text{TeV}$ at LHC in 2015 with a total integrated luminosity of $211\,\mu\text{b}^{-1}$ is used. During data taking the condition of the LHC beam as the number of colliding bunches per beam cycle and the instantaneous luminosity, can vary significantly. In particular, the analyzed data was recorded during the run 247637 with a pile-up, or interaction probability, at 0.06. Besides, during the relevant data taking period the magnet field in CMS was at zero tesla.

dataset	run	events
`/ZeroBias1/Run2015A-PromptReco-v1/RECO`	247637	2586719
`/ZeroBias2/Run2015A-PromptReco-v1/RECO`	247637	2609851
`/ZeroBias3/Run2015A-PromptReco-v1/RECO`	247637	2645131
`/ZeroBias4/Run2015A-PromptReco-v1/RECO`	247637	2645047
`/ZeroBias5/Run2015A-PromptReco-v1/RECO`	247637	2644913
`/ZeroBias6/Run2015A-PromptReco-v1/RECO`	247637	2581011
`/ZeroBias7/Run2015A-PromptReco-v1/RECO`	247637	2645209
`/ZeroBias8/Run2015A-PromptReco-v1/RECO`	247637	2645174
in total		21003055

Table 5.1: List of data sets used for the analysis of the very forward jet spectrum.

The selected events from the data sets are shown in table 5.1 and triggered by the unbiased trigger. For the unbiased trigger the prescale was 137 at L1 trigger level and 12 at HLT level. Since all eight datasets recorded the events independently, the effective prescale of all the triggered events is $208.5 = 137(\text{L1}) \times 12(\text{HLT})/8$. The CMS sofware release

CMSSW_7_4_4_patch1 is used for the reconstruction of the data. The parameters for the reconstruction stored in the database are accessed via the global tag GR_P_V55, which is a database tag that describes the exact detector condition and calibration for a specific range of runs. The global tag GR_P_V55 includes special information for 0 T field as well as the latest bad channel list and intercalibration constants of the CASTOR detector. The energy scale of CASTOR as the intercalibration of the channels in the detector is based on the latest results from LHC beam-halo muon response and LED test pulses.

With the aim of correcting the measured jet spectra to hadron level several dedicated Monte Carlo samples are produced. Also for the estimation of important systematic uncertainties, specific MC samples are generated with CMSSW using the release CMSSW_7_1_20_patch3 and the global tag MCRUN2_71_V1 with a realistic distribution of the beam spot position. A list of the generated Monte Carlo samples can be seen in table 5.2. These samples are then reconstructed with CMSSW_7_6_3_patch2 release and the specific global tag 76X_mcRun2_asymptotic_v14 with an updated bad channel list for CASTOR. An identical list of bad channels observed during data taking is removed from the reconstruction of the Monte Carlo and the physics data samples.

MC generator	magnetic field	events	CASTOR position
Pythia8 Tune MBR	3.8T	998920	nominal
Pythia8 Tune MBR	0T	997146	nominal
Pythia8 Tune MBR	0T	4862000	as measured
Pythia8 Tune CUETP8M1	3.8T	999330	nominal
Pythia8 Tune CUETP8M1	0T	4917500	as measured
Pythia8 Tune CUETP8M1	0T	960000	measured + systematic shift
Pythia8 Tune CUETP8M1	0T	980000	measured - systematic shift
EPOS-LHC	0T	4978400	measured

Table 5.2: Summary of Monte Carlo samples with some of the main parameters.

Because of a zero Tesla field in CMS during data taking it is especially important to have Monte Carlo samples with no magnetic field with higher statistic. Due to the position of the CASTOR detector in the forward rapidity space it is also important to simulate the events with the measured position of the detector and not the nominal position that is symmetric around the beam pipe and a realistic beamspot position. Next to the samples discussed here several additional productions are used for the validation of the impact of the detector position and the magnetic field on the final hadron level corrected jet spectrum.

5.3 Jet energy calibration

Since CASTOR is not a self-compensating calorimeter, the energy deposits of the jets must be corrected for non-compensation effects. The response on hadrons in the calorimeter is reduced compared to the one on electrons. This was tested with data taken during a test beam setup [78] with electrons and pions. Therefore a leading order calibration of the reconstructed jet p_{T} is derived from Monte Carlo simulations by determining the relation between $p_{\mathrm{T}}^{\mathrm{GEN}}$ and $p_{\mathrm{T}}^{\mathrm{RECO}}$ in terms of $p_{\mathrm{T}}^{\mathrm{CAL}} = f\left(p_{\mathrm{T}}^{\mathrm{RECO}}\right)$.

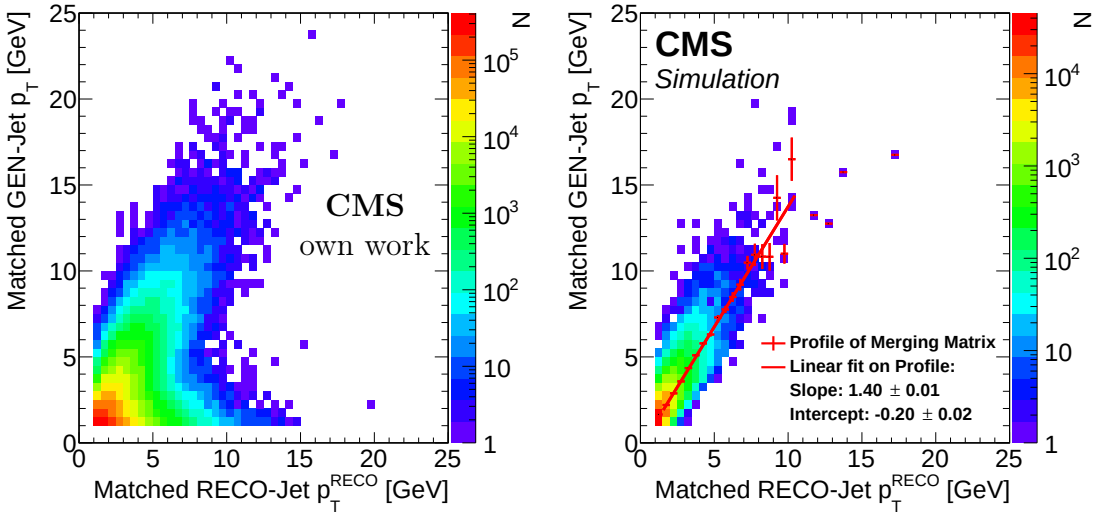


Figure 5.2: The correlation between the reconstructed jet p_{T} to generated jet p_{T} for a special selection of tight isolated matched jets generated with EPOS-LHC [96] is shown. On the left side this correlation is illustrated for the closest generator jet inside a $\Delta\phi$ and $\Delta\eta$ square around the reconstructed jet. The same correlation with an additional cut on non-jet generator level particles not to be present inside the reconstructed jet area can be seen on the right side. The red dots on the right side show the average jet p_{T} value for generated jets with the RMS as error. A linear fit is performed to parametrize the red profile in order to obtain the calibration factor from the slope of the fit function. The fit result is additionally shown in the legend of the figure on the right.

For the purpose of determining the relationship between the generated jet p_{T} and reconstructed jet p_{T} a matching procedure with a very tight isolation criteria based on a EPOS-LHC generated sample (see table 5.2) is performed. To avoid the impact of noise only reconstructed and generated jets with a $p_{\mathrm{T}} > 1\,\mathrm{GeV}$ are considered. In each event only the reconstructed CASTOR jet with the highest p_{T} is matched to a generator jet. The full energy of the generated jets should be contained in CASTOR to reduce effects on the energy correction by leakage. Such effects are considered to be corrected by the unfolding of the whole jet spectrum in the section 5.4 below. For the matching only generator jets

with an η between -6.1 and -5.7 are taken to ensure the containment in the detector. Note that this is derived from the borders of the CASTOR detector in η from -6.6 to -5.2 minus the size of the jet radius clustered via the anti-k_t algorithm with $R = 0.5$. From the remaining generator jets in the event the closest one to the reconstructed jet is registered as possible matching candidate. Additionally, the closest generator jet must have at least a $\Delta\phi < 0.2\,\text{rad}$ to the reconstructed jet, which is around half of the width ($\pi/8$) in ϕ of a CASTOR tower. Otherwise the whole event is rejected. The relationship in p_T between this matching candidate at generator level and the hottest p_T calorimeter jet is shown on the left side of figure 5.2. A tail where for low-p_T generator jet is assigned to a detector level jet with a high-p_T value can be seen there. This occurs if particles not included in the generator jet hit the area of the detector level jet and the affected penetrated tower is clustered in the reconstructed jet. To ensure that the energy of the reconstructed jet has a clean relation only to the generator jet found as matching candidate, no other generator particles than the ones belonging to the generator jet are allowed to hit the tower area of the reconstructed jet in the detector. If this criterion is fulfilled for this event, the generator jet is finally matched to the reconstructed one. The correlation in p_T of the matched jets is shown on the right of figure 5.2 where p_T^RECO is the reconstructed p_T value of the hottest (in p_T) detector level jet in an event compared to the matched generator jet p_T.

In addition, the distribution shown on the right of figure 5.2 is binned in values of p_T^RECO and in each bin the mean generator jet p_T is calculated. The result is plotted as red profile on top of the distribution on the right side of figure 5.2 and the RMS of the generator jet p_T as error. The red line is a linear fit ($a\,p_\text{T} + b$) on the profile in the range of $1.5\,\text{GeV} < p_\text{T}^\text{RECO} < 10.5\,\text{GeV}$ and the result is shown on the right side of figure 5.2 in the bottom right corner. It shows that any reconstructed jet p_T needs to be scaled up by a factor of about 1.4 to get a calibrated momentum corresponding to the detector level jets. This first leading order correction of the calorimeter jet p_T is applied on all upcoming plots in this thesis.

5.4 Unfolding

The purpose of unfolding is to regain the forward hadron level jet spectrum from the measured one in CASTOR by reversing the effects of smearing and imperfect efficiency of the components in the CMS detector. For example the structure of the beam pipe prevents in a certain η-region that particles can pass through and therefore some particles do not reach the CASTOR detector. After the leading order energy correction for CASTOR jets is applied, a dedicated response matrix that describes the migration between the reconstructed and the generated jets in p_T is needed to unfold the measured jet spectrum in CASTOR and recover the hadron level jet spectrum. For this purpose, the unfolding procedure uses the response matrix to calculate a correction of the binned measured jet spectrum to get hadron level one. In this analysis the unfolding is done by means of an iterative method by D'Agostini [97] which is based on the Bayes theorem.

5.4.1 Validation

For a good acceptance correction of the CMS experiment and especially the CASTOR calorimeter, the Monte Carlo must describe the detector as precisely as possible. Therefore, a comparison of key parameters of reconstructed CASTOR jets between unbiased data and Monte Carlo samples from table 5.2 is performed to validate the generated data samples used in the unfolding. All jets have an additional cut of $p_T > 3\,\text{GeV}$ to be well above noise but also in a p_T range where the matching between generator and detector level jets works well (see sec. 5.4.2).

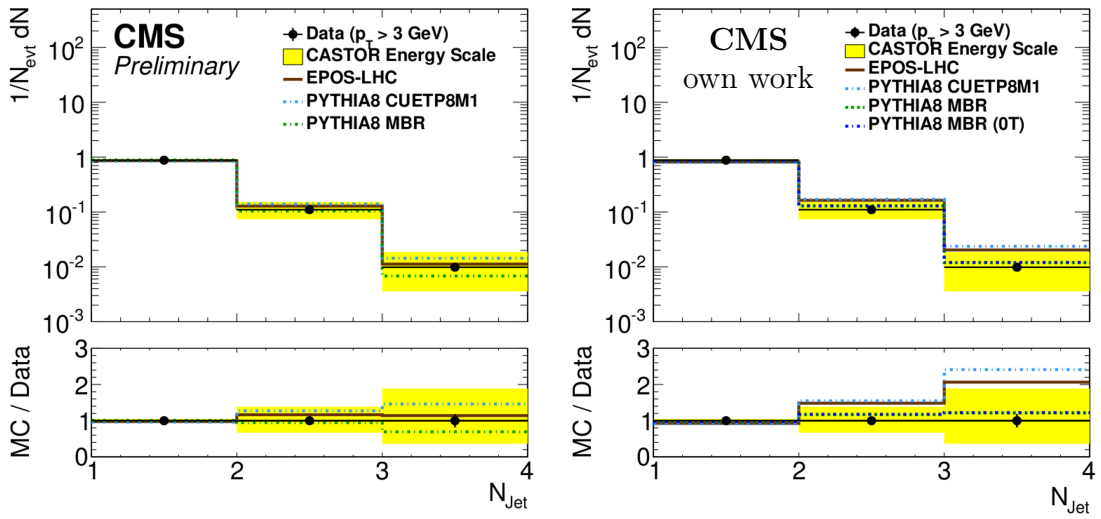


Figure 5.3: Average multiplicity per event of detector level jets in CASTOR for jets with a $p_T > 3\,\text{GeV}$. On the left side the data is compared to models without magnet field and the CASTOR detector at the measured position. The right side compares data to models with the detector at the nominal position and 3.8 T field.

In figure 5.3 the multiplicity of reconstructed CASTOR jets per event with a p_T larger than 3 GeV is shown. The yellow error band shows uncertainty due to the uncertainty on the total energy calibration factor of the detector interacting with the 3 GeV cut in jet p_T. The left side shows the jet multiplicity for unbiased data compared to Monte Carlo samples with no magnetic field in CMS and the detector at the measured position. These samples are closest to the condition during data taking, while other samples with the CASTOR detector at nominal position symmetric around the beam pipe are compared on the right side of figure 5.3. On the right of figure 5.3 all samples are generated with a magnetice field of 3.8 T except for PYTHIA8 MBR sample. On the left side EPOS-LHC is representing the data at best but also the other PYTHIA8 based samples are inside the energy scale uncertainty. The right side shows that the same samples with magnetic field are above the data except for two PYTHIA8 MBR samples, which also demonstrates that the magnetic field in the very forward direction has no big effect.

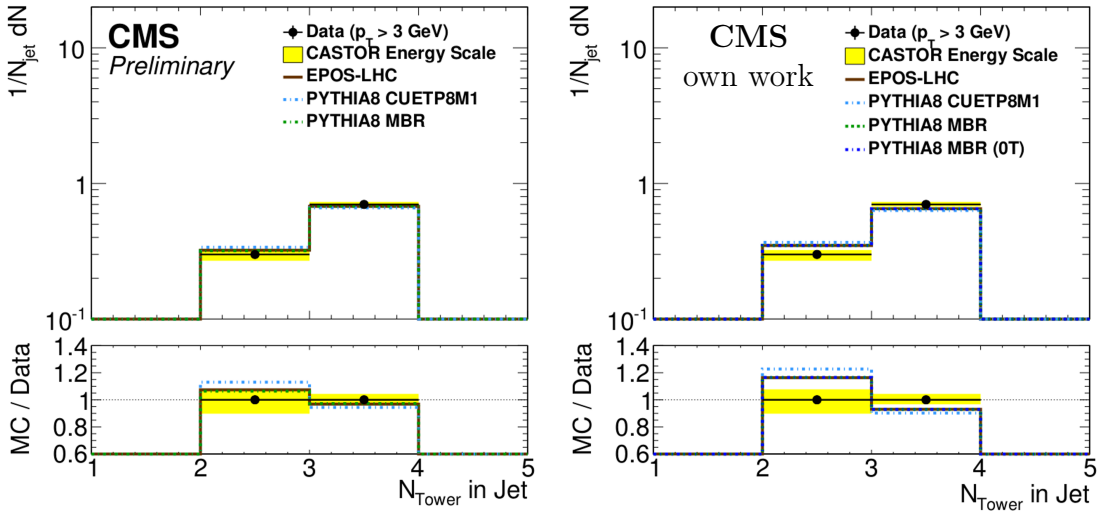


Figure 5.4: Average number of towers per jet clustered into a CASTOR jet with a $p_{\mathrm{T}} > 3\,\mathrm{GeV}$. On the left side the data is compared to models without magnetic field and the CASTOR detector at the measured position. The right side compares data to models with the detector at the nominal position and 3.8 T.

The average number of towers per jet which are combined into a jet by the anti-k_t algorithm (see. sec. 5.1.1) can be seen in figure 5.4. On the left the ZeroBias data is compared to Monte Carlo samples with CASTOR at measured position and no magnetic field while on the right the simulation is done with the detector at the nominal position and magnetic field. The samples on the left side show a better behaviour compared to the ones on the right side. The right side of figure 5.4 shows again no relevant effect due to the magnetic field comparing the two PYTHIA8 samples with MBR. Because of the p_{T} cut of the jets there is no single jet with less than two combined towers. Due to the geometry of the CASTOR detector with a distance between towers of $\pi/8 \approx 0.39\,\mathrm{rad}$ it appears very rarely that more than three towers are clustered to a jet with a radius parameter of $R = 0.5$ in the anti-k_t algorithm. In figure 5.3 and figure 5.4 the Monte Carlo samples generated without magnetic field and CASTOR at the measured position show a good comparison to data which indicates that the reconstruction of jets is consistent for data and Monte Carlo.

The top row of figure 5.5 shows how deep the shower from a jet is penetrating the CASTOR calorimeter. This longitudinal depth (see eq. 5.7) is given in mm distance from the interaction point. Because CASTOR is placed on the negative side of z, the highest value in the jet depth represents the front of the detector and the lowest value the back as seen from the interaction point. The bottom row shows the longitudinal width of the jet shower in CASTOR according to eq. (5.8). Also here the yellow error band shows the uncertainty due the energy calibration of CASTOR. On the left and on the right the data is compared to the same samples as in the previous figures. It can be seen that the depth of the jet as

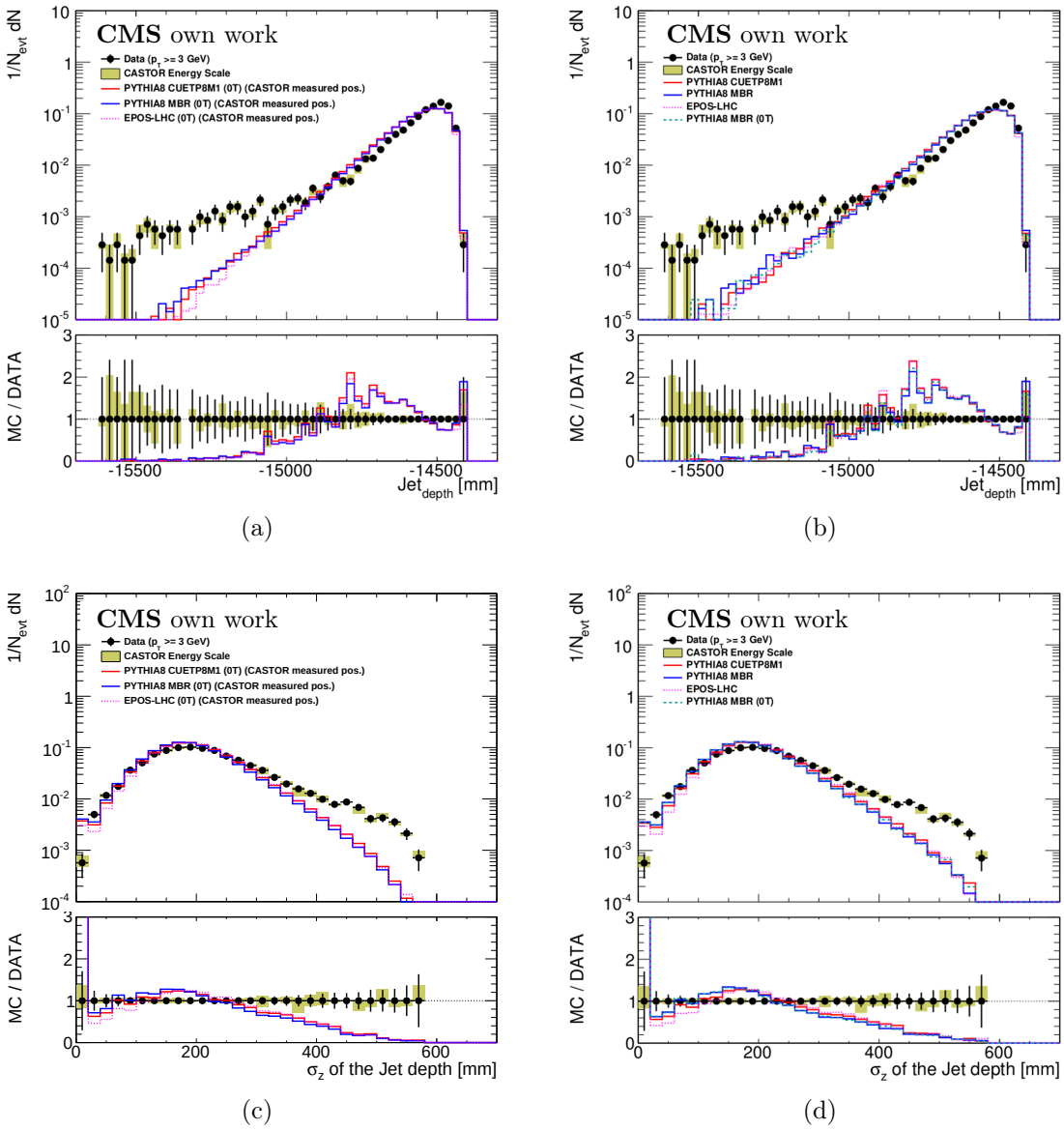


Figure 5.5: Energy weighted longitudinal mean depth of a reconstructed jet in CASTOR on the top and the longitudinal RMS of the jet shower distribution in CASTOR on the bottom. Both only for jets with an $p_{\mathrm{T}} > 3\,\mathrm{GeV}$. The depth is given in mm away from the interaction point. On the left side the data is compared with models where CASTOR is at the measured position and on the right with models where the calorimeter is at nominal position.

shown in figure 5.5a and figure 5.5b is described by the Monte Carlo samples up to a value of -15 m. Also the width of the jet shower is described by the Monte Carlo samples up to jets with a longitudinal width of 300 mm. In any case it is very reassuring that there is

almost no difference in the jet shower development if the simulation is done with or without magnetic field or if the CASTOR detector is at the nominal or the measured position.

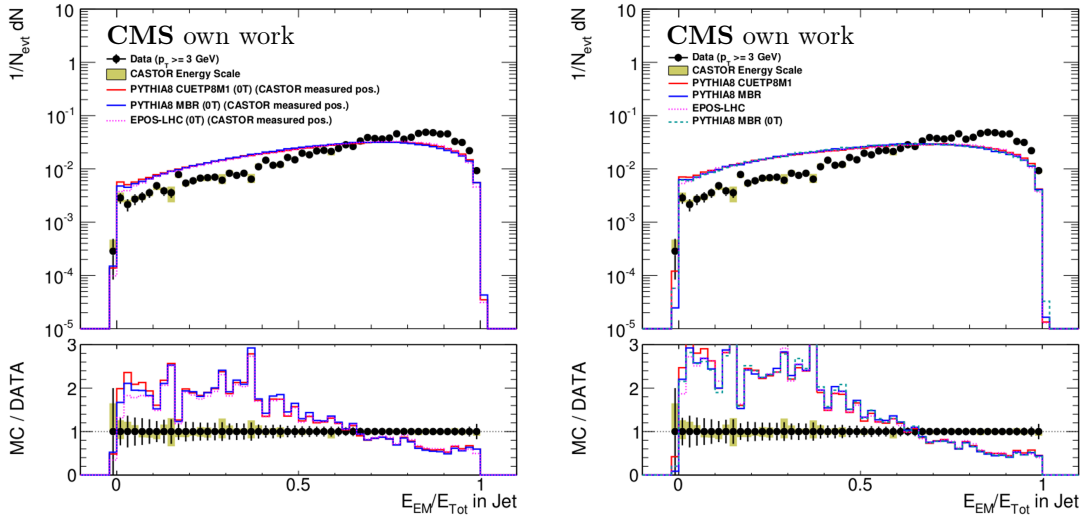


Figure 5.6: Ratio between the electro magnetic energy and the total energy in CASTOR for a jet. On the left side the data is compared with models where CASTOR is at the measured position and on the right with models where the calorimeter is at nominal position.

The ratio of the electromagnetic energy to the total energy of a jet (see eq. 5.4) is shown in figure 5.6 with the left and the right side the figure 5.6 comparing the data to the same generated samples as the previous figures. Compared to Monte Carlo the data show a more pronounced electormagnetic content of the jets. Also here the Monte Carlo samples also describe the data with similar precision for different conditions of the magnet field or the detector position.

The top row of figure 5.7 the ϕ distribution of the reconstructed jets in CASTOR is shown. The width of the bins is according to the sector size of the CASTOR detector. Because the phi distribution depends critically on the position of the detector in figure 5.7a and figure 5.7b, in addition to the energy scale uncertainty in yellow also the position uncertainty band in magenta is shown. The position uncertainty contributes, as expected, significantly to the uncertainty in the ϕ distribution of the detector level jets. In figure 5.7b the models where CASTOR is at the nominal position show the expected behaviour of an almost flat distribution in ϕ. On the other hand, in figure 5.7a the effect of simulation with the detector at the measured position is clearly shown. Even Monte Carlo samples with CASTOR at the measured position cannot perfectly reproduce the ϕ distribution in data. This is ascribed to the limited precision of the intercalibration of the CASTOR channels.

To study the influence of the ϕ distribution on the p_{T} spectrum a systematic check is performed by reweighting the data events. The ratio between data and Monte Carlo for every ϕ-bin is taken as ϕ depending weighting factor. The weight of an event is then

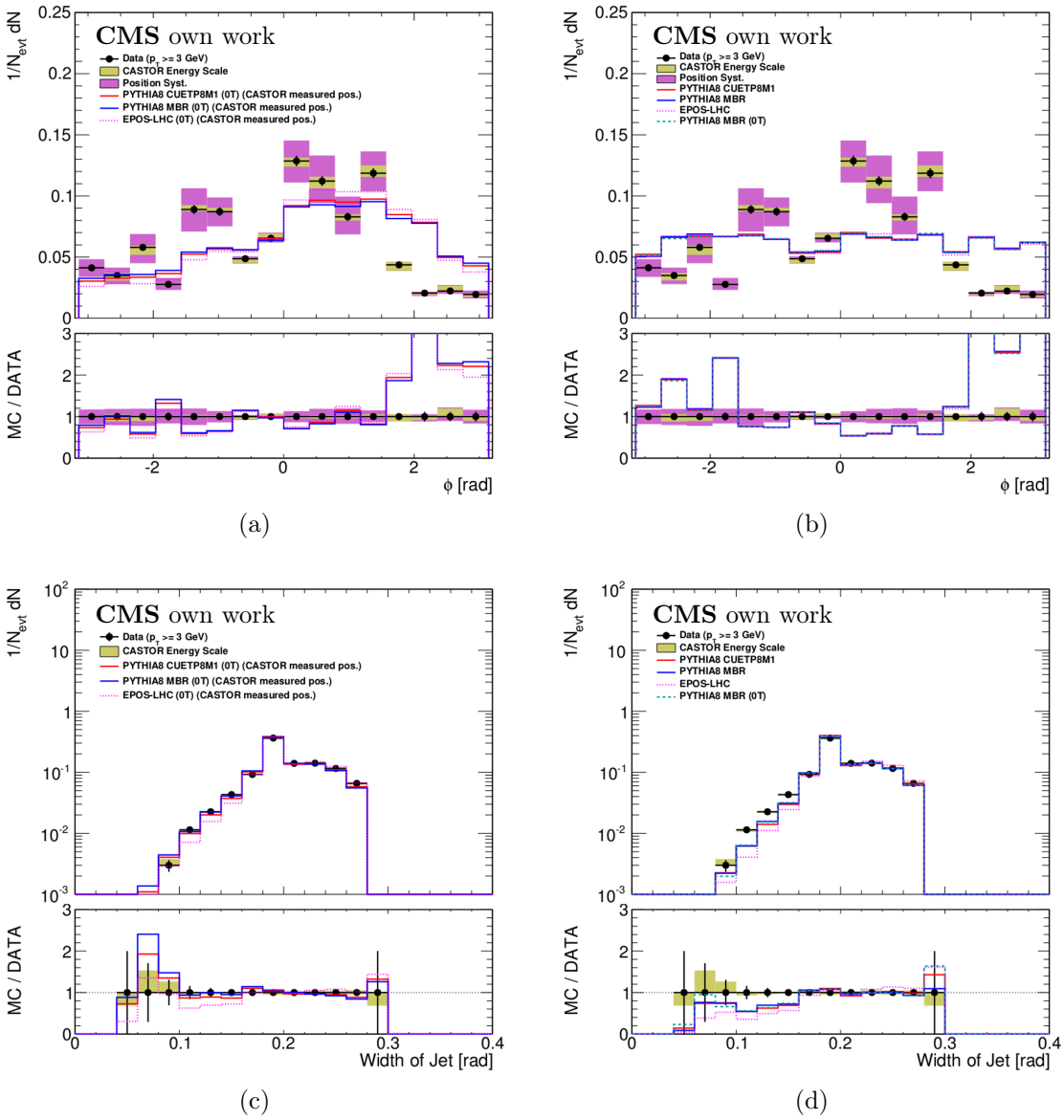


Figure 5.7: The distribution in ϕ of reconstructed jets in CASTOR is shown on the top and on the bottom the width in ϕ. On the left side the data is compared with models where CASTOR is at the measured position and no magnetic field in CMS and on the right with models where the calorimeter is at nominal position and with magnetic field in CMS.

derived by the average of the separate jet weighting factors. Since an event can consist of more than one jet, the reweighted ϕ spectrum will not perfectly follow Monte Carlo predictions. In a second iteration the new spectrum is taken to derive new factors. The resulting ϕ spectrum after the second iteration is shown on the left of figure 5.8. On the

right of figure 5.8 one can see the difference between the nominal reconstructed jet p_{T} spectrum and the spectrum with the ϕ dependent weighted events. One can see that the ϕ distribution has no major influence on the reconstructed spectrum and is equal inside the uncertainties. A similar study has been done with the $E_{\mathrm{EM}}/E_{\mathrm{tot}}$ distribution, also resulting in no major difference between the unweighted and reweighted jet p_{T} distribution.

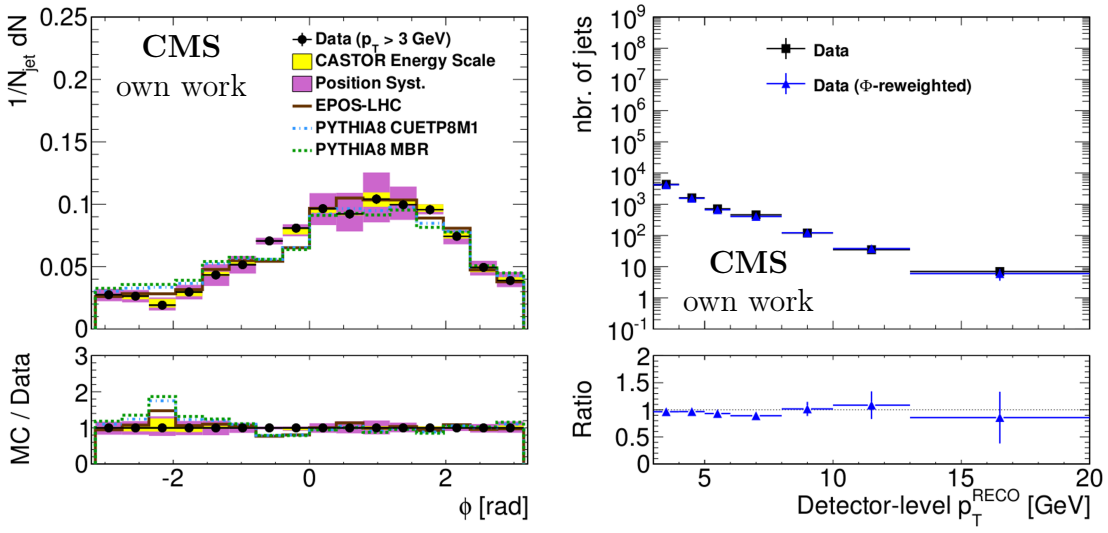


Figure 5.8: The distribution in ϕ of reconstructed jets for in ϕ reweighted events on the left. On the right the reconstructed jet p_{T} spectrum is compared to the jet p_{T} spectrum with reweighted events.

The bottom row of figure 5.7 shows the width in ϕ (see. eq. 5.6) of the reconstructed jets. In figure 5.7c the simulations without magnetic field and correct positioning of the detector show a good agreement over a wide range of the jet width in ϕ. The simulations with the nominal positioned detector in figure 5.7d show a worse agreement to the data compared to Monte Carlo samples in figure 5.7c. Here the pure magnetic field effect by comparing the two PYTHIA8 samples with MBR is very small as well. The effect of an imperfect intercalibration which was very pronounced in the jet ϕ distribution has no major effect in the width in ϕ of the jets.

The jet p_{T} spectrum at detector level can be seen in figure 5.9 where on the left side the data is compared with simulations without magnetic field and a detector at measured position and on the right side with magnetic field and the detector at nominal position. On both sides of figure 5.9 EPOS-LHC describes the data very well compared to the other PYTHIA8 samples. The PYTHIA8 samples tend to overestimate the data with higher jet p_{T} independent of the detector position or magnetic field condition. Therefore EPOS-LHC without magnetic field simulations and a correctly positioned CASTOR detector is later used for the creation of the nominal response matrix to unfold the detector effects. The two other PYTHIA8 samples with the same conditions are used for systematic studies.

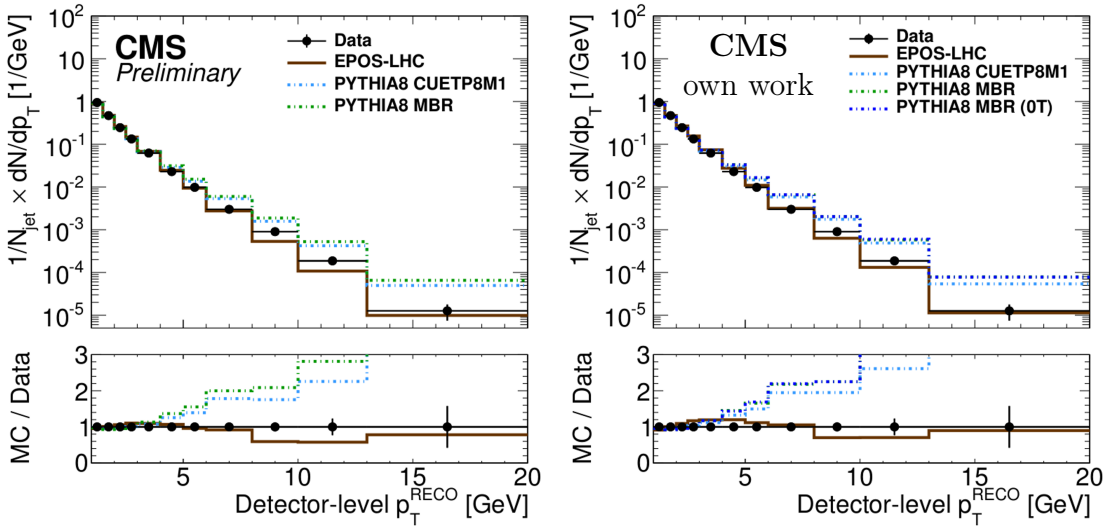
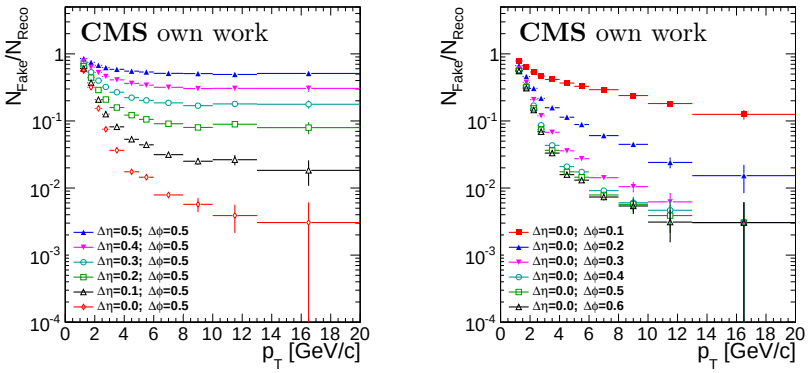


Figure 5.9: Reconstructed jet p_{T} spectrum compared to Monte Carlo simulations with CASTOR at measured position and no magnetic field in CMS on the left side and with a nominally positioned detector and 3.8 T field in CMS on the right side. The spectra are normalized to the visible number of jets.

5.4.2 Matching procedure for response matrix creation

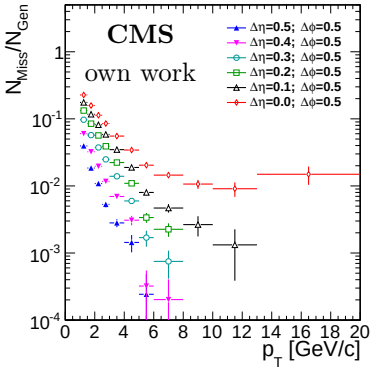
For the response matrix needed for the unfolding, in contrast to section 5.3 a specific jet-matching procedure must be performed with the aim of minimizing failures during the matching instead of having a procedure with a very clean and isolated correlation between generator and detector level jets. It is important to avoid mismatching because each generator level jet that is not matched to a detector level jet and vice versa must be considered as a pure Monte Carlo correction to the final corrected jet spectrum. Generator level jets that could not be matched are called misses and not matched detector level jets fakes. To find the best matching parameters a PYTHIA8 CUETP8M1 sample as before with the measured detector conditions is used.

For the jet matching all generator and detector level jets are sorted in p_{T} beginning with the largest p_{T}. Only generator level jets are taken into account which are in the range of $-6.6+\Delta\eta < \eta < -5.2-\Delta\eta$ where the value of $\Delta\eta = 0$ covers the whole CASTOR range in η. Beginning with the first of the sorted reconstructed jets it is looped over the sorted list of generator jets to find one inside a window of $\Delta\phi$ around the reconstructed jet-ϕ. In case no such generator jet is found the reconstructed jet is marked as a fake. Otherwise, when one or more generator jets are in the window, the hottest in p_{T} is matched to the reconstructed one. Both matched jets are then removed from the list of remaining generator and detector level jets. This procedure is repeated by taking the next reconstructed jet of the sorted list. After all CASTOR jets are matched to a generator level jet or marked as fake the remaining jets on generator side are marked as misses.

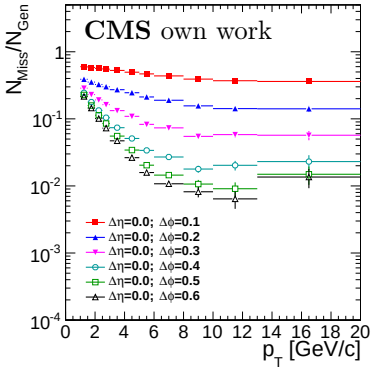


(a)

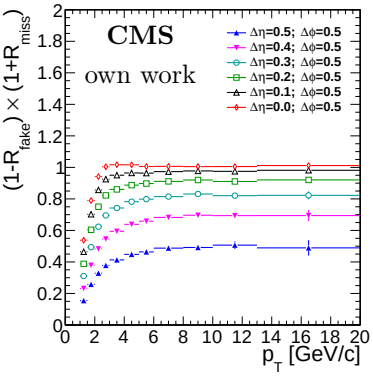
(b)



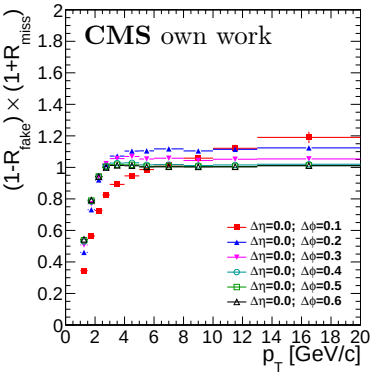
(c)



(d)



(e)



(f)

Figure 5.10: Dependence of fakes and misses on jet p_{T} for different cuts of $\Delta\phi$ and $\Delta\eta$ during the matching process. In the top row the relative number of fakes and in the middle row the relative number of misses is shown. The bottom row shows the ratio of the number of unfolded jets to reconstructed jets $N_{\mathrm{unf}}/N_{\mathrm{RECO}} = (1 - R_{\mathrm{fake}})(1 + R_{\mathrm{miss}})$ depending on fakes and misses. On the left side the variation in $\Delta\eta$ and on the right side the variation in $\Delta\phi$ is shown.

For different values of the matching parameters $\Delta\eta$ and $\Delta\phi$ the behaviour of the relative amount of misses and fakes depending on the reconstructed jet p_{T} can be seen in figure 5.10. Thereby the effect of different parameter values in the matching process is shown for $\Delta\eta$ on the left side and for $\Delta\phi$ on the right side. In figure 5.10a the amount of fakes reduces overall with a higher p_{T} value but is also reduced significantly by a smaller $\Delta\eta$. This is understandable since with a smaller $\Delta\eta$ a wider range of generator jets in η are considered as possible matching candidates for the detector level jets. On the other hand in figure 5.10c the amount of misses is rising over the whole p_{T} range with a smaller $\Delta\eta$ parameter.

To find a trade-off between misses and fakes by choosing the right value of $\Delta\eta$ it is helpful to look at the overall correction on the number of jets in the hadron level spectrum, which has to be done by the unfolding. The amount of jets after unfolding N_{unf} differs from the measured number of jets N_{RECO} by first removing the fakes from the reconstructed spectrum and afterwords adding the misses to the unfolded hadron level spectrum. Therefore the relative number of hadron level jets is

$$\frac{N_{\mathrm{unf}}}{N_{\mathrm{RECO}}} = (1 - R_{\mathrm{fake}}) \times (1 + R_{\mathrm{miss}}) \tag{5.9}$$

where $R_{\mathrm{fake}} = N_{\mathrm{fake}}/N_{\mathrm{RECO}}$ and similar $R_{\mathrm{miss}} = N_{\mathrm{miss}}/N_{\mathrm{RECO}}$. This ratio should be around one to minimize the effective number of jets that are added to or removed from the final spectrum. In this sense the best results are produced by choosing $\Delta\eta = 0$ as seen in figure 5.10e. Surely a high number of misses and fakes can also compensate one another but as shown in figure 5.10a and figure 5.10c for $\Delta\eta = 0$ and a $p_{\mathrm{T}} > 3\,\mathrm{GeV}$ the relative number of fakes and misses is below 10 %.

On the right side of figure 5.10 the fakes and misses are shown for different $\Delta\phi$ cuts in the matching process. In this case both fakes and misses are overall reduced for a wider matching window in ϕ. Considering that the jet clustering algorithm is operating with a radius parameter of $R = 0.5$ the matching parameter $\Delta\phi = 0.5$ is a natural choice. In figure 5.10f the $\Delta\phi$ values of 0.5 and 0.6 show the best result to the net jet number correction in the unfolding.

In addition, one can look at the response matrix (figure 5.12) itself to check how well the resulting matrix is conditioned. The condition of a matrix A quantifies in general how sensitively in a problem like $Ax = b$ (e.g. unfolding with a response matrix A and the measured spectrum b) the solution of x depends on changes in b [98] and is represented by the condition number κ with

$$\kappa(A) = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}, \tag{5.10}$$

where $\sigma_{\max}$ and $\sigma_{\min}$ respectively are the maximal and the minimal singular value of the matrix A. Thereby a small value of $\kappa < 10$ represents a well conditioned matrix where the approximation of the solution x could be obtained by an unfolding method without large corrections. Especially by using unfolding with a singular value decomposition [99] approach the condition number κ is directly correlated to the uncertainties of the solution

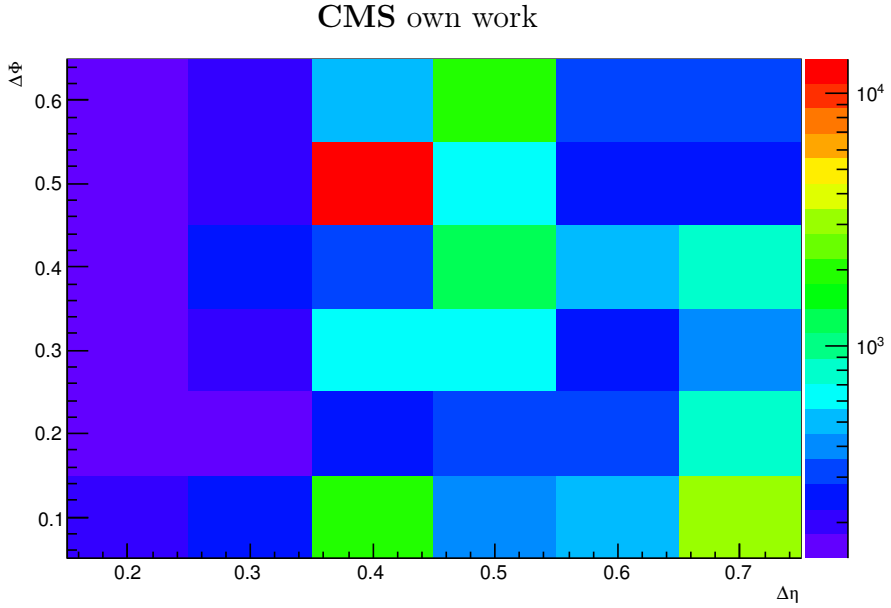
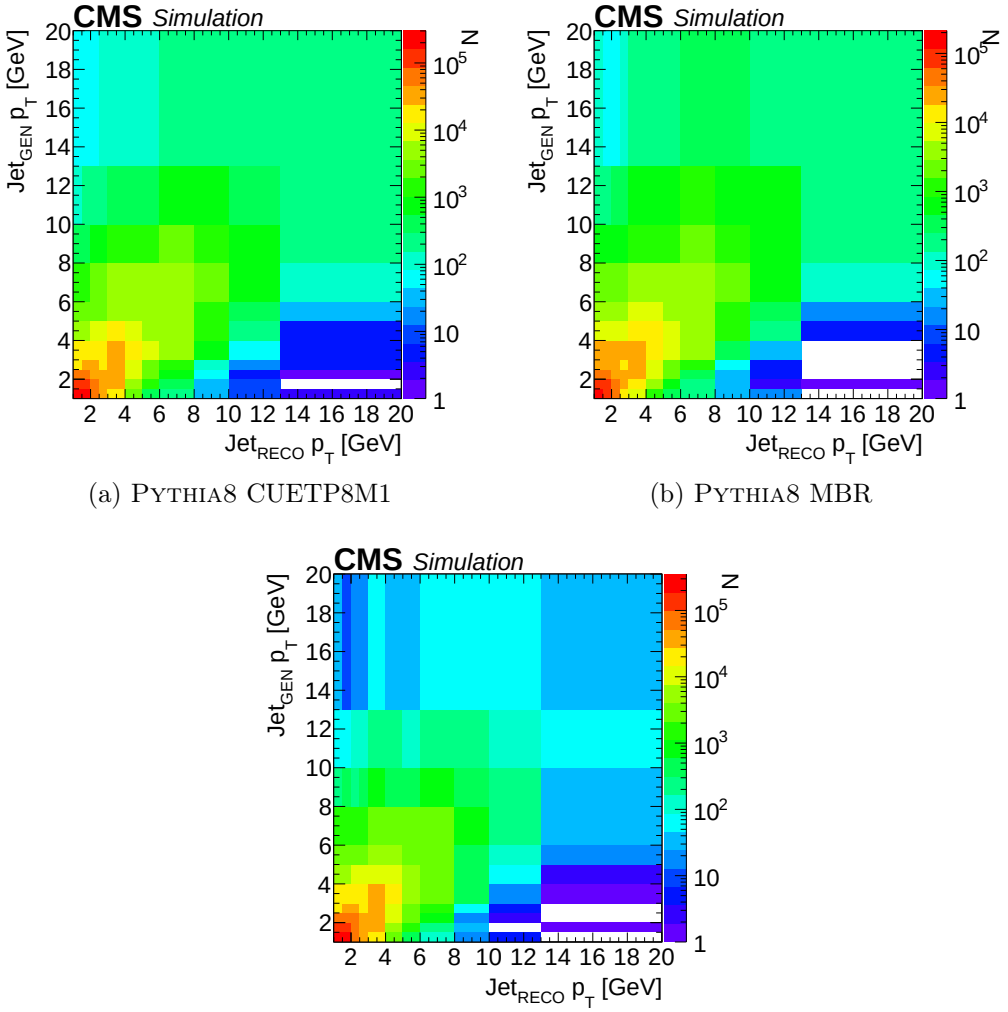


Figure 5.11: Condition number of the response matrix depending on the matching parameters $\Delta\phi$ and $\Delta\eta$.

by giving an upper bound on it [100]. However, since in this analysis this method is not used the condition number is only an indication as to how the uncertainty of the unfolded result depends on the response matrix. For the different matching parameters the condition number of the resulting response matrix is shown in figure 5.11. The smallest value of κ is reached for large $\Delta\eta$ where the matching area is very restricted. In this case the response matrix looks similar to the one in figure 5.2 and better than in figure 5.12 because the generator level jets are completely contained in CASTOR. Therefore, also the condition number of the matrix is much better or smaller than in case of matching jets in the whole CASTOR rapidity range where $\Delta\eta = 0$. But as mentioned before figure 5.10a and figure 5.10e show that a large value of $\Delta\eta$ is not preferred due to many fakes and the corresponding correction based only on Monte Carlo simulation that has to be done by the unfolding. For a fixed $\Delta\eta = 0$ the best conditioned matrix results from choosing $\Delta\phi = 0.5$. This confirms the selection of the matching parameters alone from figure 5.10 derived.

From the simulation in figure 5.10 and figure 5.11 it is concluded to use $\Delta\phi = 0.5$ and $\Delta\eta = 0$ as matching parameter. The final response matrices for different generator samples can be seen in figure 5.12 and are later used to gain the unfolded result and systematic uncertainties. Additionally it can be seen especially in figure 5.10e and figure 5.10f that a cut in $p_{\mathrm{T}} > 3\,\mathrm{GeV}$ is optimal to minimize as much as possible the pure Monte Carlo correction concerning fakes and misses in the unfolding procedure. This cut should be done on the final unfolded spectra so that jets with a $p_{\mathrm{T}} < 3\,\mathrm{GeV}$ are still included in the calculation of migration into higher p_{T} bins of the jet spectrum.



(a) PYTHIA8 CUETP8M1

(b) PYTHIA8 MBR

(c) EPOS-LHC

Figure 5.12: Response matrix showing the relation between the reconstructed jet p_T and the generated jet p_T. On the left for PYTHIA8 CUETP8M1-Tune, on the right for PYTHIA8 with MBR and in the middle for EPOS-LHC. All three samples are simulated with $B = 0\,\mathrm{T}$ and CASTOR at measured position.

5.4.3 Performance

With the response matrix illustrated in figure 5.12, the reconstructed jet spectrum and information about the amount of fakes and misses (see figure 5.10) the unfolding can be performed. For this purpose the iterative method of D'Agostini with early stopping [97] is used as it is implemented in the ROOUNFOLD [101] package.

Starting from the predictions of Monte Carlo simulations in each iterative step the result is corrected for migrations and detector efficiency effects. In the bin H_i of the hadron level

spectrum the new estimated bin content is

$$\hat{n}(H_i) = \frac{1}{\epsilon_i} \sum_{j=1}^{N_D} n(D_j) P(H_i|D_j) \tag{5.11}$$

where N_D, $n(D_j)$, $P(H_i|D_j)$ and ϵ_i respectively are the number of bins in the detector level spectrum, the number of detector level jets in the bin D_j, the conditional probability of a jet to be in bin H_i when it was detected in bin D_i and the efficiency

$$\epsilon_i = \sum_{j=1}^{N_D} P(D_j|H_i) = 1 - P(\text{miss}|H_i) \tag{5.12}$$

in the hadron level bin H_i to account for the probability of a miss $P(\text{miss}|H_i)$ in this bin. By using the Bayes theorem the conditional probability $P(H_i|D_j)$ can be written as

$$P(H_i|D_j) = \frac{P(D_j|H_i)\, P_0(H_i)}{\sum_{k=1}^{N_H} P(D_j|H_k)\, P_0(H_k)} \tag{5.13}$$

where N_H is the number of bins in the hadron level spectrum, $P(D_j|H_i)$ the normalized response matrix from figure 5.12 and $P_0(H_i)$ the initial probability of a hadron level jet to be in the bin H_i. The initial probability

$$P_0(H_i) = \frac{n_0(H_i)}{\sum_{k=1}^{N_H} n_0(H_k)} \tag{5.14}$$

is obtained by the hadron level spectrum $n_0(H_i)$ and the total number of hadron level jets generated with the same Monte Carlo simulations to create the response matrix.

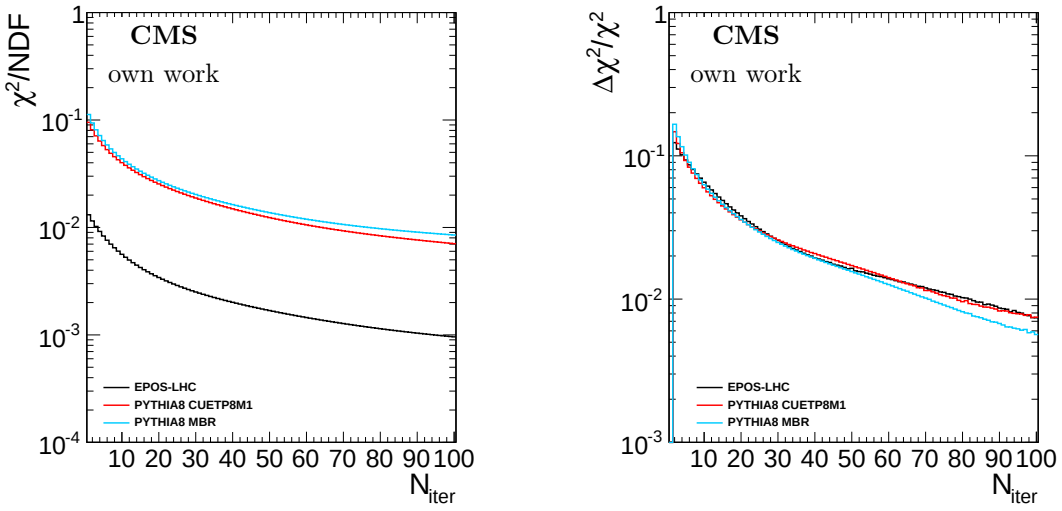
In the first iterative step of the D'Agostini method the number of hadron level jets $\hat{n}_1(H_i)$ in bin H_i is then estimated by eq. 5.11 with $P_0(H_i)$ from eq. 5.14. After determine the unfolded spectrum $\hat{n}_1(H_i)$ in the next iterative step the corresponding probability

$$P_1(H_i) = \frac{\hat{n}_1(H_i)}{\sum_{k=1}^{N_H} \hat{n}_1(H_k)} \tag{5.15}$$

is replacing $P_0(H_i)$ in eq. 5.13 to estimate a new unfolded spectrum $\hat{n}_2(H_i)$. This is then repeated in each iteration so that the unfolded spectrum $\hat{n}_k(H_i)$ after k iterations is derived from the previous result $\hat{n}_{k-1}(H_i)$. With an infinite number of iterations the result converts to the MLE (Maximum Likelihood Estimator) of the problem. This limit corresponds to solving the unfolding problem by matrix inversion, which is characterized by small bias but high variance. The regularization is achieved by stopping after a certain amount of iterations has been performed. As fewer iterations are done as higher the regularization of the unfolding problem. In the limit of a single iteration, the result is essentially equivalent to a simple bin-by-bin correction.

The described procedure does not mention fakes so far while misses are introduced via efficiency (e.q. 5.12). In RooUnfold [101] the fakes are taken into account by adding a bin H_f to the hadron level spectrum and the response matrix with the total number of fakes $n_0(H_f)$ and the probability of a fake jet $P(D_j|H_f)$ in a certain bin D_j of the detector level spectrum. Therefore the number of fakes $\hat{n}_k(H_f)$ is part of the iterative estimation of the unfolded spectrum and will change during the iterations.

Uncertainties in RooUnfold [101] are treated after [97] as Poisson errors of the bin content $n(D_j)$ corresponding to mean and variance in the detector level spectrum. The covariance of the unfolded spectrum $\hat{n}_k(H_i)$ is derived via standard error propagation. Statistical uncertainties in the response matrix and the hadron level spectrum of the Monte Carlo generator due to the finite number of generated events are not considered in the error estimation of the unfolded spectrum. It is the duty of the analyzer to handle that and to estimate as well as minimize their impact.



(a) (b)

Figure 5.13: On the left the χ^2 distribution based on the difference between the backfolded to the reconstructed jet p_{T} spectrum (e.q. 5.16) depending on the number of iterations used in the D'Agostini method with early stopping is shown. On the right the relative difference of χ^2 between one iterative steps is shown.

Using the D'Agostini method the regularization of the unfolding and hence the number of iterations needs to be determined. Therefore in figure 5.13 the χ^2/NDF value depending on the number of iterations and with the number of bins N_D in the detector level jet p_{T} spectrum as number of degrees of freedom is shown on the left. Furthermore,

$$\chi^2 = \sum_{j=1}^{N_D} \left[n(D_j) - \nu(D_j)\right]^2 / \nu(D_j) \tag{5.16}$$

where $n(D_j)$ and $\nu(D_j)$ are the number of detector level jets and the number of backfolded jets in the bin D_j. The backfolded spectrum is the back-smeared spectrum of the unfolded result obtained by D'Agostini after a certain number of iterations, which is obtained by subtracting misses, multiplying the response matrix, and adding fakes. For the unfolding and back-smearing the same response matrix is used. Also for the calculation of χ^2 only Poisson errors $\sqrt{\nu(D_j)}$ of the back-smeared spectrum are taken into account. The right side of figure 5.13 shows the relative change of χ^2 going from $N-1$ to N iteration depending on the absolute number of iterations used in the unfolding process. After 80 iterations the relative change of χ^2 is below 1 % for all Monte Carlo generated response matrices shown in figure 5.12. Variations of the number of iterations around 80 have therefore no big impact on the unfolded result.

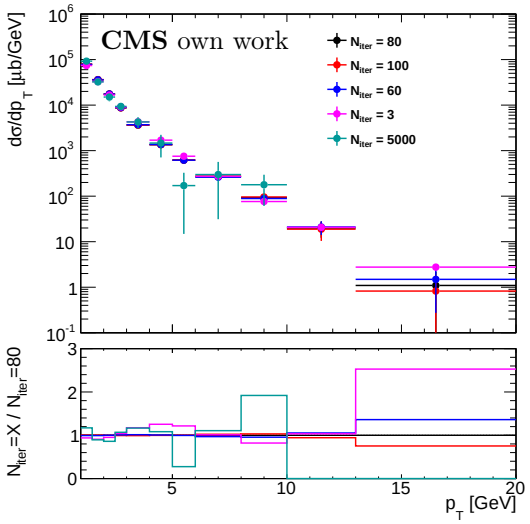


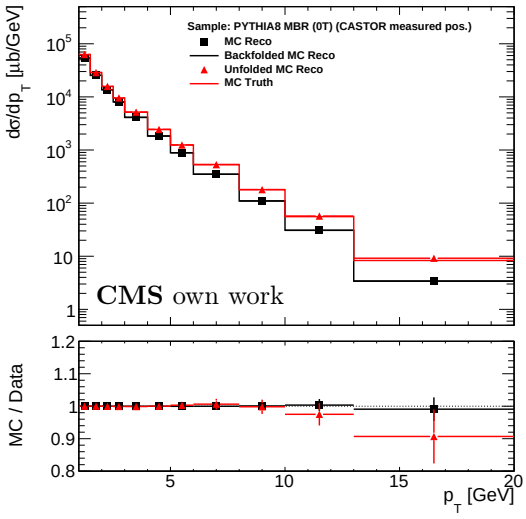
Figure 5.14: Unfolded jet p_{T} spectrum taken by the average unfolded spectra for the three different Monte Carlo generator samples Pythia8 CUETP8M1, Pythia8 MBR and EPOS-LHC for different iterations of D'Agostini with early stopping. The bottom plot shows the ratio to the result obtained by using 80 iterations.

The stability of the unfolded result depending on the number of iterations used during unfolding is also demonstrated in figure 5.14. Here the resulting spectra obtained from data after unfolding with a different number of iterations are compared. At the bottom of figure 5.14 the ratio between these resulting spectra to the one obtained by using D'Agostini with 80 iterations is shown. The difference of the unfolded spectra using 60, 80 or 100 iterations is small except for the highest p_{T} bin. For three iterations the outcome is still considerably different to the one produced with 60, 80 and 100 iterations. The shape of the spectrum is not yet stable, indicating a modulation with a peak deviation of almost 20 % around 4 GeV and 9 GeV. This is an artefact of a too strong impact of the regularization after a small number of iterations. Using 5000 iterations the result shows the expected fluctuations since in this case the D'Agostini method corresponds to matrix inversion.

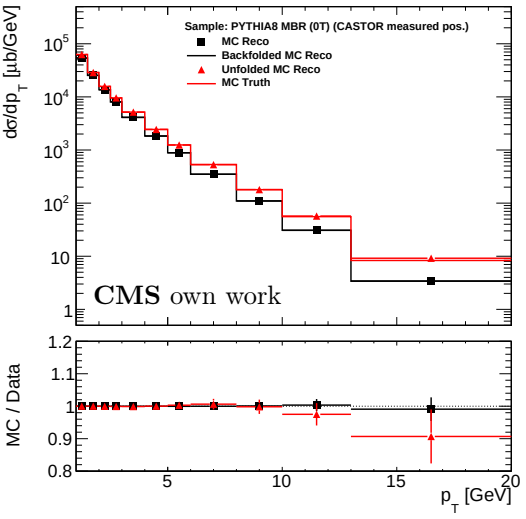
Together with figure 5.13b 80 iterations for the unfolding by D'Agostini with early stopping are chosen as regularization parameter. The impact of the regularization in this regime is stable and the variation in the highest p_{T} bin is likely a borderline effect. This bin is not used the final physics result.

For three different Monte Carlo samples in figure 5.15 the unfolded detector level spectrum is compared with the generator level spectrum. The unfolding is done by D'Agostini with early stopping, using 80 iterations. The ratios of the generator level divided by the unfolded spectrum for the generated samples can be seen as red curve at the bottom of figure 5.15. All Monte Carlo samples show that this ratio falls at higher p_{T}. This happens because of the upper limit of the response matrix at 20 GeV no migration of higher p_{T} jets, which would show up in the lower visible p_{T} bins, can be take into account. Therefore the last p_{T} bin from 13 GeV to 20 GeV is not shown in the final result due to the lack of migration. In figure 5.15 in black the back-smeared spectrum of the unfolded one in red and the detector level spectrum of the generated sample are shown. At the bottom one can see also in black the ratio of the backfolded divided by the detector level spectrum. Both agree very well since both spectra are produced by the same Monte Carlo generated events and show the unfolding is working completely self-consistently as expected.

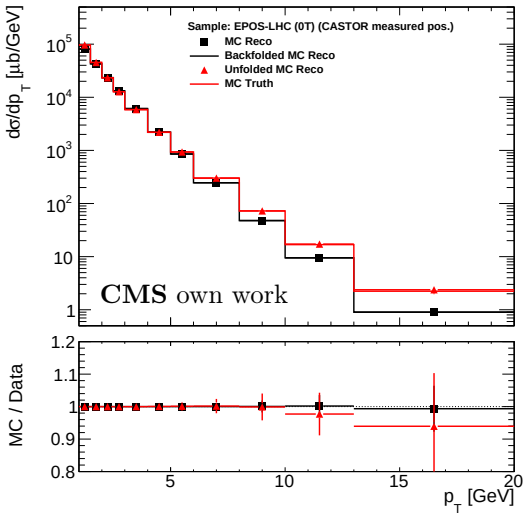
Being aware of the fakes and misses in figure 5.10 and the missing migration effects at the border of the spectra as shown in figure 5.15 the final result of the unfolded spectra is shown in range of 3 GeV up to 13 GeV in p_{T}. Nevertheless, the whole range in p_{T} from 1 GeV to 20 GeV of the reconstructed spectrum is used in the unfolding to account for the migration effect into the final physics range. The unfolding is then done using 80 iterations in the D'Agostini method with early stopping where the unfolded result is stable versus the exact number of iterations as shown in figure 5.13 and figure 5.14.



(a) Pythia8 CUETP8M1



(b) Pythia8 MBR



(c) EPOS-LHC

Figure 5.15: Comparison of generator level jet p_{T} with unfolded jet p_{T} in red and detector level jet p_{T} with the backfolded jet p_{T} spectrum in black for three different generator samples. The compared generator and reconstructed jets in one subfigure are based on the same Monte Carlo generator. The reconstructed jet p_{T} spectrum is unfolded via the D'Agostini method with early stopping using 80 iterations. All three samples are simulated with $B = 0\,\mathrm{T}$ and CASTOR is at measured position.

5.5 Inclusive jet spectrum

The final unfolding of the jet spectrum is done with the RooUnfold package using the D'Agostini method with early stopping and 80 iterations. Since EPOS-LHC describes the data best at on the detector level (see left figure 5.9), the main result is obtained by unfolding with the response matrix figure 5.12c generated by EPOS-LHC with CASTOR at measured position and 0 T magnetic field in CMS (see table 5.2). Fakes and misses are also determined by this generator sample.

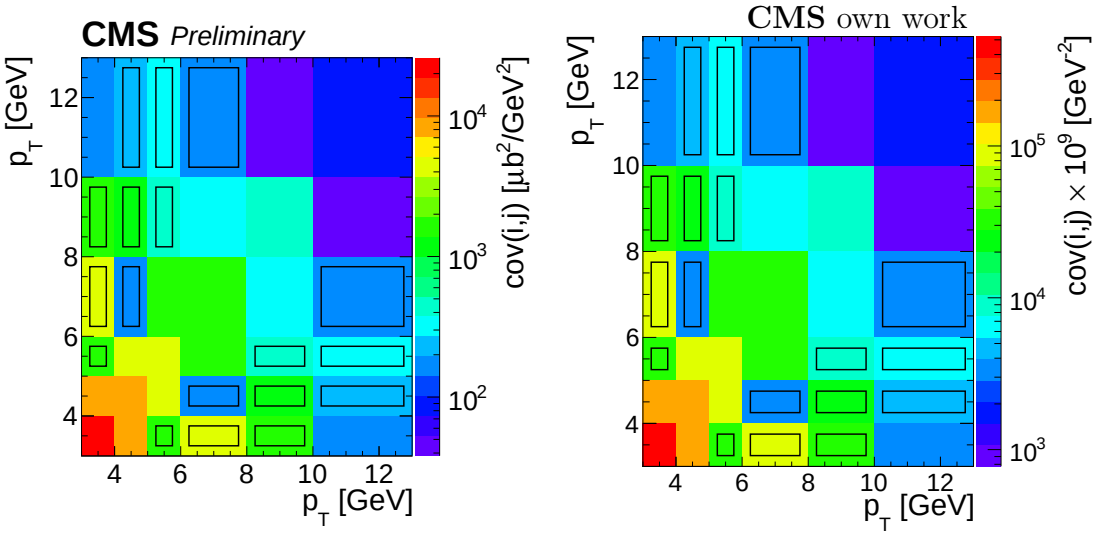


Figure 5.16: Covariance matrix of the unfolded jet p_{T} spectrum using the response matrix of EPOS-LHC. Black boxes indicate negative entries of the covariance matrix. On the left, uncertainties are scaled by luminosity and on the right by the number of visible jets of the spectra.

The statistical uncertainties in the detector level spectrum are assumed to be Poisson fluctuations and are propagated with RooUnfold package to the unfolded spectrum. The covariance matrix of the unfolded jet p_{T} spectrum is shown in figure 5.16 with the uncertainties scaled by luminosity on the left and scaled by the number of visible jets of the spectra on the right.

In figure 5.17 the final unfolded result of the jet p_{T} spectrum in CASTOR is shown. At the top the spectrum is scaled by the luminosity of the recorded data of $0.212\,\mathrm{nb}^{-1}$ and shows the cross section depending on the p_{T} of jets in CASTOR. The bottom shows the shape of the jet p_{T} normalized by dividing the spectrum by the number of jets in the visible p_{T} range. Monte Carlo models based on the Gribov Regge field theory and on various tunes of Pythia are compared to data on the left side of figure 5.17. For different parton density functions used in the tune CUETP8M1 with Pythia8 a comparison to data is shown on the right side. The latter procedure has the advantage of different, and smaller, systematic

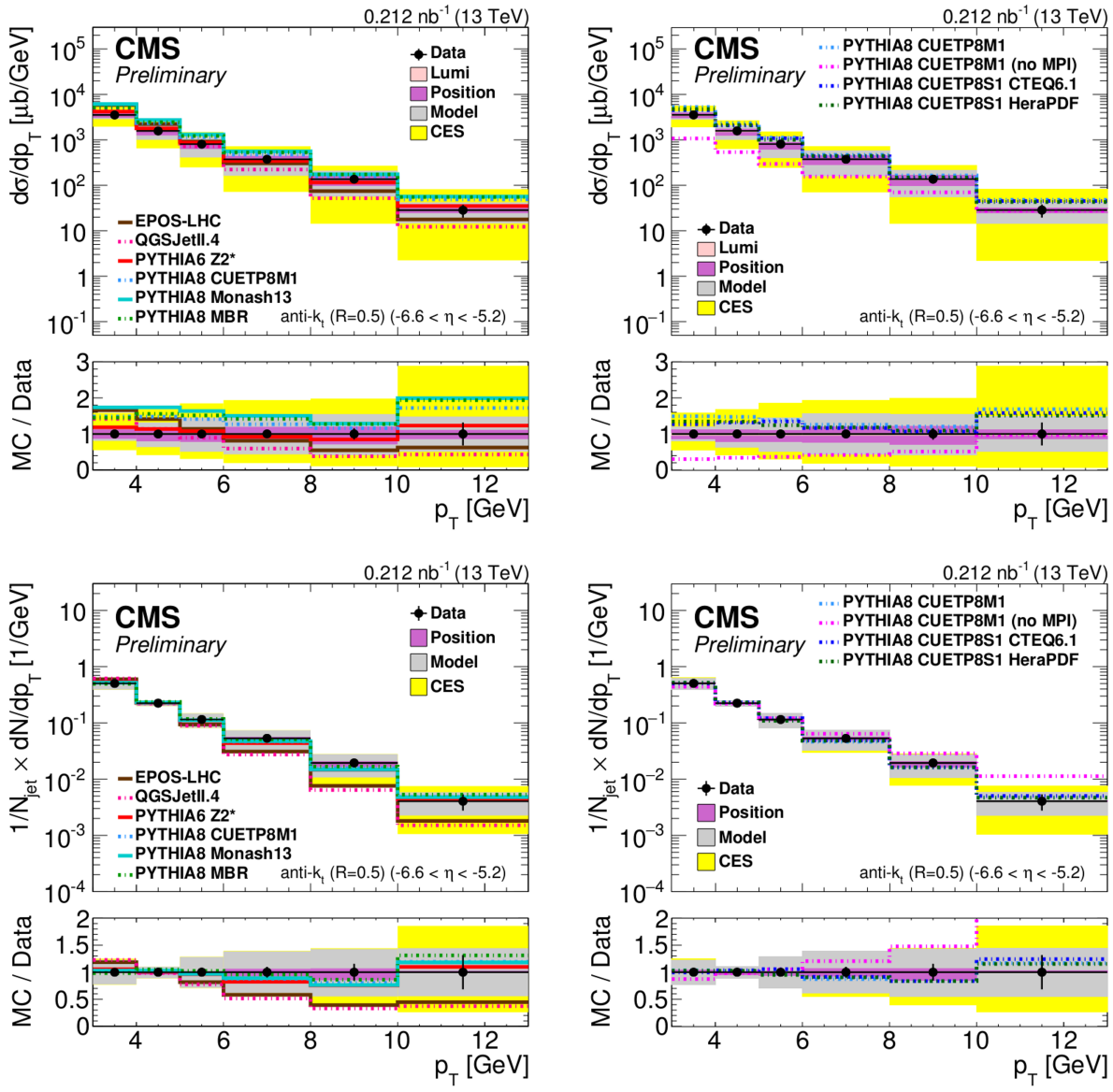


Figure 5.17: Final unfolded p_{T} spectrum of jets inside the CASTOR detector scaled by luminosity on the top and scaled by the visible number of jets inside the p_{T} range on the bottom. The colored bands indicate different systematic uncertainties.

uncertainties. Furthermore, it turns out that the focus on the shape of the spectrum is even more powerful to discriminate between different models.

The coloured band in figure 5.17 shows the size of the different systematics. For the luminosity of recorded data an overall uncertainty of 2.9 % in each bin is taken and shown as pink band in figure 5.17. In the unfolded result it is one of the modest systematic

uncertainties. In case the spectrum is normalized by the visible jet yield no uncertainties from the luminosity are present. The measured luminosity was determined from an inelastic proton-proton cross section study [102] and is based on the HF calorimeters.

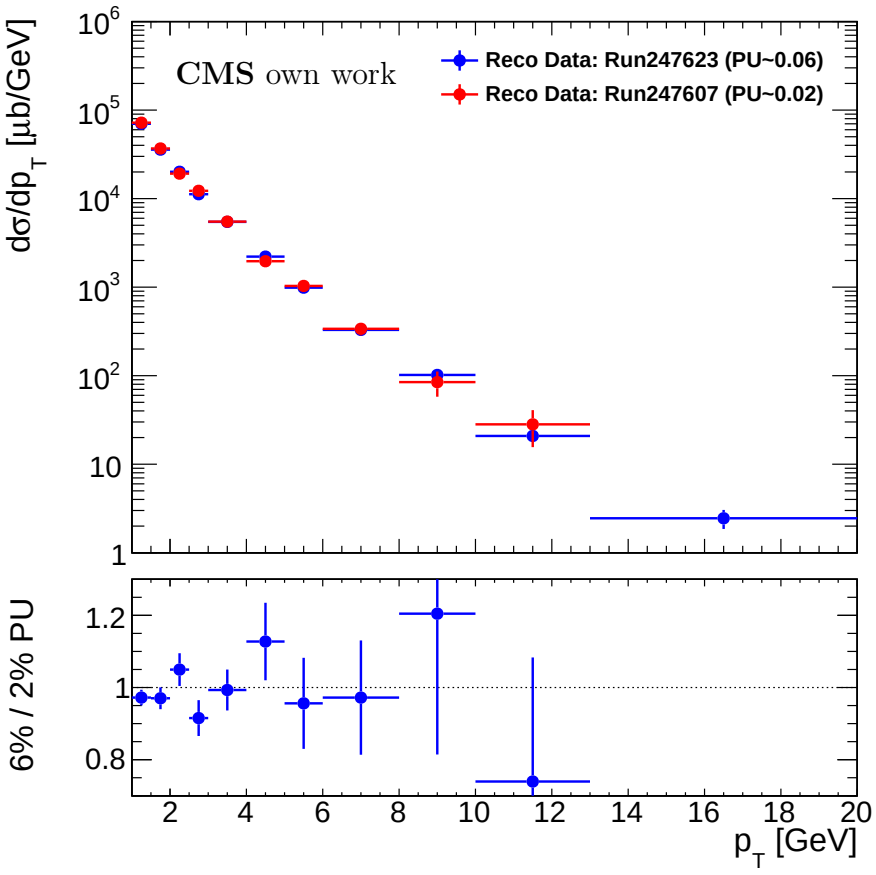


Figure 5.18: Final unfolded p_{T} spectrum of jets inside CASTOR scaled by luminosity for different datasets with 2 % interaction probability in red and 6 % interaction probability in blue.

To resolve an uncertainty in the dependence of the result on the models used in the unfolding two additional Monte Carlo samples are applied in the unfolding procedure of the measured jet spectrum. These are Pythia8 tune CUTETP8M1 and Pythia8 tune 4C with MBR as listed in table 5.2. The measured spectrum is then separately unfolded with the two corresponding response matrices from figure 5.12. The systematic uncertainty originated from the model (grey band in figure 5.17) is then derived by taking the bin-by-bin difference between the main unfolded result from EPOS-LHC and the minimal and maximal result obtained of the two unfolded spectra using Pythia8. The corresponding difference is evaluated after the unfolded spectra are normalized by luminosity or by the jet yield.

The geometry and exact location of CASTOR has a very important impact on the simulated

response to jets. The effect related to the limited precision of the alignment of CASTOR is shown as magenta band in figure 5.17. It is determined by two dedicated generated samples where the CASTOR detector is simulated at the measured position plus and minus a shift in xy-direction equivalent to the uncertainty of the detector position measurement. These samples are produced with PYTHIA8 tune CUETP8M1 as event generator. Using the CUETP8M1 tune, the data is unfolded independently with the generated samples where the CASTOR detector is at nominal position and two times where the detector is shifted corresponding to the position uncertainties. The bin-by-bin difference between the result of the shifted and nominal detector position is calculated after the three unfolded spectra have been normalized either by luminosity or by the number of visible jets. The relative difference according to the PYTHIA8 tune CUETP8M1 sample with the detector at measured position is applied to the final result as position uncertainty.

The largest contribution to the systematic uncertainties is derived from the 15 % uncertainty in the CASTOR energy scale calibration. Considering a jet in the CASTOR detector reconstructed from the energy deposit in the detector towers (see sec. 5.1) the energy calibration has a direct impact on the p_{T} distribution. To evaluate the influence of the systematics in the energy calibration of CASTOR two new jet p_{T} spectra are generated from data by scaling up and down the jet p_{T} by 15 %. These two spectra with the up and down scaled jet p_{T} are unfolded with the EPOS-LHC sample. The relative difference to the main unfolded result delivers the uncertainty from the CASTOR energy scale (CES) and is shown as yellow band in figure 5.17.

To investigate the effect of different interaction probabilities in figure 5.18 the data as specified in table 4.2 are compared to data from a different run with a lower interaction probability of the colliding proton beams. The recorded and unfolded data in figure 5.17 have an interaction probability of 6 % where they are compared to data with an interaction probability of 2 % [102]. Inside the statistical uncertainties there is no difference between the two datasets. Therefore no additional systematics for the interaction probability (also called pileup) is added to the final result in figure 5.17. The higher statistics data with interaction probability of 6 % is not affected by pileup in a statistically significant way. Thus, at the level of precision of the current analysis no systematic uncertainty is assigned.

A summary of the systematic uncertainties is shown in table 5.3 and table 5.4. In the first case the systematic uncertainties are shown for the jet spectrum normalized by luminosity and in the second case for the normalization with the visible number of jets.

Table 5.3: Summary of systematic uncertainties in different p_T bins if the result is normalized by luminosity.

p_T-bin [GeV]	lumi [%]	uncertainty w/o lumi & CES [%]	energy scale [%]	total [%]
3-4	±2.9	−19/+18	−40/+55	−45/+58
4-5	±2.9	−36/+32	−46/+60	−58/+68
5-6	±2.9	−49/+45	−49/+72	−69/+85
6-8	±2.9	−56/+55	−58/+76	−80/+93
8-10	±2.9	−59/+55	−67/+81	−89/+98
10-13	±2.9	−49/+48	−78/+183	−92/+189

Table 5.4: Summary of systematic uncertainties in different p_T bins when result is normalized by number of events.

p_T-bin [GeV]	uncertainty w/o CES [%]	energy scale [%]	total [%]
3-4	−22/+23	−5/+10	−23/+25
4-5	−11/+10	−2/+0	−11/+10
5-6	−29/+28	−5/+5	−29/+28
6-8	−38/+39	−22/+8	−44/+39
8-10	−45/+43	−40/+11	−60/+44
10-13	−44/+44	−59/+73	−73/+85

5.6 Very forward dijet events

In order to enhance the sensitivity to actually very small-x parton dynamics the event sample of very forward dijet events is analyzed in more detail now. In a hard parton collision the two scattered partons create two jets which are orientated in ϕ opposite to each other. In figure 5.19 a scheme of this process is shown where parton p_1 collides with p_2 and creates two jets j_1 and j_2 after hadronization. If such events are seen in the very

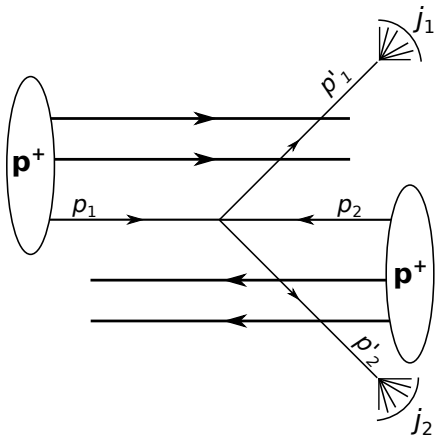


Figure 5.19: Scheme of a parton-parton collision creating two jets after hadronization.

forward region of the CASTOR detector the correlated longitudinal momentum of the two involved partons

$$p_1 = \begin{pmatrix} x_1 E_p \\ 0 \\ 0 \\ x_1 E_p \end{pmatrix} \text{ and } p_2 = \begin{pmatrix} x_2 E_p \\ 0 \\ 0 \\ -x_2 Ep \end{pmatrix} \tag{5.17}$$

with x as the fraction of the proton energy E_p must be extremely asymmetric. The created jets themselves,

$$j_1 = \begin{pmatrix} E_j \\ E_j/\cosh\eta \\ 0 \\ E_j \tanh\eta \end{pmatrix} \text{ and } j_2 = \begin{pmatrix} E_j \\ -E_j/\cosh\eta \\ 0 \\ E_j \tanh\eta \end{pmatrix} \tag{5.18}$$

are assumed to have the same energy E_j and the same pseudorapidity η. Due to energy and momentum conservation $p_1 + p_2 = j_1 + j_2$ this leads to

$$\begin{aligned} (x_1 + x_2)E_p &= 2E_j \\ (x_1 - x_2)E_p &= 2E_j \tanh\eta \end{aligned} \Rightarrow \begin{aligned} x_1 &= E_j/E_p(1 + \tanh\eta) \\ x_2 &= E_j/E_p(1 - \tanh\eta) \end{aligned} \quad \text{or} \quad x_2 = x_1 \frac{1 - \tanh\eta}{1 + \tanh\eta}. \tag{5.19}$$

With a minimum jet p_T of 3 GeV the jet energy is $E_j \geq 605\,\mathrm{GeV}$ at $\eta = 6$. For a center-of-mass energy of 13 TeV in a proton-proton collision ($E_p = 6.5\,\mathrm{TeV}$) it follows that the

momentum fraction of the first parton x_1

$$x_1 \geq \frac{605\,\text{GeV}}{6.5\,\text{TeV}}(1+\tanh 6) = 0.186 \tag{5.20}$$

and of the second parton

$$x_2 \geq 0.186\frac{1-\tanh 6}{1+\tanh 6} = 1.14\times 10^{-6}. \tag{5.21}$$

At this value the parton interactions are completely dominated by the density of gluons. On the other hand, effects such as ISR and FSR will still weaken the correlation between two jets, whereas MPI and remnant fragmentation will have an important impact, too. In the following figures the $\Delta\phi$ distribution of two reconstructed jets in CASTOR is shown and compared to different model prediction. Also different PDFs using the PYTHIA8 Tune CUETP8M1 are tested.

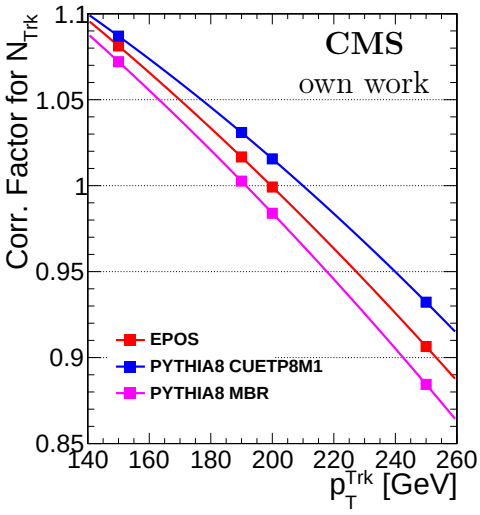


Figure 5.20: Factor to correct from reconstructed tracks in the number of charged final state particles in the range $|\eta| < 2$ depending on the minimum p_T of the generated particles. At four p_T values the ratio of N_{trk} to generated particles is determined and fitted with a second degree polynomial.

In the following, the results regarding the correlation of two jet events in CASTOR are presented with bin-by-bin correction, which is acceptable since the migration in ϕ is much smaller compared to the one in p_T. Corrections to a stable particle level are performed by comparing variables of reconstructed jets with the hadron level jets for different models. As in the previous chapter the uncertainty on the CASTOR jet energy scale is taken into account with 15 %. At hadron level only jets are counted which are inside $-6.55 < \eta < -5.25$ and have a $p_T > 3\,\text{GeV}$. For this choice the correction factor for reconstructed two jet events has an uncertainty smaller than $\approx 5\,\%$. Furthermore, in the following figures

the number of central tracks is used as an estimator of the number of charged particles at generator level. This is used to classify events according to the centrality of the collision. In figure 5.20 the correction factor between tracks and generated charged particles in the range $|\eta| < 2$ is shown depending on the p_T cut on the generated particles. Since there was no magnetic field in CMS during data taking the track p_T was not measured. One advantage of this is that the normal magnetic particle cutoff of around 100 MeV does not apply, thus also very low-p_T particles can be reconstructed. Figure 5.20 shows for three different models that for a $p_T > 200$ Mev of generated particles the correction factor becomes one. The variation of this procedure between the different models is around 5 %, and it is taken into account as systematic uncertainty of the measurement. If the results are divided by luminosity, the additional 4 % luminosity uncertainty is added to the total uncertainty of the presented results.

For dijet events one can see in figure 5.21 the ϕ correlation of the two jets in CASTOR. In the left panel the data is compared to different models at cross section level, while in the right panel the result is divided by the total observed number of two jet events in CASTOR. The upper left of figure 5.21 shows that the different PYTHIA8 tunes are roughly comparable to the data. The behaviour of the different tunes is very similar but the PYTHIA8 CUETP8M1 tune using the NNPDF2.3 PDF has a higher cross section compared to the other PYTHIA8 CUETP8S1 tunes with different PDFs. Also the 4C tune of PYTHIA8 using the CTEQ6L1 PDF is at the same level as the PYTHIA8 CUETP8S1 tunes. When looking at the upper right of figure 5.21 one can clearly see that the shape of the $\Delta\phi$ distribution is equal for all PYTHIA8 tunes. This shows that the different PDF sets have no influence on the actual shape of the dijet correlation but on the total cross section of produced dijet events. Compared to the data, the different PYTHIA8 tunes generate more collinear jets with $\Delta\Phi$ around π compared to the distribution in data which is more flat. This indicates that in data the overall production of jets by MPI in the forward direction is more pronounced than in PYTHIA8. Comparing the cosmic ray model QGSJetII to data in the lower right of figure 5.21 the difference is much stronger than for PYTHIA8. Since in QGSJetII MPI effects are neglected it generates much more collinear dijet events compared to data. EPOS-LHC, on the contrary, includes strong effects of MPI, and describes the flat shape of the data very well. SIBYLL2.3 is similar to PYTHIA8, showing a slight enhancement of opposite dijets compared to data. At cross section level, the cosmic ray models including the QGSJetII model have a significant spread, but are all comparable to data inside the uncertainties.

Since the $\Delta\Phi$ distribution of the dijet events in CASTOR is still influenced by MPI or by the beam remnants, one should also observe a dependence of the number of dijet events on the hardness of the proton-proton collision. In order to more than roughly test the region that is most sensitive to the gluon PDF the cross section for dijet production is explicitly measured for the back-to-back topology. The "background" from MPI and similar effects are not subtracted. To account for the hardness of the proton-proton collision the number of events with two jets in CASTOR and a $\Delta\Phi > \pi/2$ is shown in figure 5.22 as function of the central tracks ($|\eta| < 2$) multiplicity. In the left column the distributions are normalized

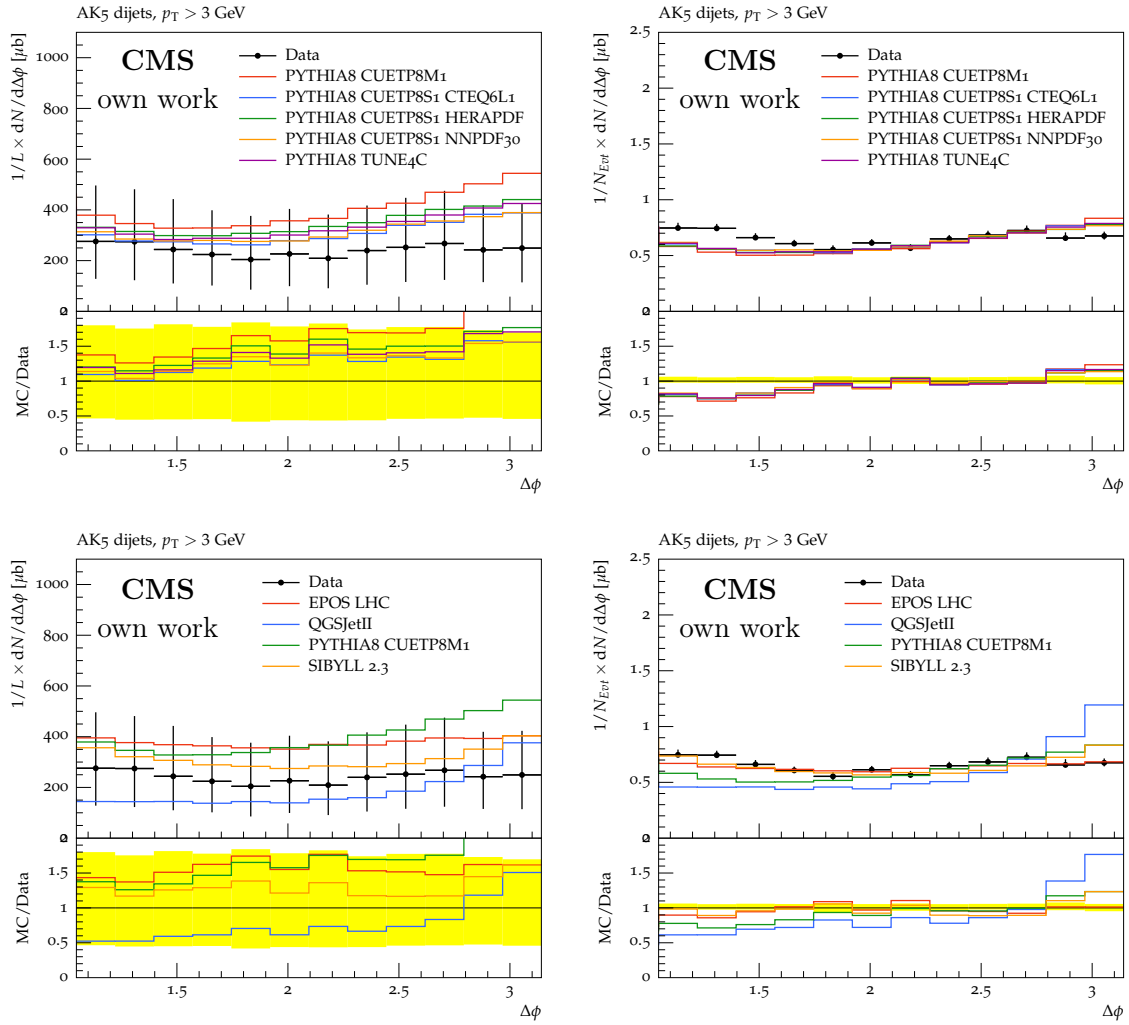


Figure 5.21: Jet $\Delta\phi$ distribution for events with two jets in CASTOR. On the left the distribution is normalized by the luminosity while on the right it is normalized to the number of events with a $\Delta\phi > \pi/2$.

to the luminosity and in the right to the event yield, while in the top row PYTHIA8 tunes are compared to data and in the bottom row to cosmic ray models. The different PYTHIA8 tunes follow the data at low central tracks multiplicity but while in the data a peak is reached at around $N_{\text{trk}} \approx 35$, the PYTHIA8 models have the peak at around $N_{\text{trk}} \approx 55$. Also in this representation the $\Delta\phi$ distribution of PYTHIA8 CUETP8M1 has an overall higher cross section compared to the other PYTHIA8 tunes. On the top right panel of figure 5.22 the shape of the distribution for the PYTHIA8 tunes is compared to data, it shows that no tune is able to describe the data. Here the number of dijet events is peaking in data at lower track multiplicity than predicted by the PYTHIA8 model also. All the different tunes have a very similar shape and the different PDF sets show no major impact

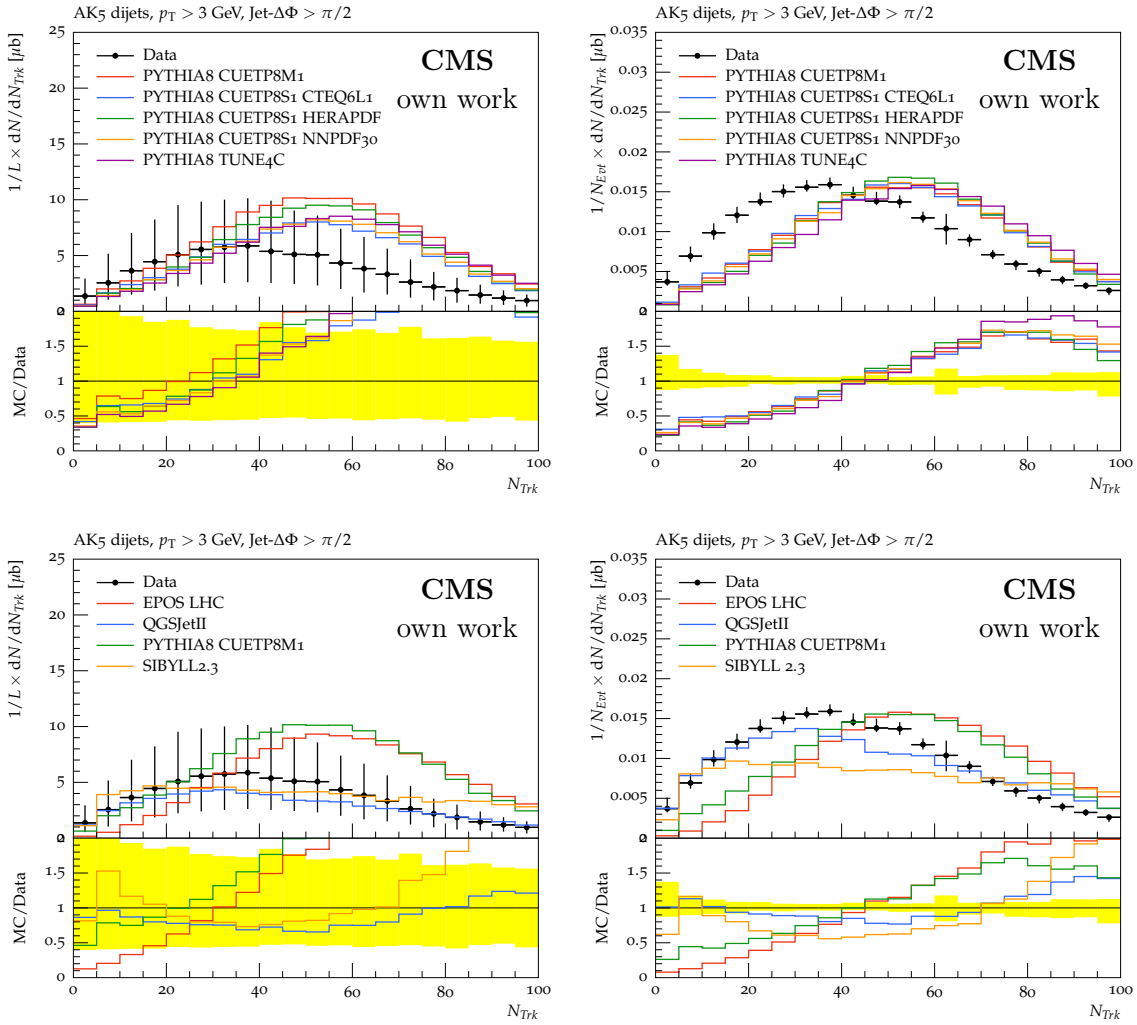


Figure 5.22: Number of events with two jets in CASTOR and $\Delta\Phi > \pi/2$ depending on the number of central tracks ($|\eta| < 2$). On the left the distribution is normalized to the luminosity and on the right to the total number of selected events.

on the prediction. Only the PYTHIA8 tune 4C is a little bit flatter than the others at high track multiplicities. EPOS-LHC on the lower left-hand corner of figure 5.22 indicates a higher cross section for dijet events at high central track multiplicities than in data while at low multiplicities the data is underestimated. The QGSJetII model, instead, describes the data relatively well despite the bad description of the $\Delta\phi$ distribution by QGSJetII in figure 5.21. SIBYLL2.3 shows almost no dependency on the track multiplicity or on the collision hardness except for very low multiplicities in the first bin. In the lower right of figure 5.22 only QGSJetII is close to the data.

The $\Delta\phi$ distribution of two jets is explicitly shown in figure 5.23 for a low central track multiplicity on the left ($N_{\mathrm{trk}} < 10$) and for a high track multiplicity ($N_{\mathrm{trk}} > 80$) on the right.

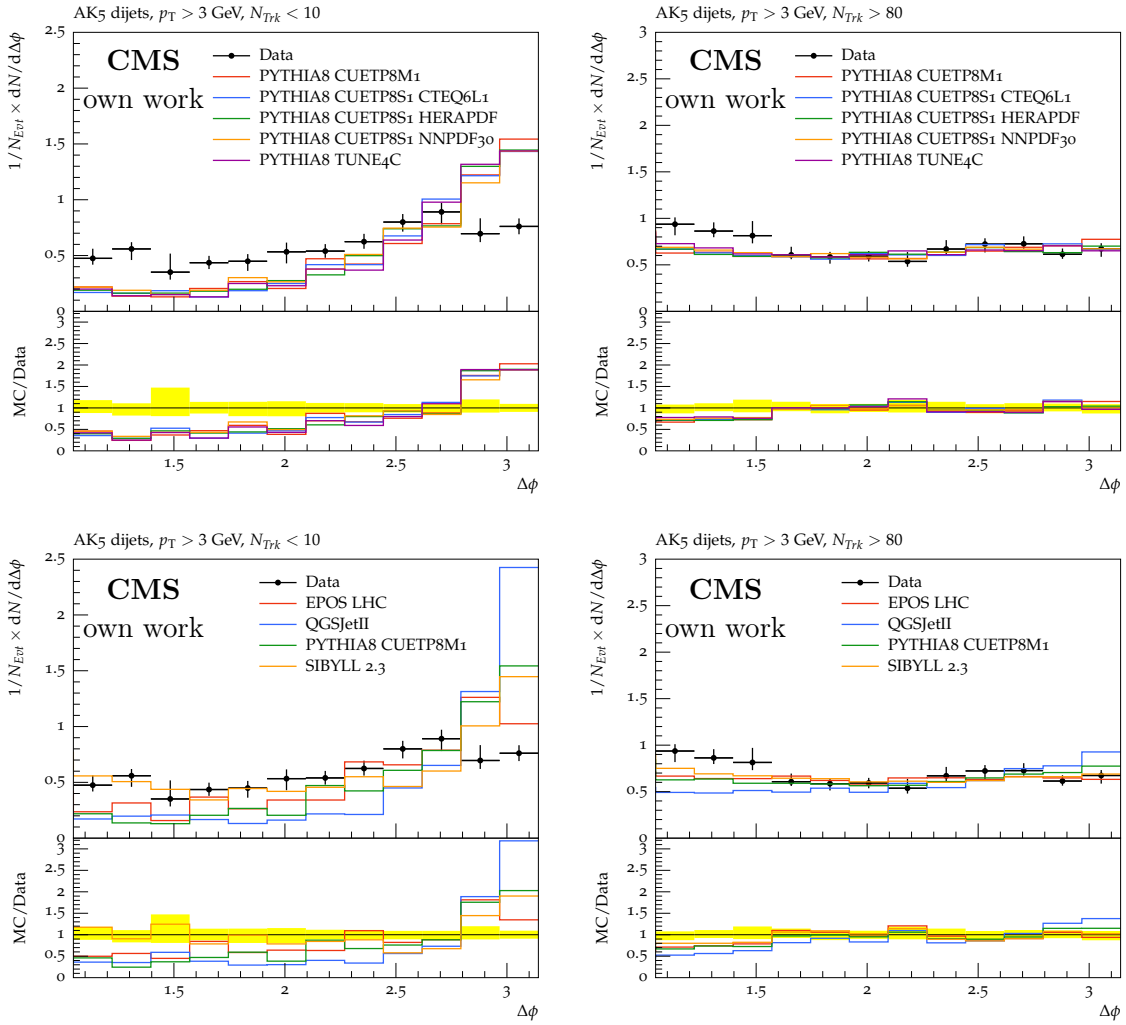


Figure 5.23: Jet $\Delta\phi$ distribution for events with two jets in CASTOR normalized to the number of events with a $\Delta\phi > \pi/2$. The two figures on the left show the distribution only for events where the number of central tracks ($|\eta| < 2$) is smaller than 10 while on the right side the distribution is shown only for events with more than 80 central tracks.

The distributions are normalized to the total dijet event yield. For less hard collisions with $N_{\text{trk}} < 10$ the correlation of opposite jet events in $\Delta\phi$ is strongly pronounced in the models compared to the data. The data implies an enhancement of back-to-back jet events in this event class. The different PYTHIA8 tunes shown in the top left panel of figure 5.23 are all extremely similar to each other. On the lower left of figure 5.23 only SIBYLL2.3 is close to the data but also predicts slightly more back-to-back jet events than EPOS-LHC and the PYTHIA8 tune CUETP8M1. QGSJETII has for $N_{\text{trk}} < 10$ a very pronounced peak at $\Delta\phi \approx \pi$. This overestimates the measurements by a large margin. For very hard collisions

with $N_{\mathrm{trk}} > 80$ MPI becomes a more important effect and flattens the distributions for data as well as for the different models. At this point no enhanced back-to-back jet production is visible in CASTOR anymore. All models with the exception of QGSJetII agree with this observation.

6 Summary

In this thesis the first measurement of very forward jets with CASTOR in CMS has been performed. Such jets can probe partons down to values of $x_{\text{Bjorken}} = 10^{-6}$. This is unique at LHC. The main outcomes are the inclusive jet spectrum as well as a study of events with exactly two reconstructed jets in CASTOR. It was found that these data are very sensitive to low-x soft QCD effects like MPI, beam remnant and high-order radiation. The relationship of these jets to the parton distribution functions at low-x was explicitly investigated. The data was interpreted using simulation performed with PYTHIA8 using different PDFs as well as different tunes, and furthermore cosmic ray models like EPOS-LHC and QGSJETII. Some of the core results of this measurement were presented at the DIS 2016 conference in Hamburg.

To measure jets in the very forward direction with the CASTOR detector a good reconstruction in a wide range of jet energies is needed. Since the deposited energy in the very forward direction can become very high, up to substantial portions of the beam energy, the saturation of the recorded channel energy can be a major problem. This is caused by the digitization of the signal which is restricted to a certain signal range. The energy at which a channel saturates differs from channel to channel because it depends on parameters such as gain efficiency and high voltage. In this thesis I have developed an algorithm to correct the saturation effect in data:

- To reconstruct the signal of a saturated channel the pulse shape of the PMT is used. It was shown that at high energies the shape is very stable and the signal tail can be used to reconstruct the desaturated signal.
- By analysing the shape at different times it was found that the values for this desaturation algorithm are relatively stable.
- The replacement of several PMTs in 2012 showed a major impact on the particular pulse reconstruction values.
- This new desaturation algorithm was included into the standard CMS event reconstruction.
- The needed parameters for the CMS database were provided as part of this thesis for all operation periods of CASTOR from 2010 to 2016.

Furthermore, another major development of this thesis was the development and commissioning of the first physics trigger with the CASTOR detector. Such triggers were used

to trigger the CMS experiment in 2013, 2015 and 2016. The following triggers based on CASTOR were developed to record special events:

- For photon/electron detection in CASTOR a trigger was developed which selects events requiring isolated clusters in the electromagnetic region of the detector. With a special trigger signal L1SA the trigger information was synchronized between CMS and TOTEM and made it possible to record and reconstruct data with the information of both experiments.
- The electron trigger was tested and a mismatch of channels was observed in the trigger software. This was repaired and the trigger was used again in the next run period.
- Also, a trigger to detect jets was developed to record events with very high energy jets in the very forward detector of CASTOR.
- With this trigger around a thousand times more event statistics at high jet energies could be recorded compared to the unbiased data set. The efficiency of the jet trigger reached the 100 % plateau at around 2 TeV of jet energy.

In particular with the CMS-TOTEM electron trigger it was possible to record events with single electrons in CASTOR. These events were used for the first and so far unique alignment measurement of CASTOR:

- The precise track reconstruction with the TOTEM T2 detector directly in front of the CASTOR detector was used for this purpose.
- To determine the uncertainty a bootstrap method was developed which is based on the measured data itself.
- The x and y shifts as well as the ϕ rotation of the two independent halves of CASTOR have been measured with a precision of about 1 mm and 1 degree. This is so far the most precise measurement of the CASTOR position in CMS.

For the reconstruction of the jets the anti-k_t algorithm was used at detector level. The radius parameter of $R = 0.5$ in the anti-k_t algorithm was chosen because it fits well with the CASTOR granularity. The same parameter was also chosen for the definition of jets at hadron level.

This made CASTOR capable not only to measure very forward jets in a wide energy range but also to unfold the distribution to a hadron level definition of jets to compare it with different MC models. It is a very important precondition for any jet measurement to determine the jet energy scale. A method to achieve this was developed within this thesis. It was found that a jet needs a typical correction of about 40 % to correct for the noncompensation of the calorimeter. For the inclusive jet spectrum the unfolding method according to D'Agostini was used, which is based on the Bayes theorem. All important sources of systematic uncertainties have been propagated to the final results. The uncertainties are dominated by the energy scale uncertainty of CASTOR.

It was shown that within the experimental uncertainties all model predictions of the inclusive jet production are comparable with the data. In order to enhance the sensitivity of the analysis the data was also normalized by the jet yield. This is more sensitive to the shape of the distribution. It was found that EPOS-LHC and QGSJetII predict a slightly softer spectrum than the data suggest. Moreover, the effect of switching off multi-parton interactions in Pythia8 is clearly incompatible with the measurement.

To increase the sensitivity to low-x partons exclusive dijet events in CASTOR were measured as well. Parton scattering with very asymmetric values of x_{Bjorken} produce back-to-back jet topologies in the very forward direction. The following observations were made:

- For the $\Delta\phi$ distribution the EPOS-LHC model shows a very good agreement with the data while QGSJetII shows a too strong correlation between the two jets, which is much stronger than in the data. This indicates that in data MPI is more dominant since EPOS accounts for it and in QGSJetII these effects are not completely implemented. Moreover, the predicted shapes in $\Delta\phi$ from different PDFs in Pythia8 are very similar. But a change of the overall dijet cross section is visible. For the $\Delta\phi$ distribution also a small enhancement at $\Delta\phi = \pi$ is predicted by the models.
- To distinguish between the hardness of the different proton-proton collisions, the number of correlated dijet in a $\Delta\phi \geq \pi/2$ window were measured depending on the number of charged particles in the central tracker. Here the QGSJetII model shows the best description of the data compared to the other models. EPOS-LHC and Pythia8 reach their maximum later in the number of produced central tracks than seen in the measurement. Here, too, the different PDF sets used in Pythia8 have a minor effect on the shape of the distribution.
- The $\Delta\phi$ of the two jets in the low and the high central multiplicity region represent entirely different regimes. For very low multiplicity the correlation of the two jets is clearly seen in data and for the models, while the models still overestimate this effect. At very high multiplicity and therefore for very hard proton-proton collisions these correlated jets vanish since here radiation and underlying event become a major effect.

It is another important observation that while the ϕ-shape of jets, and also global event properties in CASTOR are well described by the simulations, the development of jets when they penetrate the calorimeter are significantly different in simulations compared to the data. The hadronic event generators obviously need to improve their description of very forward jet fragmentation. The relative importance of "underlying event" and "minijet-like" particle production is not correct in any of the available models. In particular, it is absolutely possible, but it remains to be demonstrated that this observation will have an important impact on the interpretation of ultra-high energy air shower measurements.

The CASTOR detector was installed in CMS to take 13 TeV data only in the very first few weeks of the low luminosity phase during the startup of LHC Run-2. During this time the CMS magnet was not working, which limits the physics reach of CASTOR data since

correlations and combined analysis with normal central CMS objects are very difficult. If CASTOR were taking more data with functioning CMS magnet, the analysis presented in this thesis would reach its ultimate performance. However, this is not foreseen.

This thesis presents the first results on very forward jets at a center-of-mass energy of 13 TeV. The unique place of the CASTOR detector in CMS at $\eta \approx -6$ made it possible to probe proton-proton collisions at a very low-x by using jets. It was proved that the CASTOR detector is capable of reconstructing jets in this phase space and at very high energies. It was found that all models describe the inclusive jet production within experimental uncertainties. Another finding was that by requiring exactly two very forward jets with back-to-back topology and a level of extra activity in the rest of CMS, one has a very powerful tool to test models. The sensitivity to the low-x parton dynamics is maximized and the importance of the underlying event is minimized.

7 Bibliography

[1] Rick D. Field. The Underlying event in hard scattering processes. *eConf*, C010630: P501, 2001.

[2] H. Abramowicz et al. Summary of the Workshop on Multi-Parton Interactions (MPI@LHC 2012). 2013.

[3] Vardan Khachatryan et al. Event generator tunes obtained from underlying event and multiparton scattering measurements. *Eur. Phys. J.*, C76(3):155, 2016. doi:10.1140/epjc/s10052-016-3988-x.

[4] Vardan Khachatryan et al. Measurement of the underlying event activity using charged-particle jets in proton-proton collisions at sqrt(s) = 2.76 TeV. *JHEP*, 09: 137, 2015. doi:10.1007/JHEP09(2015)137.

[5] Serguei Chatrchyan et al. Jet and underlying event properties as a function of charged-particle multiplicity in proton-proton collisions at $\sqrt{s} = 7$ TeV. *Eur. Phys. J.*, C73(12):2674, 2013. doi:10.1140/epjc/s10052-013-2674-5.

[6] Morad Aaboud et al. 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*, 03:157, 2017. doi:10.1007/JHEP03(2017)157.

[7] Georges Aad et al. Measurement of the underlying event in jet events from 7 TeV proton-proton collisions with the ATLAS detector. *Eur. Phys. J.*, C74(8):2965, 2014. doi:10.1140/epjc/s10052-014-2965-5.

[8] Serguei Chatrchyan et al. Study of the underlying event at forward rapidity in pp collisions at $\sqrt{s} = 0.9$, 2.76, and 7 TeV. *JHEP*, 04:072, 2013. doi:10.1007/JHEP04(2013)072.

[9] Jan Ebr, Petr Necesal, and Jan Ridky. Soft Particle Production in Very High Energy Hadron Interactions. *Astropart. Phys.*, 90:37–49, 2017. doi:10.1016/j.astropartphys.2017.02.001.

[10] Ralf Ulrich, Ralph Engel, and Michael Unger. Hadronic Multiparticle Production at Ultra-High Energies and Extensive Air Showers. *Phys. Rev.*, D83:054026, 2011. doi:10.1103/PhysRevD.83.054026.

[11] R. D. Parsons, C. Bleve, S. S. Ostapchenko, and J. Knapp. Systematic uncertainties in air shower measurements from high-energy hadronic interaction models. *Astropart. Phys.*, 34:832–839, 2011. doi:10.1016/j.astropartphys.2011.02.007.

[12] Laura Collica. Measurement of the Muon Production Depths at the Pierre Auger Observatory. *Eur. Phys. J. Plus*, 131(9):301, 2016. doi:10.1140/epjp/i2016-16301-6.

[13] Alexander Aab et al. Muons in air showers at the Pierre Auger Observatory: Mean number in highly inclined events. *Phys. Rev.*, D91(3):032003, 2015. doi:10.1103/PhysRevD.91.059901, 10.1103/PhysRevD.91.032003. [Erratum: Phys. Rev.D91,no.5,059901(2015)].

[14] Alexander Aab et al. Evidence for a mixed mass composition at the ankle in the cosmic-ray spectrum. *Phys. Lett.*, B762:288–295, 2016. doi:10.1016/j.physletb.2016.09.039.

[15] Pedro Abreu et al. Interpretation of the Depths of Maximum of Extensive Air Showers Measured by the Pierre Auger Observatory. *JCAP*, 1302:026, 2013. doi:10.1088/1475-7516/2013/02/026.

[16] K. A. Olive et al. Review of Particle Physics. *Chin. Phys.*, C38:090001, 2014. doi:10.1088/1674-1137/38/9/090001.

[17] Serguei Chatrchyan et al. Measurement of the ratio of the inclusive 3-jet cross section to the inclusive 2-jet cross section in pp collisions at $\sqrt{s} = 7$ TeV and first determination of the strong coupling constant in the TeV range. *Eur. Phys. J.*, C73(10): 2604, 2013. doi:10.1140/epjc/s10052-013-2604-6.

[18] B. Povh, K. Rith, C. Scholz, and F. Zetsche. *Teilchen und Kerne: Eine Einführung in die physikalischen Konzepte.* Springer Berlin Heidelberg, 2004. ISBN 978-3-642-37821-8. doi:10.1007/978-3-642-37822-5.

[19] R. P. Feynman. The behavior of hadron collisions at extreme energies. *Conf. Proc.*, C690905:237–258, 1969.

[20] Raymond Brock et al. Handbook of perturbative QCD: Version 1.0. *Rev. Mod. Phys.*, 67:157–248, 1995. doi:10.1103/RevModPhys.67.157.

[21] John C. Collins, Davison E. Soper, and George F. Sterman. Factorization of Hard Processes in QCD. *Adv. Ser. Direct. High Energy Phys.*, 5:1–91, 1989. doi:10.1142/9789814503266_0001.

[22] Francis Halzen and Alan Martin. *Quark & Leptons: An introductory course in modern particle physics.* John Wiley & Sons, New York, 1984. ISBN 978-0-471-88741-6.

[23] H. Abramowicz and A. Caldwell. HERA collider physics. *Rev. Mod. Phys.*, 71: 1275–1410, 1999. doi:10.1103/RevModPhys.71.1275.

[24] Curtis G. Callan, Jr. and David J. Gross. High-energy electroproduction and the constitution of the electric current. *Phys. Rev. Lett.*, 22:156–159, 1969. doi:10.1103/PhysRevLett.22.156.

[25] F. D. Aaron et al. Combined Measurement and QCD Analysis of the Inclusive e+- p Scattering Cross Sections at HERA. *JHEP*, 01:109, 2010. doi:10.1007/JHEP01(2010)109.

[26] V. N. Gribov and L. N. Lipatov. Deep inelastic e p scattering in perturbation theory. *Sov. J. Nucl. Phys.*, 15:438–450, 1972. [Yad. Fiz.15,781(1972)].

[27] Guido Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl. Phys.*, B126:298–318, 1977. doi:10.1016/0550-3213(77)90384-4.

[28] Yuri L. Dokshitzer. Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics. *Sov. Phys. JETP*, 46:641–653, 1977. [Zh. Eksp. Teor. Fiz.73,1216(1977)].

[29] L. N. Lipatov. Reggeization of the Vector Meson and the Vacuum Singularity in Nonabelian Gauge Theories. *Sov. J. Nucl. Phys.*, 23:338–345, 1976. [Yad. Fiz.23,642(1976)].

[30] E. A. Kuraev, L. N. Lipatov, and Victor S. Fadin. The Pomeranchuk Singularity in Nonabelian Gauge Theories. *Sov. Phys. JETP*, 45:199–204, 1977. [Zh. Eksp. Teor. Fiz.72,377(1977)].

[31] I. I. Balitsky and L. N. Lipatov. The Pomeranchuk Singularity in Quantum Chromodynamics. *Sov. J. Nucl. Phys.*, 28:822–829, 1978. [Yad. Fiz.28,1597(1978)].

[32] P. Cipriano, S. Dooling, A. Grebenyuk, P. Gunnellini, F. Hautmann, H. Jung, and P. Katsas. Higgs boson as a gluon trigger. *Phys. Rev.*, D88(9):097501, 2013. doi:10.1103/PhysRevD.88.097501.

[33] F. D. Aaron et al. Inclusive Deep Inelastic Scattering at High Q^2 with Longitudinally Polarised Lepton Beams at HERA. *JHEP*, 09:061, 2012. doi:10.1007/JHEP09(2012)061.

[34] H. Abramowicz et al. Measurement of high-Q2 neutral current deep inelastic e^+p scattering cross sections with a longitudinally polarized positron beam at HERA. *Phys. Rev.*, D87(5):052014, 2013. doi:10.1103/PhysRevD.87.052014.

[35] Voica Radescu. Combination and QCD analysis of the HERA inclusive cross sections. *PoS*, ICHEP2010:168, 2010.

[36] A. M. Cooper-Sarkar. PDF Fits at HERA. *PoS*, EPS-HEP2011:320, 2011.

[37] S. Alekhin, J. Blumlein, and S. Moch. Parton Distribution Functions and Benchmark Cross Sections at NNLO. *Phys. Rev.*, D86:054009, 2012. doi:10.1103/PhysRevD.86.054009.

[38] V. M. Abazov et al. Measurement of the electron charge asymmetry in $p\bar{p} \to W + X \to e\nu + X$ events at $\sqrt{s} = 1.96$-TeV. *Phys. Rev. Lett.*, 101:211801, 2008. doi:10.1103/PhysRevLett.101.211801.

[39] V. M. Abazov et al. Measurement of the muon charge asymmetry from W boson decays. *Phys. Rev.*, D77:011106, 2008. doi:10.1103/PhysRevD.77.011106.

[40] Hung-Liang Lai, Marco Guzzi, Joey Huston, Zhao Li, Pavel M. Nadolsky, Jon Pumplin, and C.-P. Yuan. New parton distributions for collider physics. *Phys. Rev.*, D82:074024, 2010. doi:10.1103/PhysRevD.82.074024.

[41] Jun Gao, Marco Guzzi, Joey Huston, Hung-Liang Lai, Zhao Li, Pavel Nadolsky, Jon Pumplin, Daniel Stump, and C.-P. Yuan. CT10 next-to-next-to-leading order global analysis of QCD. *Phys. Rev.*, D89(3):033009, 2014. doi:10.1103/PhysRevD.89.033009.

[42] Pavel Nadolsky, Jun Gao, Marco Guzzi, Joey Huston, Hung-Liang Lai, Zhao Li, Jon Pumplin, Dan Stump, and C.-P. Yuan. Progress in CTEQ-TEA PDF Analysis. In *Proceedings, 20th International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS 2012): Bonn, Germany, March 26-30, 2012*, pages 417–420, 2012. doi:10.3204/DESY-PROC-2012-02/301. URL `https://inspirehep.net/record/1118434/files/arXiv:1206.3321.pdf`. [,417(2012)].

[43] Georges Aad et al. Measurement of the inclusive $W^{\pm}$ and Z/gamma cross sections in the electron and muon decay channels in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Phys. Rev.*, D85:072004, 2012. doi:10.1103/PhysRevD.85.072004.

[44] Georges Aad et al. Measurement of inclusive jet and dijet production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector. *Phys. Rev.*, D86:014022, 2012. doi:10.1103/PhysRevD.86.014022.

[45] Serguei Chatrchyan et al. Measurement of the electron charge asymmetry in inclusive W production in pp collisions at $\sqrt{s} = 7$ TeV. *Phys. Rev. Lett.*, 109:111806, 2012. doi:10.1103/PhysRevLett.109.111806.

[46] R Aaij et al. Inclusive W and Z production in the forward region at $\sqrt{s} = 7$ TeV. *JHEP*, 06:058, 2012. doi:10.1007/JHEP06(2012)058.

[47] Richard D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015. doi:10.1007/JHEP04(2015)040.

[48] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur. Phys. J.*, C63:189–285, 2009. doi:10.1140/epjc/s10052-009-1072-5.

[49] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne. Parton distributions in the LHC era: MMHT 2014 PDFs. *Eur. Phys. J.*, C75(5):204, 2015. doi:10.1140/epjc/s10052-015-3397-6.

[50] CMS Collaboration. Measurement of the inclusive jet cross section in pp collisions at sqrt(s) 2.76 TeV. 2015.

[51] Bo Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. Parton Fragmentation and String Dynamics. *Phys. Rept.*, 97:31–145, 1983. doi:10.1016/0370-1573(83)90080-7.

[52] Donald H. Perkins. *Introduction to high energy physics*. Cambridge University Press, New York, 1999.

[53] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008. doi:10.1016/j.cpc.2008.01.036.

[54] Richard Corke and Torbjorn Sjostrand. Interleaved Parton Showers and Tuning Prospects. *JHEP*, 03:032, 2011. doi:10.1007/JHEP03(2011)032.

[55] Studies of Diffractive Enhanced Minimum Bias Events in ATLAS. 2010.

[56] Track-based underlying event measurements in pp collisions at sqrts = 900 GeV and 7 TeV with the ATLAS Detector at the LHC. 2010.

[57] K. Aamodt et al. Charged-particle multiplicity measurement in proton-proton collisions at $\sqrt{s} = 7$ TeV with ALICE at LHC. *Eur. Phys. J.*, C68:345–354, 2010. doi:10.1140/epjc/s10052-010-1350-2.

[58] Robert Ciesielski and Konstantin Goulianos. MBR Monte Carlo Simulation in PYTHIA8. *PoS*, ICHEP2012:301, 2013.

[59] Peter Skands, Stefano Carrazza, and Juan Rojo. Tuning PYTHIA 8.1: the Monash 2013 Tune. *Eur. Phys. J.*, C74(8):3024, 2014. doi:10.1140/epjc/s10052-014-3024-y.

[60] Richard D. Ball et al. Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO. *Nucl. Phys.*, B855:153–221, 2012. doi:10.1016/j.nuclphysb.2011.09.024.

[61] Richard D. Ball et al. Parton distributions with QED corrections. *Nucl. Phys.*, B877:290–320, 2013. doi:10.1016/j.nuclphysb.2013.10.010.

[62] T. Pierog and K. Werner. EPOS Model and Ultra High Energy Cosmic Rays. *Nucl. Phys. Proc. Suppl.*, 196:102–105, 2009. doi:10.1016/j.nuclphysbps.2009.09.017.

[63] S. Ostapchenko. QGSJET-II: Towards reliable description of very high energy hadronic interactions. *Nucl. Phys. Proc. Suppl.*, 151:143–146, 2006. doi:10.1016/j.nuclphysbps.2005.07.026.

[64] H. J. Drescher, M. Hladik, S. Ostapchenko, T. Pierog, and K. Werner. Parton based Gribov-Regge theory. *Phys. Rept.*, 350:93–289, 2001. doi:10.1016/S0370-1573(00)00122-8.

[65] Adrian Dumitru, Kevin Dusling, Francois Gelis, Jamal Jalilian-Marian, Tuomas Lappi, and Raju Venugopalan. The Ridge in proton-proton collisions at the LHC. *Phys. Lett.*, B697:21–25, 2011. doi:10.1016/j.physletb.2011.01.024.

[66] Jean-Luc Caron. Overall view of LHC experiments.. Vue d'ensemble des experiences du LHC. AC Collection. Legacy of AC. Pictures from 1992 to 2002., May 1998. URL http://cds.cern.ch/record/841555.

[67] Lyndon Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001, 2008. doi:10.1088/1748-0221/3/08/S08001.

[68] AC Team. Diagram of an LHC dipole magnet. Schéma d'un aimant dipôle du LHC. Jun 1999. URL http://cds.cern.ch/record/40524.

[69] CERN Bulletin. Powering CERN and the LHC. Energie pour le LHC et le CERN. (BUL-NA-2011-016. 05/2011):6, Jan 2011. URL https://cds.cern.ch/record/1324541.

[70] Cinzia De Melis. The CERN accelerator complex. Complexe des accélérateurs du CERN. Jul 2016. URL http://cds.cern.ch/record/2197559. General Photo.

[71] S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008. doi:10.1088/1748-0221/3/08/S08004.

[72] CMS Collaboration. Detector Drawings. CMS Collection., Mar 2012. URL https://cds.cern.ch/record/1433717.

[73] Siona Ruth Davis. Interactive Slice of the CMS detector. Aug 2016. URL http://cds.cern.ch/record/2205172.

[74] G. Anelli et al. The TOTEM experiment at the CERN Large Hadron Collider. *JINST*, 3:S08007, 2008. doi:10.1088/1748-0221/3/08/S08007.

[75] Ekaterina Kuznetsova. Performance and Calibration of CASTOR Calorimeter at CMS. *Phys. Procedia*, 37:356–363, 2012. doi:10.1016/j.phpro.2012.02.386.

[76] Results on CASTOR Performance during LHC Run 2. Feb 2016. URL https://cds.cern.ch/record/2133152.

[77] Michael Eliachevitch. Simulation of the energy response of the CASTOR very forward calorimeter at the CMS experiment. 2015.

[78] Peter Göttlicher. Design and test beam studies for the CASTOR calorimeter of the CMS experiment. *Nucl. Instrum. Meth.*, A623:225–227, 2010. doi:10.1016/j.nima.2010.02.203.

[79] CMS Collaboration. Measurement of the pseudorapidity and centrality dependence of the very forward energy density in PbPb collisions at sqrt s=2.76 ATeV. 2012.

[80] Hauke Wöhrmann for the CMS Collaboration. Studies of the nuclear stopping power in PbPb collisions at 2.76 TeV with CMS. Technical Report CMS-CR-2012-256, CERN, Geneva, Oct 2012. URL http://cds.cern.ch/record/1486004.

[81] Darin Acosta. CMS Trigger Improvements Towards Run II. *Nucl. Part. Phys. Proc.*, 273-275:1008–1013, 2016. doi:10.1016/j.nuclphysbps.2015.09.158.

[82] Mirko Berretti, Stefano Lami, and Angelo Scribano. *Measurement of the forward charged particle pseudorapidity density in pp collisions at* $\sqrt{s} = 7$ *TeV with the TOTEM experiment.* PhD thesis, Siena U., Sep 2012. URL https://cds.cern.ch/record/1508392. Presented 27 Nov 2012.

[83] Hauke Wöhrmann. Alignment of CASTOR with 2013 pA TOTEM-Data. Mar 2014. URL https://cds.cern.ch/record/1697694.

[84] F. James and M. Roos. Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations. *Comput. Phys. Commun.*, 10:343–367, 1975. doi:10.1016/0010-4655(75)90039-9.

[85] S. V. Chekanov. Jet algorithms: A Minireview. In *Hadron collider physics. Proceedings, 14th Topical Conference, HCP 2002, Karlsruhe, Germany, September 29-October 4, 2002*, pages 478–486, 2002. URL http://alice.cern.ch/format/showfull?sysnb=2351331.

[86] Gerald C. Blazey et al. Run II jet physics. In *QCD and weak boson physics in Run II. Proceedings, Batavia, USA, March 4-6, June 3-4, November 4-6, 1999*, pages 47–77, 2000. URL http://lss.fnal.gov/cgi-bin/find_paper.pl?conf-00-092.

[87] S. Catani, Yuri L. Dokshitzer, M. H. Seymour, and B. R. Webber. Longitudinally invariant K_t clustering algorithms for hadron-hadron collisions. *Nucl. Phys.*, B406: 187–224, 1993. doi:10.1016/0550-3213(93)90166-M.

[88] Stan Bentvelsen and Irmtraud Meyer. The Cambridge jet algorithm: Features and applications. *Eur. Phys. J.*, C4:623–629, 1998. doi:10.1007/s100520050232.

[89] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008. doi:10.1088/1126-6708/2008/04/063.

[90] Gavin P. Salam and Gregory Soyez. A Practical Seedless Infrared-Safe Cone jet algorithm. *JHEP*, 05:086, 2007. doi:10.1088/1126-6708/2007/05/086.

[91] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Catchment Area of Jets. *JHEP*, 04:005, 2008. doi:10.1088/1126-6708/2008/04/005.

[92] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl. Instrum. Meth.*, A506: 250–303, 2003. doi:10.1016/S0168-9002(03)01368-8.

[93] Matteo Cacciari, Juan Rojo, Gavin P. Salam, and Gregory Soyez. Quantifying the performance of jet definitions for kinematic reconstruction at the LHC. *JHEP*, 12: 032, 2008. doi:10.1088/1126-6708/2008/12/032.

[94] Matteo Cacciari and Gavin P. Salam. Dispelling the N^3 myth for the k_t jet-finder. *Phys. Lett.*, B641:57–61, 2006. doi:10.1016/j.physletb.2006.08.037.

[95] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet User Manual. *Eur. Phys. J.*, C72:1896, 2012. doi:10.1140/epjc/s10052-012-1896-2.

[96] Hauke Wöhrmann for the CMS Collaboration. Measurement of the very forward inclusive jet cross section in pp collisions at $\sqrt{s} = 13$ TeV with CMS. Technical Report CMS-PAS-FSQ-16-003, CERN, Geneva, 2016. URL `http://cds.cern.ch/record/2146006`.

[97] G. D'Agostini. A Multidimensional unfolding method based on Bayes' theorem. *Nucl. Instrum. Meth.*, A362:487–498, 1995. doi:10.1016/0168-9002(95)00274-X.

[98] Francesco Spano. Unfolding in particle physics: a window on solving inverse problems. *EPJ Web Conf.*, 55:03002, 2013. doi:10.1051/epjconf/20135503002.

[99] Andreas Hocker and Vakhtang Kartvelishvili. SVD approach to data unfolding. *Nucl. Instrum. Meth.*, A372:469–481, 1996. doi:10.1016/0168-9002(95)01478-0.

[100] Volker Blobel. Unfolding Methods in Particle Physics. In *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN, Geneva, Switzerland 17-20 January 2011*, pages 240–251, Geneva, 2011. CERN, CERN. doi:10.5170/CERN-2011-006.240. URL `http://inspirehep.net/record/1478292/files/1087459_240-251.pdf`.

[101] Tim Adye. Unfolding algorithms and tests using RooUnfold. In *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN, Geneva, Switzerland 17-20 January 2011*, pages 313–318, Geneva, 2011. CERN, CERN. doi:10.5170/CERN-2011-006.313. URL `http://inspirehep.net/record/898599/files/arXiv:1105.1160.pdf`.

[102] Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-FSQ-15-005, CERN, Geneva, 2016. URL `https://cds.cern.ch/record/2145896`.

Acknowledgements

First of all, I would like to thank Prof. Dr. Thomas Müller for provide me the possibility and the time to graduate at the KIT. In addition I want to thank Prof. Dr. Johannes Blümer for volunteering as my co-referee for this work. As part of the doctoral fellow program, KSETA, I am very grateful for the support as well as for the lectures and talks organized by my referees and others.

Furthermore, I owe special thanks to my supervisor Dr. Ralf Ulrich, who invited me to be part of his Helmholtz Young Investigator group. I am especially grateful for his endless and patient help, day and night. In this group I also received a lot of support from my colleagues Dr. Colin Baus, Dr. Igor Katkov, Melike Akbyik and Sebastian Baur. At CERN the whole CASTOR group at CMS was of great assistance in any situation. The communication with the members of the Pierre Auger group at KIT helped me a lot to extend my knowledge and I am grateful I could always rely on their help. At last, I want would like to thank my family for standing by my side, also in difficult situations in my life during this time.

Declaration

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

Karlsruhe, July 2017

Hauke Wöhrmann