Aalto University
School of Science
Degree Programme of Applied Physics

Juska Pekkanen

Jets in CMS Experiment

Master's Thesis
Espoo, February 19, 2013

Supervisor: Professor Mikko Alava
Instructor: Mikko Voutilainen Dr.Sc.(Tech.), PhD

Aalto University
School of Science
Degree Programme of Applied Physics

ABSTRACT OF
MASTER'S THESIS

Author:	Juska Pekkanen		
Title:	Jets in CMS Experiment		
Date:	February 19, 2013	**Pages:**	69
Professorship:	Applied physics	**Code:**	F3005
Supervisor:	Professor Mikko Alava		
Instructor:	Mikko Voutilainen Dr.Sc.(Tech.), PhD		

Jets are collimated showers of particles originating from quarks and gluons, which are the constituents of protons and neutrons that make up the atomic nuclei. In the CERN's Large Hadron Collider (LHC) protons are collided with the highest energies ever achieved and production of jets is ubiquitous in these collision events. Measuring energies of jets is a complex process and requires sophisticated jet energy calibration methods. In this Master's Thesis a new jet composition driven method for enhancing jet calibration in the Compact Muon Solenoid experiment (CMS) is studied. We study the effects of sensitivity of different detector elements to the jet energy composition and try to find sources of observed discrepancies between composition of Monte Carlo simulated and measured jets. We test three different mis-calibration scenarios in the lightweight FastSim simulation environment and observe encouraging results that are in agreement with the hypothesized mis-calibrations. The FastSim approach proves to be a useful tool for investigating the role of detector calibration on jet composition, at least in the case of hadronic and electromagnetic calorimeters. In the near future the developed method will be applied also to other parts of CMS.

Keywords:	experimental particle physics, high energy physics, LHC, jets, jet energy scale, CMS
Language:	English

Aalto-yliopisto
Perustieteiden korkeakoulu
Tietotekniikan tutkinto-ohjelma

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä:	Juska Pekkanen		
Työn nimi:	Jetit CMS-kokeessa		
Päiväys:	19. helmikuuta 2013	**Sivumäärä:**	69
Professuuri:	Teknillinen fysiikka	**Koodi:**	F3005
Valvoja:	Professori Mikko Alava		
Ohjaaja:	Mikko Voutilainen TkT, FT		

Jetit ovat kvarkeista ja gluoneista alkunsa saavia kartiomaisia hiukkasryöppyjä. Protonit ja neutronit, siis atomiytimien rakenneosaset, koostuvat kvarkeista ja gluoneista. Euroopan hiukkastutkimuskeskus CERN:n Suuressa Hadronitörmäyttimessä LHC:ssa törmäytetään protoneja toisiinsa suurilla energioilla ja törmäyksissä syntyy runsaasti kyseisiä hiukkasryöppyjä. Näiden jettien energian mittaaminen on monimutkainen prosessi, joka vaatii pitkälle kehitettyjä kalibrointitekniikoita. Tässä diplomityössä tutkitaan uutta jettien energioiden tarkempaan mittaukseen tähtäävää tekniikkaa, joka perustuu ryöppyjen koostumuksen tutkimiseen. LHC:n CMS-kokeen puitteissa suoritettavassa tutkimuksessa mittaamme CMS:n eri havaintojärjestelmien herkkyystasojen vaikutusta jettien koostumukseen ja etsimme syitä simuloitujen ja mitattujen hiukkasryöppyjen koostumuksien eroihin. Tutkimme kolmen eri herkkyysskenaarion vaikutusta jetteihin käyttäen laskennallisesti kevyttä FastSim-simulaatioympäristöä. Saamamme tulokset ovat yhteensopivia oletettujen virhelähteiden kanssa ja rohkaisevat jatkotutkimuksiin. Käyttämämme menetelmä osoittautuu hyödylliseksi työkaluksi havaintojärjestelmien herkkyyden ja hiukkasryöppyjen koostumuksen välisen yhteyden tutkimiseen. Kehitettyä menetelmää aiotaan tulevaisuudessa soveltaa laajemmin CMS-kokeen eri mittausjärjestelmiin.

Asiasanat:	kokeellinen alkeishiukkasfysiikka, alkeishiukkasfysiikka, LHC, CERN, hiukkasryöpyt, jetit, CMS-koe
Kieli:	Englanti

Acknowledgements

First of all I wish to express my gratitude to all the colleagues at the HIP CMS project for keeping up the supporting and inspiring atmosphere where I have felt safe to misunderstand, ask for help and share my views on my first steps on the rocky road of becoming a particle physicist.

Without the unlimited patience and astonishing expertise of my instructor Mikko Voutilainen this thesis would not exist. Mikko's encouraging and comforting way of guiding me is a key reason why I wish to pursue my career in physics.

I want to thank Paula Eerola and Kati Lassila-Perini from the project coordination for making my thesis possible in terms of finances. I am especially grateful to Kati for making it possible for me to try my wings in seminars and workshops abroad and in Finland. I would also like to thank my colleague Matti Kortelainen for offering a helping hand whenever I have faced technical difficulties.

My dear parents Tiina and Jari have always given me all the support on whatever path I have taken. Thank you mom and dad for stretching your understanding even when I chose not to study the laws of humans but the laws of nature. I also wish to thank my brothers Jami and Jyri for brotherly care and constructive criticism. I have been delighted to observe a flicker of interest towards my research from my family when I have forced them to listen to my talks.

Lastly I want to thank my dear girlfriend Sari for being there for me and for helping me to remember how much more there is to life than particle physics. Having something to look forward to is what makes my life full. With you we never run out of ideas for yet another adventure to anticipate.

Helsinki, February 19, 2013

Juska Pekkanen

Acronyms and Symbols

CERN	European Organization for Nuclear Research
CMS	Compact Muon Solenoid
CMSSW	CMS Software framework
ECAL	Electromagnetic Calorimeter
HCAL	Hadronic Calorimeter
HF	Forward Hadronic Calorimeter
IP	Interaction Point
JEC	Jet Energy Corrections
LHC	Large Hadron Collider
MC	Monte Carlo
PF	Particle Flow Event Reconstruction Algorithm
SM	Standard Model
WLCG	World Wide LHC Computing Grid

E_T^{miss}	Missing transverse energy
p_T	Transverse momentum
p_T^{reco}	Reconstructed jet transverse momentum
p_T^{gen}	Monte Carlo generated jet transverse momentum
$\hat{p}_\mathrm{T}$	Transverse momentum of a hard $2 \to 2$ scattering subprocess

Contents

Chapter 1

Introduction

Nature has an unparalleled ability to astonish physicists by behaving in ways that no-one ever imagined. A good example of her capability lies in the discovery of quarks and their interactions, which are the initiators of jet physics. Intuition and experience tell that the force between two bodies feeling the same interaction weakens when they are drawn further away from each other. This is evidently true for gravitation and electromagnetism, and weak interaction affects only particles that basically touch each other. Thinking of a force that does not act this way would be ridiculous, would it not? It may be so, but this is how Nature behaves for quarks, the fundamental constituents of protons and neutrons that build up atomic nuclei. The fourth fundamental force of nature, *strong force* that binds quarks together does not dilute with distance but becomes dramatically stronger. When two bodies bound together by the strong force are drawn apart, they feel increasing attraction. If they are given enough momentum, as in particle colliders, the initial quarks can be separated, and a quark–anti-quark pair is created from the potential energy of the bond. Quarks can never exist alone, they are *confined*.

The phenomenon of confinement is confirmed in simulations and observed in collider experiments, although an exhaustive theoretical description waits to be formulated. In experiments the manifestation of confinement is seen in collimated sprays of particles, *jets*, which are produced in repeating hadronization processes.

Jets are ubiquitous especially at the Large Hadron Collider (LHC) where protons, packages of quarks and gluons, are collided. As jets are produced in the majority of collisions, knowing the energies carried away in them is essential for deep understanding of the event as a whole. The process where signals in different detector elements are assigned to the true energy of the jet, and hence to the initial quark or gluon, is called jet energy calibration.

In this Master's Thesis jet energy calibration in the Compact Muon

Solenoid (CMS) experiment is studied and methods for more accurate jet energy determination are searched for. In previous studies we have seen indications that sensitivity of some parts of the CMS particle detector may not be in their optimal level and this is a potential source of uncertainty in jet energy measurements. We test this hypothesis with simulations and ultimately aim to see if the previously observed discrepancies between observations and Monte Carlo simulations can be corrected with fine-tuning sensitivities of certain detector elements. To our knowledge this approach has not been used before for jet energy calibrations in CMS and we believe that this analysis has a high potential to enhance the calibration. An improvement of the jet energy calibration would yield a significant increase to the accuracy of many CMS physics analyses. Jet energy calibration is an essential part in the search for better understanding of Nature.

1.1 Motivation

The Standard Model (SM) of particle physics is the most successful theoretical structure describing the sub-atomic world to date, and the predictions of the SM agree with measurements with an unprecedented accuracy. Nevertheless, by no means can we say that it is the final and fundamental theory of nature; there are problems within the SM and also many phenomena that it simply does not account for.

Arguably the strongest motivation for building the LHC is the Higgs boson. Without this the Standard Model is not complete and should it not exist, the theory is undermined at its very basis. LHC is designed so that if the Higgs mechanism is part of our universe, the particle will be found. Solving the origin of mass would be the most remarkable milestone in physics in decades.

One of the biggest mysteries that the SM does not address is the domination of matter over anti-matter. Assuming the Big Bang model, all the matter and anti-matter was created from pure energy some 14 billion years ago. Conservation laws say that when matter is created from energy, there should be equal amounts of matter and anti-matter. To our understanding this is not a property of our universe. The LHC provides conditions where this problem called the *baryon asymmetry* can be studied.

According to the prevailing quark model, quarks are point-like particles without any inner structure. However this is just a hypothesis and as have happened many times in the history of physics, indivisible particles prove to be composites. Searches for substructure of quarks are possible at the LHC, thanks to the highest collision energies ever achieved.

These were three unsolved mysteries of Mother Nature that the LHC can and will shed light on, and the list could go on. Funding of expensive scientific projects from public funds traditionally stirs discussion, especially when direct applications are not yet known. While it may be difficult to see directly, it is a historical fact that scientific discoveries tend to spawn applications useful for all mankind.

1.2 Structure of the Thesis

After this introductory chapter, the theoretical landscape for the thesis is briefly reviewed in the second chapter. There a summary of the Standard Model is first given and then some main tools of experimental particle physicist are described, and in the end the protagonists of this study, jets, are introduced.

The experimental set-up, namely the Large Hadron Collider and the Compact Muon Solenoid experiment, are described in chapter three, and in the fourth chapter the process of jet reconstruction in the CMS experiment is explained.

In the fifth chapter the analysis chain is reviewed and the main method, FastSim parameter variation, is introduced. The results of this study are then presented in chapter six, and the main findings, drawbacks and limitations of the research are discussed in chapter seven, where also future prospects of the subject are reflected. The thesis is recapitulated chapter eight, where also the role of jet physics in the field of high energy physics today and tomorrow is discussed.

1.3 Definitions and terminology

In this section some used conventions and notations are introduced in no particular order.

Coordinate system

In CMS a right-handed coordinate system is defined so that x points along the radial acceleration of the particles in the accelerator ring, y points upwards with respect to the plane of the accelerator and z along the counterclockwise particle beam.

Angles in the plane perpendicular to the particle beam (in $x-y$-plane) are denoted by ϕ, and angles in the $y-z$-plane by θ. In cylindrical coordinate

systems these are called azimuthal and polar angles, respectively.

Event

An event means all the phenomena initiated by one proton bunch crossing near the origin of the detector's coordinate system. An event that is worth recording usually contains a hard interaction process between (parts of) two protons, but also numerous softer collisions, called pile-up interactions, are typically present. Duration of an event is in the scale of nanoseconds.

Transverse momentum p_T

Particles circulating in different directions in the LHC have momentum only in z-direction with a very good approximation. As a result the total energy and momentum in the plane of the cross section of the beam line, the transverse plane, is conserved. For this, the transverse momentum p_T is a convenient conserved quantity, formally defined as

$$p_T = \sqrt{p_x^2 + p_y^2}, \tag{1.1}$$

where p_x and p_y are respectively the momentum components along the x- and y-axes.

Rapidity y

In special relativity the concept of rapidity is used to measure speeds of relativistic particles, but in accelerator physics it is used as a measure of angle with respect to the beam line. Rapidity is defined as

$$y = \frac{1}{2}\ln\frac{E + p_z c}{E - p_z c}, \tag{1.2}$$

where p_z is particle's momentum in the direction of the beam line. For reasons of convenience, rapidity is often substituted by *pseudorapidity* η, that approaches y when particle's speed approaches c and its mass energy is much smaller than its total energy.

Pseudorapidity η

In high energy physics the concept of pseudorapidity is often used to measure the scattering angles of particles with respect to the beam axis, and also when

speaking of coverage of detectors. Pseudorapidity is defined as

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right), \tag{1.3}$$

where θ is the angle from the positive z-axis. This at first sight cumbersome definition has the advantage that the number N of particles flying to one unit in η is nearly constant throughout the detector, i.e. $\frac{dN}{d\eta} \approx$ constant.

Integrated luminosity $\int Ldt$

Integrated luminosity is used as a measure of the recorded data, and can be calculated by integrating the instantaneous luminosity L over time. For a storage ring as the LHC the following holds:

$$\int Ldt = \int fn\frac{N^2}{A}dt, \tag{1.4}$$

where f is the revolution frequency, n is the number of bunches in each beam, N is the number of particles in each bunch and A is the cross section of the beams at the moment of interaction. The unit used for integrated luminosity is inverse barn, $b^{-1} = (10^{-28}\ \mathrm{m}^2)^{-1}$.

Chapter 2

Theory

In this second chapter essential background information for the work is presented with an introduction to the current understanding of the sub-atomic world (section 2.1) and a foreword on experimental methodology is given in section 2.2. Experimentalists' important tools called Monte Carlo simulations together with event generators and detector simulations are introduced in sections 2.3 and 2.4. Finally in section 2.5 the objects under examination in this thesis, jets, and their production process are introduced.

2.1 The Standard Model

The branch of experimental physics called high energy physics (also called experimental particle physics) studies the properties of the smallest constituents of matter: elementary particles. The theoretical model that best summarizes the current knowledge and most accurately describes the observed phenomena is called the Standard Model of particle physics.

In this section the particles of the Standard Model are first introduced, and then the forces acting on these particles and theories behind them are named, together with references to detailed descriptions and original articles.

2.1.1 Elementary particles

Currently fifteen elementary particles are known and found in experiments, and the number doubles if also corresponding *anti*-particles are counted. These particles are typically divided to three categories: *quarks, leptons*, and *gauge bosons*.

Quarks are the fundamental constituents of protons and neutrons which in turn constitute the atomic nuclei. Every proton and neutron, or nucleon

for short, are made up of three quarks: two 'up'-quarks and one 'down' (uud) in the case of a proton, and one 'up' and two 'downs' (udd) in a neutron. For reasons yet to be understood, there are also two other 'generations' of quarks: the second generation in the quark family consists of 'charm' and 'strange' quarks (c and s) and the third is made up of 'top', t and 'bottom', b (also called 'beauty'). Masses, electric charges and *spins* of quarks are presented in figure 2.1.

The second category of elementary particles, leptons, are named after the Greek word *leptos* meaning something small and thin [17]. The most familiar and ubiquitous leptons are the electrons, which surround the atomic nuclei and together with up and down quarks form all the matter in the known universe. The electron, e, has a counterpart without which it never could exist: the electron neutrino, ν_e. It is an extremely light and elusive particle; its mass is measured to be less than one hundred thousandth of the electron mass, and it interacts only via the *weak interaction*. The lepton number (N_l) seems to be conserved in our universe and an electron ($N_l = 1$) can be created only if an electron anti-neutrino ($N_l = -1$) is born in the same process.

As it is for the quarks, there are also three generations of leptons: as an addition to the electron family (e and $\bar{\nu_e}$), there is the muon family (μ and $\bar{\nu_\mu}$) and the tau family (τ and $\bar{\nu_\tau}$). For masses (or measured upper limits), charges and spins, see figure 2.1.

The third particle category is responsible for the interactions between the nucleons and leptons introduced above: the gauge bosons are intermediators of forces. Starting from the upper right corner of figure 2.1, photons mediate the electromagnetic attraction or repulsion between charged particles. Gluons are mediators of the strong interaction that 'glues' the quarks tightly together and are also responsible for keeping nuclei of elements stable. The Z^0 and $W^{\pm}$ give rise to the weak interaction that causes radioactive decay and makes hydrogen fusion possible in stars such as the sun. [2]

One more elementary particle is foreseen by the Standard Model: the Higgs boson. It is discussed in a few lines in the following section.

2.1.2 Interactions

There is a rigorous and predictive theory, a *relativistic quantum field theory*, describing each of these fundamental forces of nature. For electromagnetic force the theory is called quantum electrodynamics, QED for short, and it is a relativistic extension to the classical theory of electromagnetism that was put together by James C. Maxwell. [9]

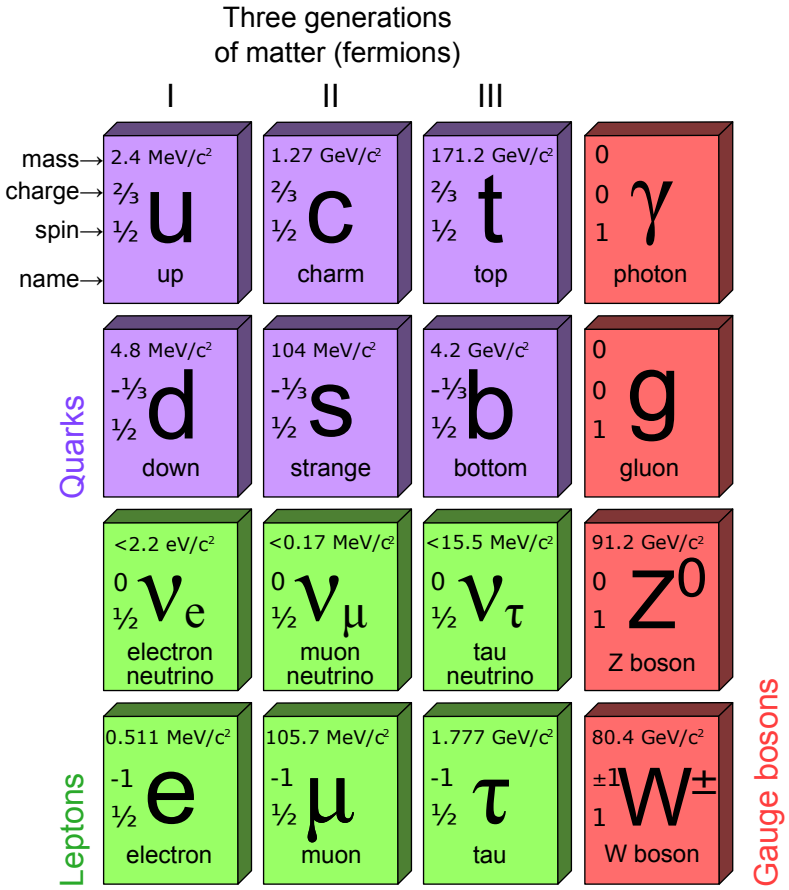

Figure 2.1: Experimentally observed elementary particles.[26]

The description of the weak force is provided by the *electroweak interaction*, and it is sometimes also called the Glashow-Weinberg-Salam model for its developers. This force is an extension of the QED to contain also reactions of weak interaction. Glashow, Weinberg and Salam found a way to combine the seemingly very different realms of QED and weak interactions to a single theory, and showed that the only difference is the mass of the corresponding force carriers. Masses of the Z^0 and $W^{\pm}$ bosons are responsible for the weakness of weak interactions. [12]

Forces between quarks are described by quantum chronodynamics, QCD, in which quarks gluons carry *color charge* and make strong binds to other colored particles. QCD gives rise to phenomena called *color confinement* and *asymptotic freedom*, which are the main reasons explaining why we observe

particle jets in detectors, and thus are of high importance for this Master's Thesis. We well return to this aspect of the QCD later in section 2.5, when the jet creation process is discussed. [14]

There is still one vital part of the Standard Model that was only briefly mentioned in section 2.1.1, the Higgs boson. For solving the puzzle of why especially Z^0 and $W^{\pm}$ bosons are massive, several theorists proposed a mechanism of *electroweak symmetry breaking* that explains why some particles are massive and others are not. This is more commonly called the Higgs mechanism and the elusive Higgs boson is the messenger particle of the hypothesised Higgs field which explains the masses. Strong evidence for the Higgs particle was presented in the summer of 2012 at CERN, and the Standard Model is about to be completed with the last missing piece. [4, 10, 13, 15]

It is worth mentioning that gravity is not part of the Standard Model and its effects are usually negligible in the realm of high energy physics.

2.2 Testing a theory

Distribution of work in the field of particle physics is as follows: theoretical physicists develop a multitude of possible scenarios of how the nature might work, and experimental physicists build experiments to directly see how the world behaves. Although the experiments are inspired by theoretical predictions, a big part of the discoveries have surprised the whole community. Experiments can be viewed as filters able to refute false theories and confirm or support viable models of nature. Extracting experimentally testable predictions from theories is a non-trivial process. Here, in between the theorists and experimentalists, are phenomenologists that turn theories into observable predictions.

Theories mentioned in section 2.1 are products of theorists. In this section the approach of experimentalists is described, or at least an attempt to communicate the spirit of the used methodology is made.

2.2.1 Nature's probabilistic nature

With the rise of quantum mechanics it started to become clear to physicists that nature can not be thought nor modeled as an entirely deterministic system. As the quantum theory, a theory that describes how things work in the sub-atomic world, is able to give only probabilistic answers, it is clear that high energy physics has to face the same restrictions and limitations.

Where quantum mechanics can tell that a particle x is in a box with the probability P_{box}, a particle physicist can tell that particles x and y are

created in a head-on collision of two z:s with the probability $P_{z+z\to x+y}$. For this reason, experimental particle physics is a profoundly statistical science. For the same reason billions of collisions are needed for claiming discoveries and preliminary results have to be considered cautiously.

Smashing protons together with the speed of light can be thought as a coin tossing experiment: the result can never be known beforehand, but each outcome has a certain probability. In collider experiments these probabilities are measured in terms of *cross-sections* that phenomenologists calculate from theories and experimentalists strive to measure.

2.2.2 Cross-sections

In the LHC protons are circulating in both directions in tightly squeezed bunches of billions of protons. For examining the process when two such bunches cross, a bunch crossing, let us consider the case in the rest coordinates of one bunch in the interaction point. Now we can study the situation as a fixed target experiment without losing generality.

If we count the rate of protons penetrating a unit area inside the target and perpendicular to the colliding beam, we have the flux, J. Flux can be expressed as a product of particle density n_b and velocity v_b of the beam with respect to the target:

$$J = n_b v_b. \tag{2.1}$$

Let N be the number of particle in the fixed target, and we can write the reaction rate W_r for an arbitrary reaction r as

$$W_r = JN\sigma_r, \tag{2.2}$$

where the coefficient σ_r represents the probability for the reaction r to occur. This measure of probability is called the cross-section for r and, as can be seen from equation 2.2, has the dimension of an area. For historical reasons cross-sections are measured in *barns*, b, and one barn is equal to 10^{-28} m^2.

Terms that contain dependence of geometry of the beam and target and their densities are often combined to one measurable quantity, luminosity L:

$$W_r = L\sigma_r. \tag{2.3}$$

Luminosity is used in high energy physics as a measure of collision rate; when L is integrated over time we get the *integrated luminosity* that conveniently absorbs the collision rate, total run time and level of bunch compression to a single quantity with a dimension of inverse barns, b^{-1}.

Cross-sections for different processes can on one hand be calculated from the theories and on the other hand measured in a particle collider experiment,

such as the CMS. Basically cross-sections can be estimated in a particle detector simply by deducing from observed number of events how many reactions r occurred in a run of integrated luminosity $\int_0^t Ldt$.

To summarize the previous, calculating and measuring cross-sections provides an interface to compare the theoretical predictions to real world phenomena, and it is the key method in many analyses. [18]

2.3 Monte Carlo event generation

Monte Carlo (MC) methods are methods that use random numbers for calculating approximate results. With the Monte Carlo scheme analytically unsolvable problems can be numerically solved to an arbitrary precision, the common example being difficult integrals. The origin of the name of the method lies in the famous Monte Carlo casino of Monaco; casinos are places where probabilities and randomness are at the essence.

One of the numerous application areas of the MC methods is particle physics. MC event generation is a process where the possible outcomes of a particular collision can be calculated, given that the beam properties and cross-sections are known. Using this approach simulated events from any predictive theory, such as the Standard Model, can be produced. Schematic view of how an event could look like at the generator level is in figure 2.2.

In CMS two most widely used MC event generators are PYTHIA and HERWIG++. While PYTHIA is the reference MC generator in the majority of analyses, it is important to have other generators with different underlying assumptions and techniques for cross-checking the results of the simulations. Exhaustive descriptions of these tools are given in references [1, 21, 22].

2.4 Detector simulation

As the Monte Carlo event generators produce simulated events that basically take place in vacuum without any disturbing force fields nor bulk material of the detectors, the simulation clearly is not yet sufficient for mimicking real world events. This is where detector simulations come to the picture. Detector simulation is an extremely detailed and computationally demanding procedure, where effects of all the aspects of the detector and surrounding circumstances are added to the generated events. This includes for example the strong magnetic field, noise in the detector electronics, effects of electronic wires and so on. Detector simulation includes naturally also all the subdetectors which are introduced in chapter 3. Figure 2.3 illustrates a generated

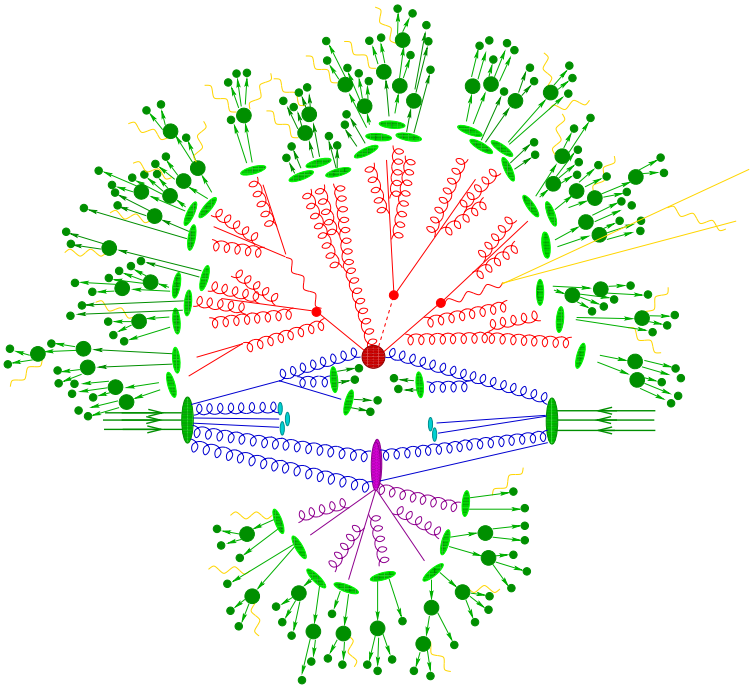

Figure 2.2: A sketch of a proton-proton collision in an event generator. The ellipses with three arrows represent the colliding protons. The hard scattering process is illustrated in the top part of the figure, where final state particles are produced via parton showering (red spirals), hadronization (green ellipses) and finally decays of unstable particles (big green disks). The lower part of the sketch represents a process arising from the underlying event.[20]

collision inside a simulated CMS detector.[11]

At the CMS experiment full detector simulation is done with the Geant4 simulation platform, which also incorporates Monte Carlo methods. The Geant4 package is composed of nearly half a million lines of source code and follows the open source philosophy as do PYTHIA and HERWIG++. More information on the platform and the Geant4 collaboration behind it can be found in reference [19].

The simulations done specifically for this study were run in the FastSim simulation environment, which is a computationally lighter version of the full CMS software (CMSSW) that is normally used for simulations and reconstruction. The platform corresponding to the Geant4 in the FastSim environment is called the GFlash. Physics results of GFlash and Geant4 are essentially similar, but GFlash uses parameterizations instead of individual particle tracking when modeling particle showers.

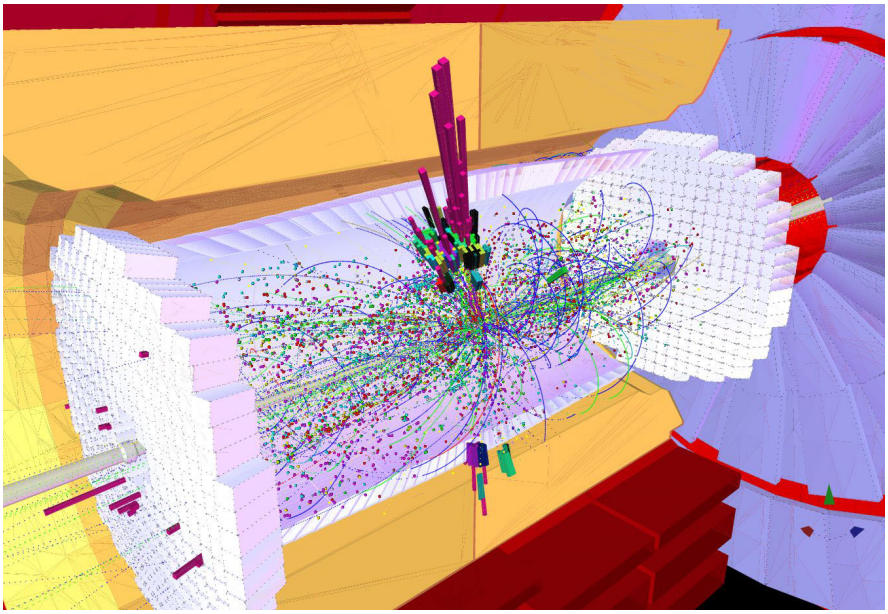

Figure 2.3: A generated event in simulated CMS detector. [11]

2.5 Hadronization and Jet production

The LHC collides protons which are made out of quarks and the force carriers gluons. As mentioned in section 2.1.2, quarks and gluons are assigned with an additional quantum number, imaginatively named the *color* or *color charge*. Experiments have shown that colored particles never seem to exist by themselves, but are always in a bound state with another colored particles. This experimental fact has guided the theory of QCD to include color confinement, a rule that says that no colored particle can exist alone, but must be accompanied with a particle or particles that make the composite particle colorless or white. In practice this means that every proton comprises one red, one blue and one green quark and, in analog to light rays, appears white. Mesons, on the other hand, are particles that consist of one quark and one anti-quark. Quarks in mesons have to form colorless combination; one color and a corresponding anti-color, for example a π^+-meson can be a bound state of an anti-blue down anti-quark ($\bar{d}_{\overline{blue}}$) and a blue up-quark ($u_{blue}$). QCD explains this extraordinary property of color interactions by assigning color charges also to gluons. The mediators of the strong force then also interact with each other, and this property is believed to be the origin of confinement. Colored force carrier is something that electro-weak

interactions do not have and hence confinement is not part of that theory.

Another aspect of QCD that was mentioned earlier is the asymptotic freedom. The rule of asymptotic freedom states that quarks and gluons can move in freedom as long as they stay in vicinity of the other colored particles that make the configuration colorless. Here 'vicinity' means distances of the order of 10^{-15} meters, which is roughly the diameter of a nucleon.

When color confinement and asymptotic freedom are taken into the realm of protons colliding at the speed of light, two observations can be made instantly: firstly, quarks and gluons, also called *partons*, inside protons are free to collide with partons of the other proton. Secondly, even when two quarks collide head-on with energy corresponding to rest-mass energy of thousands of protons, they can not fly out freely because of confinement. What really happens here is *hadronization*, also known as *fragmentation*. When a parton – be it a sea or valence quark or a gluon – gets high momentum to a direction away from the particle(s) that make the package colorless, a new particle–anti-particle pair is created solely from the energy of the strong confining bond, and a new *color singlet* is formed and is free to exit the system. In high-energy collisions the initial color singlets are often in a *virtual state* and decay further to hadrons until energies of single hadrons are low enough for forming a real particle. In this process numerous hadrons are created from one initial parton.

This phenomenon is pictured in the Lund string model as if there was a rubber band between the initial partons and when the stretching force is strong enough, the rubber-band snaps and a new particle pair is created solely from the stored energy of the bond. Illustrations of this model is presented in figures 2.4 and 2.5.

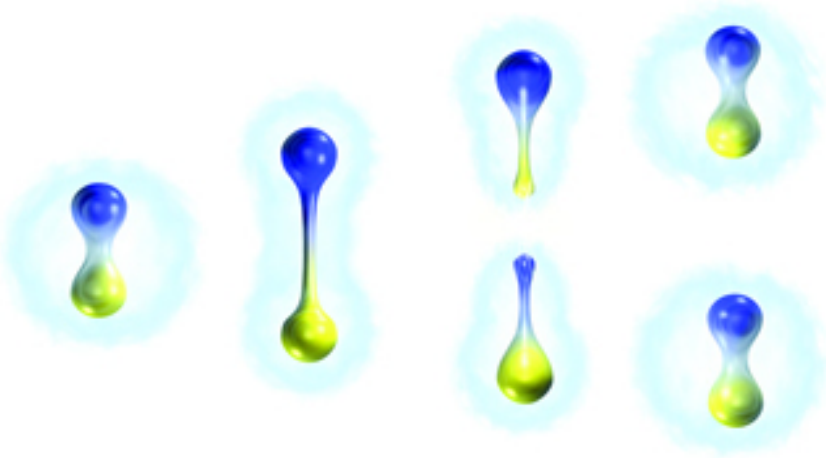

Figure 2.4: An illustration of the hadronization process in the Lund string model. Different colors refer to opposite charges. [23]

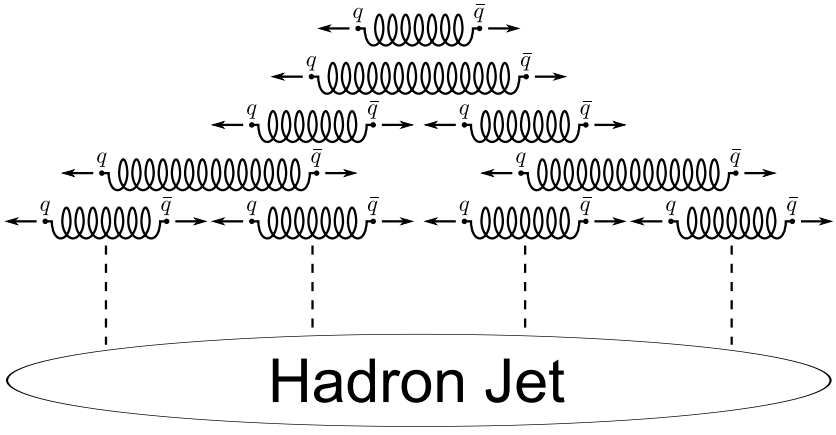

Figure 2.5: Another illustration of the Lund rubber-band hadronization idea. [26]

When a composite particle is born in the hadronization process, it almost never is stable: none of the known mesons are stable, and just one of the *baryons*, the composites of three quarks, is stable. The particles born in the high energy collision are thus quite prone to decay to other particles, all the way to stable particles: photons, electrons and protons. In high energy experiments, though, it is possible to observe particles that live just a tiny fraction of a second; with the speed of light, it takes less than ten nanoseconds to traverse the CMS tracker and calorimeters. For example charged pions and kaons live long enough to be measured, but still many of the particles born in the hadronization process undergo decay before even hitting the first layers of the tracker just a few centimeters away from the interaction point. A typical jet in the barrel region of the cylindrical CMS detector contains around 10 to 100 particles before secondary particles are born in interactions with detector materials.

Figure 2.6 illustrates the jet production process from the hard collision to the signals in the calorimeters via the hadronization process.

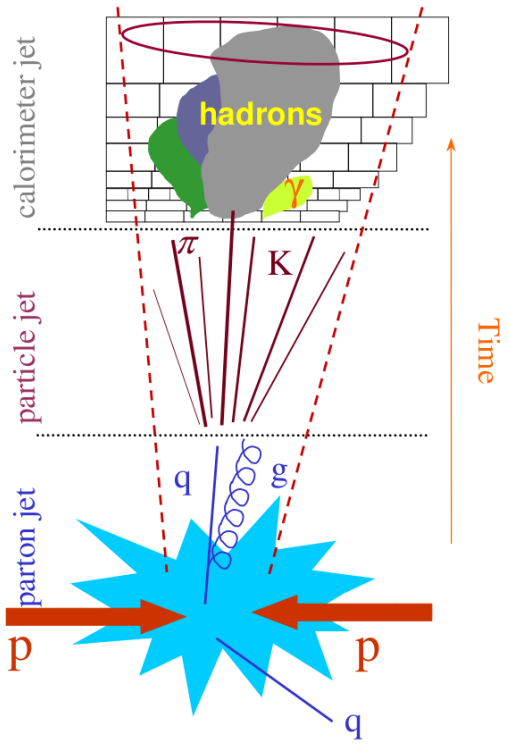

Figure 2.6: An illustration of the jet production process and measurement. [25]

Chapter 3

Experimental setup

This chapter is dedicated for introducing the experimental framework in which the jets under study are observed and reconstructed. The LHC machine is introduced in the first part of this chapter, after which the CMS experiment is described with all the major sub-detectors in section 3.2. In order to provide the reader with a realistic picture of data taking in a modern high-energy collider experiment, the important concept of *triggering* is briefly covered in the last section.

3.1 Large Hadron Collider

The Large Hadron Collider is the largest machine in the world, with circumference of 27 km in an underground tunnel at the border of Switzerland and France. The tunnel (Fig. 3.1), inherited from the previous LEP accelerator, is at its deepest point 175 meters under the Franco-Swiss territory, and is partly situated under the mountain range of Jura.

A process which begins by separating protons from hydrogen atoms and ends colliding them at a fraction to the speed of light cannot be done by one accelerator alone. Hence the LHC machine comprises a multitude of pre-accelerators, which ramp up the energy step by step. The CERN accelerator complex together with the pre-accelerators of the LHC is illustrated in Figure 3.2.

The LHC is built to the same tunnel previously used for the Large Electron Positron collider LEP that operated 1989-2000. Dug under the Franco-Swiss border area, the tunnel goes between 45 and 170 meters underground. The total length of the ring is 26.7 kilometers and roughly 80% of the tunnel is under French territory. The plane spanned by the LHC is inclined 1.4% towards the Lake Geneva, which is practical for redirecting water coming from

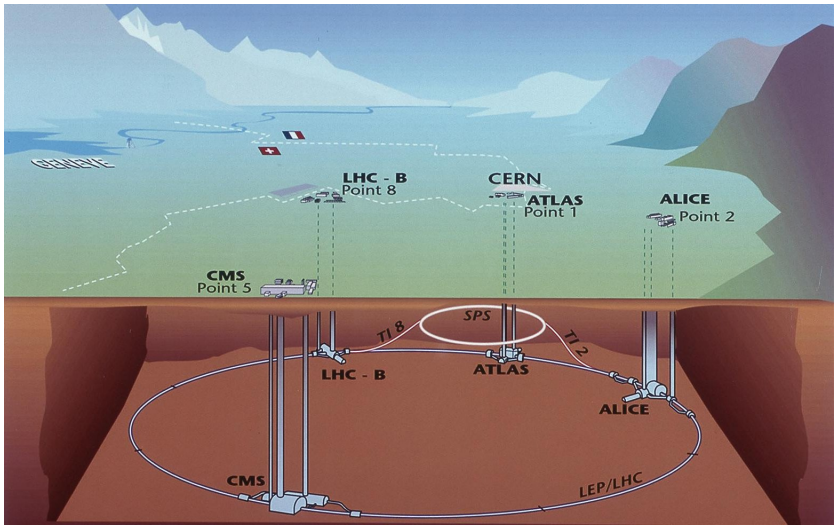

Figure 3.1: An overall view of the LHC accelerator tunnel.[11]

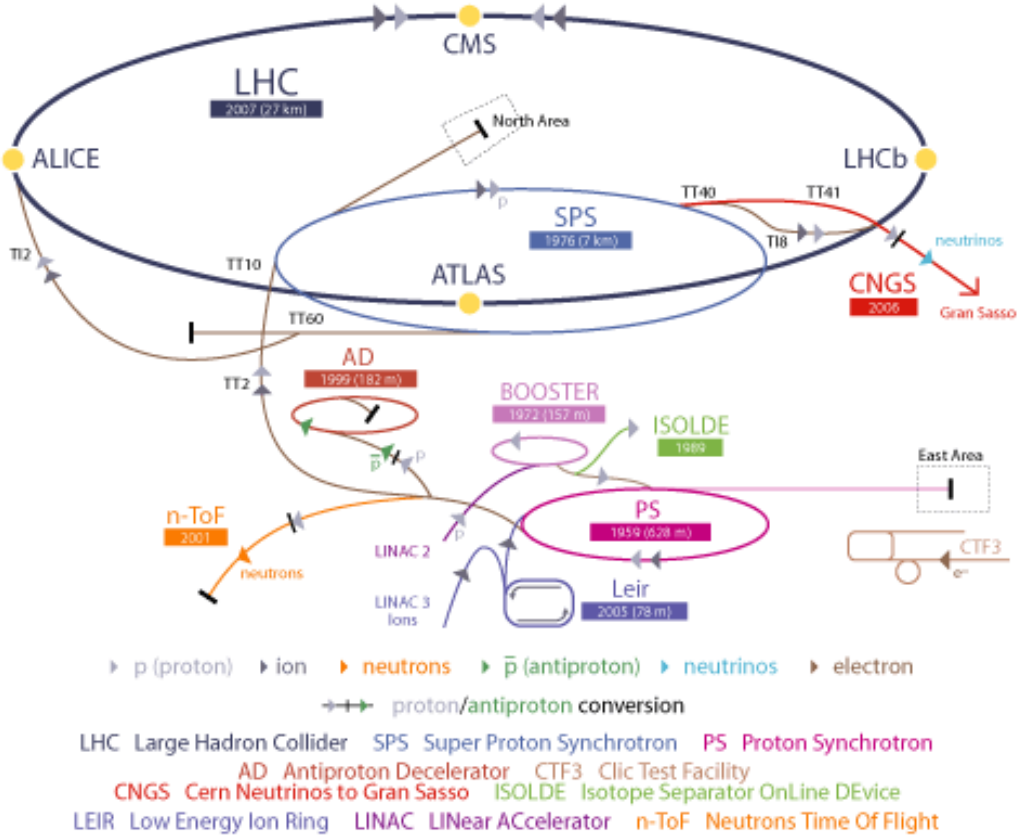

Figure 3.2: The CERN accelerator complex.[11]

the Jura out of the tunnel. The tubular tunnel itself is 3.8 meters in diameter and roomy enough for the *dipole* and *quadrupole* magnets that house the beam pipes, alongside which there is enough space for simple maintenance traffic.

The LHC is divided to eight octants and the convention is that the center of Nth octant is called Point N (see fig. 3.3). In each of the Points, or Access Points more precisely, is an elevator with the first and last stop down at the tunnel. The CMS experiment is located at Point 5, ATLAS at Point 1, ALICE at Point 2 and LHC-B at Point 8. The accelerator facility is at Point 4 and the beam dump that is responsible for safely terminating the runs is at Point 6. Points 7 and 3 are dedicated to beam quality management apparatus.

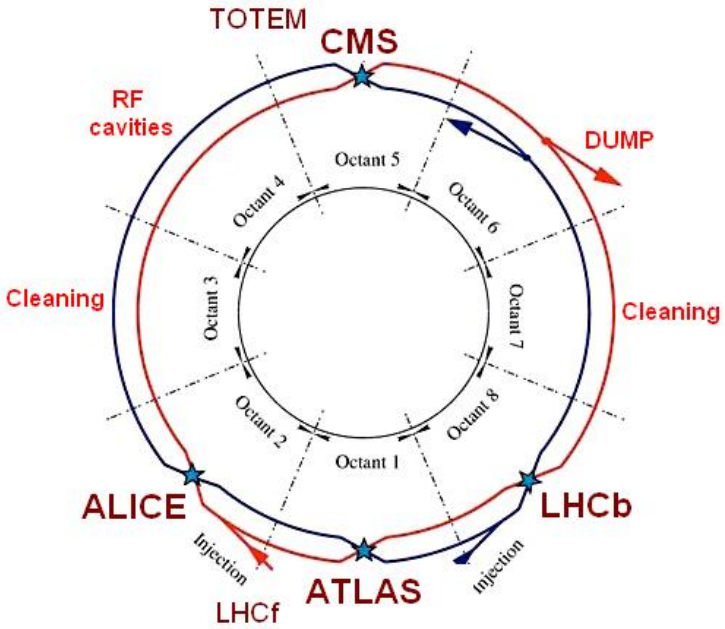

Figure 3.3: Different sections of the LHC. [24]

3.2 Compact Muon Solenoid experiment

The Compact Muon Solenoid experiment is in terms of weight the greatest detector at the LHC. Despite its superior weight of 12 500 tonnes, by physical size of 15 m (diameter) by 21.5 m (length) CMS is just the third biggest apparatus along the ring. This is why it is called compact. The other two words in the name refer to the detector's excellent muon detecting capability and exceptionally strong solenoid magnet.

As illustrated in figure 3.4, the CMS consists of five separate layers. Starting from the center, a traversing particle faces first the tracker that can measure trajectories of charged particles. Second layer from the interaction point is the electromagnetic calorimeter, ECAL. The ECAL is capable of measuring energies of photons and electrons and often also contributes to the measurement of hadronic particles. After ECAL comes the hadronic calorimeter, HCAL, that stops and measures all hadrons coming through the ECAL. The fourth layer of the CMS is the superconducting solenoid that induces a strong magnetic field throughout the detector. Magnetic field plays a key role in particle detectors, because it bends particle trajectories and hence makes momentum measurements possible. The last and outermost layer consists of the muon detecting system that doubles as a return yoke for the magnetic field and triples as a supporting structure for the whole assembly.

The scale and longitudinal structure of the CMS experiment can be seen in Figure 3.5

All of the detector layers are introduced in more detail in the following sections, and also the concept of triggering is described towards the end of this chapter.

3.2.1 Tracker

The first thing along a particle's path after penetrating the beryllium beam pipe is the tracker. Despite the tracker's rather modest size compared to the other detector elements, it is the most important sub-detector. Tracker provides high accuracy momentum determination without affecting significantly the particles' energy and trajectory. The key challenges in the development of a tracker are that it has to be almost transparent to the particles or it would affect the tracks significantly, and it has to stand extreme radiation conditions, as it is the closest element to the interaction point. The tracker is subjected to pulses of highly relativistic particles at the rate of up to 40 MHz.

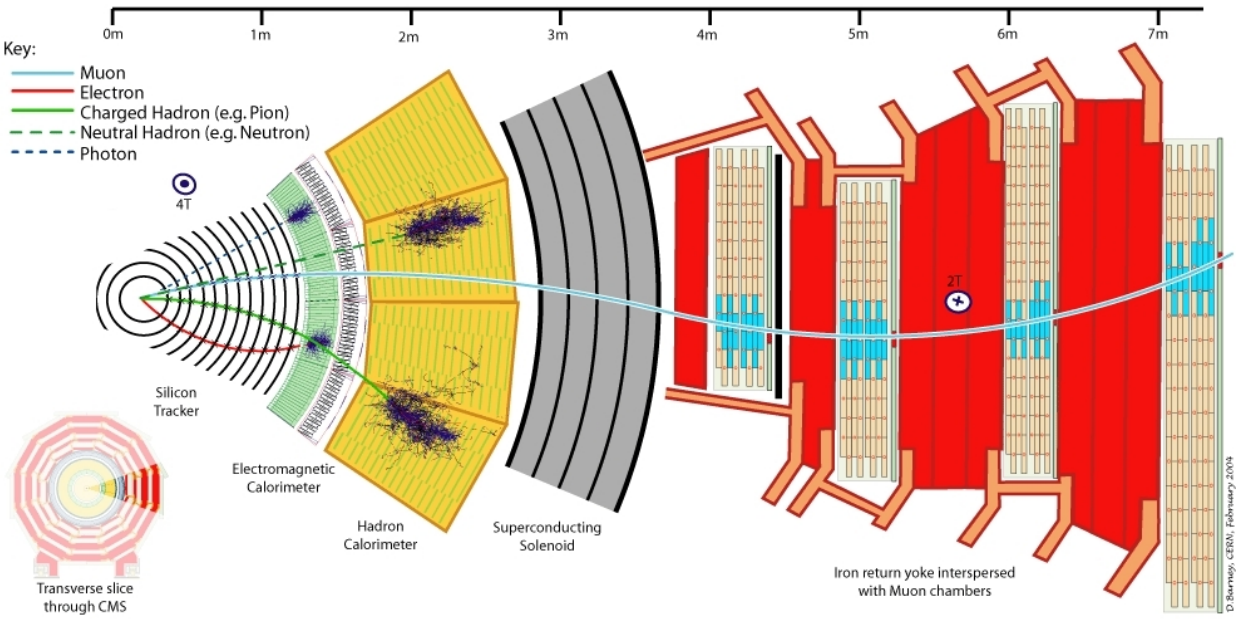

Figure 3.4: Cross-section slice of the CMS detector. [11]

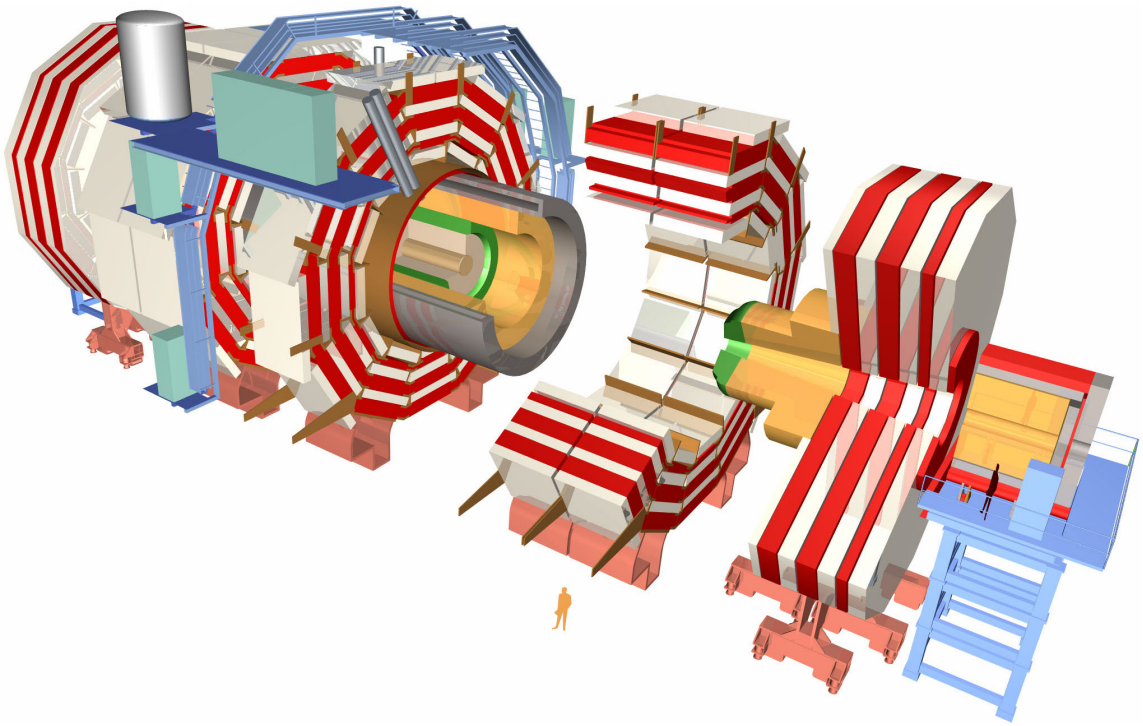

Figure 3.5: Exploded view of the CMS experiment. [11]

As illustrated in figure 3.6, the CMS tracker is a combination of silicon pixel and silicon strip detectors. The pixel tracker is the innermost detector and just the size of a shoebox, but still measures charged particles in three layers and records signals of a total of 65 million pixels. Each pixel is 100×150 μm^2 in size and is electrically connected to a read-out system.

Outer layer of the tracker, the silicon strip tracker, performs tracking measurements with ten million strips that are arranged in ten layers. Silicon strips do not provide as high spatial accuracy as pixels, but when strip hits are combined with pixel signals, measurement accuracy of 10 microns is achieved for each track.

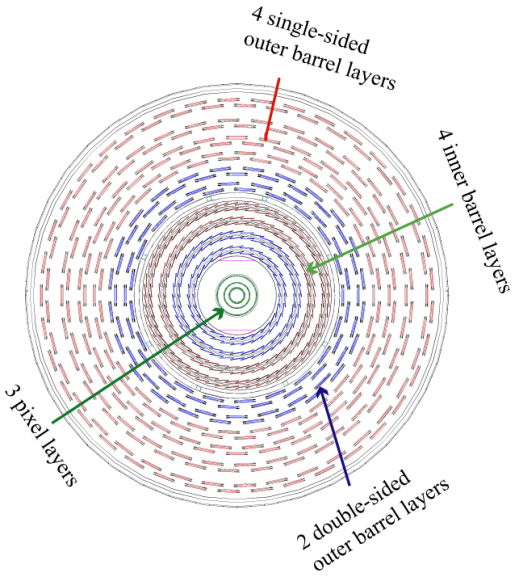

Figure 3.6: Cross-section view of the CMS tracker. Total diameter is 2.6 meters. [11]

For both the pixel and strip detector the measurement mechanism is ionization in a semi-conductor. When a charged particle flies through silicon, it knocks off electrons from the depletion region of the semi-conducting junction. These free electrons then drift to anodes at the n-type end of the semi-conductor due to the 300V voltage applied over the p-n junction. These anodes are connected to the readout system from which tracker hits can be read after amplification. [5]

3.2.2 Electromagnetic calorimeter

After traversing the tracker, optimally with minimal interaction, particles arrive to the ECAL. The ECAL stops and measures energies of light electromagnetically interacting particles: electrons, positrons and photons. Charged hadrons also sometimes interact with a nucleus in the ECAL but the calorimeter is not thick enough to measure more than the first part of the resulting cascade. Muons mostly pass the ECAL as minimum ionizing particles and do not produce secondary particles.

The CMS ECAL is made of lead tungstate ($PbWO_4$) crystals. Lead tungstate is an extraordinary material with exceptionally high density (8.28 g/cm^3) , excellent optical transparency and extremely short excitation time.

This combination makes lead tungstate a very good medium for a calorimeter. High density and nuclear charge ensure high interaction probability, transparency makes collecting scintillation light possible and short excitation time allows measuring events every 25 ns.

Electromagnetic crystal calorimeter works as follows: incoming energetic photon (electron) undergoes pair creation (bremsstrahlung) in the vicinity of a heavy nucleus, where the probability for the process is high due to the strong electric field. Then, in the case of an initial photon, the created electron-positron pair emit bremsstrahlung photons when moving close to nuclei, and in the case of an initial electron, the emitted bremsstrahlung photons undergo pair creation. This alternating process goes on until the bremsstrahlung photons or secondary electrons do not have enough energy for further pair creation and. Secondaries then excitate medium molecules and when this excitation is relaxed (in nanoseconds), the emitted photon is in the wavelength region that the photodetectors at the end of each calorimeter crystal can detect.

The electromagnetic calorimeter consists of 75 848 crystals, each 23 cm long in the barrel and 22 cm long in the end caps. Crystals are arranged to a single layer next to the silicon strip tracker. In barrel the crystals' cross-section is 2.4 cm x 2.4 cm and in the end caps 3 cm x 3 cm, which ensures good spatial resolution.

Radiation length of the lead tungstate crystals is only 8.8 mm. This means that a $PbWO_4$ crystal is about 26 radiation lengths long. High-energy electron loses $1-\frac{1}{e} \approx 63\%$ of its energy in one radiation length, so the average energy loss in a crystal is 0.999994 %. Approximately the same is true for photons that convert to positrons and electrons promptly after entering the crystal. Electrons and photons are thus efficiently stopped by the CMS ECAL. [5]

3.2.3 Hadronic calorimeter

ECAL stops efficiently only electrons and photons because they are light particles and hence for them the probability for bremsstrahlung is high. All the other particles penetrate ECAL and have to be detected with something more massive. For this, the third layer of the CMS is the HCAL that is able to stop all known particles except muons and neutrinos.

The hadronic calorimeter is, unlike the ECAL, a sampling calorimeter. This means that there are separate layers where particles interact to create secondaries and layers where measurable scintillations are created by the secondary particles. In order to have good stopping capability, a HCAL needs to be thick and be made of dense material, because only then sufficient

interaction probability is achieved also for hadrons. In CMS HCAL the absorber layers are made of brass which is an alloy of copper and zinc. While the density of the absorber material is close to that of lead tungstate, the crucial difference between the ECAL and the HCAL is that the latter is about one meter thick.

When hadrons collide with medium nuclei, follows a hadronization process due to color confinement. Hadronization results in a cascade of secondary particles which then hit the scintillation layers of the sampling calorimeter. In CMS these scintillation layers are made of plastic and are optically connected to readout photodetectors. As already implied in the previous section, the operational principle of a scintillator is that photons are emitted in a relaxation process after medium molecules have been excited by passing secondaries.

The HCAL is a massive structure organized to four distinct parts: barrel part (HB), outer barrel part behind the solenoid (HO), end cap plates (HE) and forward calorimeters (HF). For example the barrel part consists of 36 wedges each weighing 26 tonnes. So where the nearly 80 000 dense ECAL crystals weigh some 90 tonnes, the barrel part of the HCAL alone weighs almost 1000 tonnes. [5]

3.2.4 Superconducting solenoid

All of the detector layers introduced in previous sections are immersed in extraordinarily strong magnetic field induced by the superconducting solenoid magnet. The solenoid did not make it all the way to the name of the experiment for light reasons; it is the most powerful magnet system of its size range ever built and the five solenoid modules are the biggest superconducting magnets in the world. It is a hollow cylinder of 12.5 m length and 6 m diameter.

The solenoid provides the tracker and calorimeters with a magnetic field of up to 4 T flux density with nearly uniform configuration (Fig. 3.7). Such a field is able to bend trajectories of charged particles that would have a nearly straight track in weaker fields. Thus, the CMS solenoid enables high precision curvature and momentum measurements. Although the designed 4 T field density is achieved in tests, the magnet is operated at 3.8 T to ensure reliable operation throughout the experiment's lifespan.

Relying on superconducting technology, all the solenoid modules are inside a cryostat where circulating liquid helium keeps the temperature at 4.2 K. The current of 19 500 A that induces the exceptional field is too high for superconductors relying on more conventional liquid nitrogen cooling. Non-sufficient cooling leads to *quenching*, which is a phenomenon where conductor

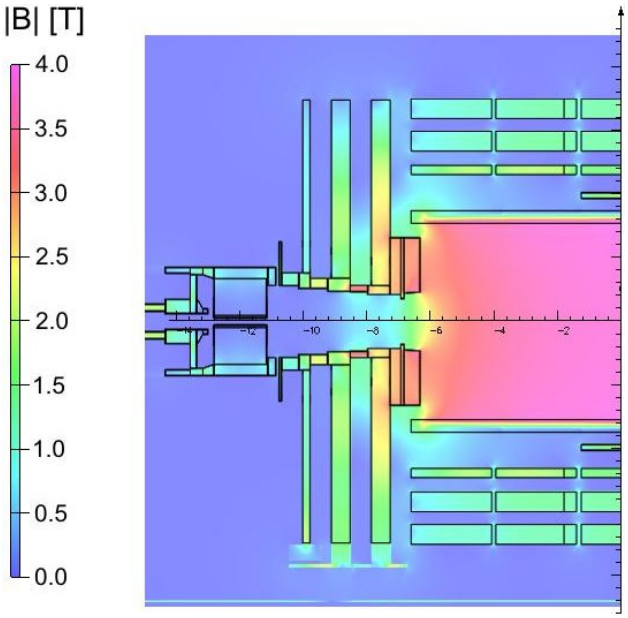

Figure 3.7: Simulated field configuration with 3.8 T flux density. [6]

moves from a superconductive to a resistive state and the strong current generates a lot of heat in a fraction of a second. Uncontrolled quenching in a solenoid of this size would have dramatic consequences, therefore the coils are accompanied by a system for discharging the solenoid safely in the case of increasing temperature. [5]

3.2.5 Muon chambers & Iron return yoke

The outermost and perhaps the most characteristic part of the CMS experiment are the muon detecting system and the return yoke that surround the detector. This fifth layer, which is coloured white and red in Figure 3.5, is a multi-purpose assembly that takes care of detecting muons, guiding the magnetic field, stopping any residual hadrons and finally supporting the whole experiment.

As deducible already from its name, muon detection plays a key role in CMS physics strategy. Not least because muons are the protagonists in many detection channels of the Higgs boson. Thus, the muon system is a sophisticated device that exploit three different techniques: drift tubes (DT), cathode strip chambers (CSC) and resistive plate chambers (RPC). The barrel region is covered with DTs with an additional layer of RPCs. End caps of the muon system use CSCs instead of DTs, but also incorporate

RPCs for the same reason as the barrel: sufficient redundancy for efficient muon triggering, i.e. for deciding whether an event is worth recording.

The importance of serving as a general support structure is self-explanatory, but the fourth task of the system, field guidance, is less evident. The field configuration of a solenoid is normally one that reaches its maximal magnitude in the center and then weakens rather fast towards mouths of the cylinder. Field like this is not optimal for momentum measurements, as curvature of tracks would diminish when particles move in the z-direction. Thanks to the iron return yoke, however, the field configuration inside the CMS solenoid is nearly uniform (Fig. 3.7). This is so because ferromagnetic medium, such as iron in this case, layered on top and at the mouths of a solenoid greatly reduces spatial spreading of the magnetic field outside the solenoid. This, in turn, causes the field lines to be more parallel inside the cylinder.

3.2.6 Trigger system

At the design luminosity of the LHC, 40 million bunch crossings takes place every second at the center of the CMS experiment. Recording signals of an event takes about one megabyte of storage space, so recording all collisions of one hour would fill roughly 1.5×10^{11} MB of disk space, corresponding to almost 150 000 hard drives of one terabyte. This is clearly a practically impossible task, and this is why triggering is unavoidable.

Only a tiny fraction of the collision events actually contain signals of interesting physics processes, and the task of the triggers is to extract these very rare events from all the rest. Trigger has to decide in a few microseconds whether an event is worth recording, and thus consists of dedicated hardware and extremely fast reconstruction algorithms that run on microprocessors placed in the close vicinity of the detector.

In CMS the trigger system makes the decision in three microseconds and reduces the rate of events for recording from 40 million per second to about 100 events per second. In terms of required disk space this is around 100 MB/s, which is still a lot but manageable.

Chapter 4

Jet reconstruction

When quarks and gluons hadronize to jets instantly after a pp collision, the result is a shower of particles that is spatially spread before hitting different layers of the detector. Jet reconstruction is a task where signals of the detector are combined for understanding which hits belong to a jet and which is just a single particle born in the collision. In CMS this process begins with the particle-flow (PF) event reconstruction (section 4.1) where the event and all its particles are reconstructed, after which the anti-k_T jet clustering algorithm (section 4.2) finds which of the particles belong to jets. In the last section of this chapter the concept of jet energy composition – perhaps the most important single concept of this study – is described.

4.1 Particle-flow event reconstruction

The particle-flow event reconstruction scheme is a modern approach for event reconstruction that makes use of all the sub-detectors of the CMS experiment for reconstructing collision events to the level of single particles. In contrary to other reconstruction techniques in high-energy physics experiments, particle-flow produces a list of particles in the event with classification of the type of the particle. The PF algorithm classifies particles as muons, electrons, photons, charged hadrons and neutral hadrons.

Before proceeding to particle-level reconstruction, the PF algorithm divides a collision event to "blocks" of detector signals using a link algorithm. These blocks consist of signals that probably are produced by one or a few physical objects. This approach is essential for having computationally light algorithm that can efficiently reconstruct highly complex events.

After the division to blocks, the algorithm proceeds by extracting muons from detector signals by combining information of the tracker and muon

chambers. Tracks of reconstructed muons are then removed from the block from disturbing further stages. Electron reconstruction and identification follows. Electrons are identified by combining tracks to matching energy deposits in the ECAL, and once identified, the corresponding detector hits are removed from further stages as was done for muons.

The remaining tracks are then reconstructed as originating from charged hadrons, as all other charged particles are already taken care of. Momenta of charged hadrons are compared to calorimeter energy measurement, and if the calorimeter energy deposit is significantly higher than that expected from the corresponding track (assuming the deposition is made by charged pion), the excess in the calorimeter energy is reconstructed primarily as photons and secondarily as neutral hadrons. Processed signals are again removed from blocks.

After all tracks are have been processed and the corresponding calorimeter signals are taken into account, the remaining isolated calorimeter deposits have to be due to photons or neutral hadrons. As photons are efficiently stopped by the ECAL and neutral hadrons deposit energy mostly in the HCAL, the remaining ECAL clusters are reconstructed as photons and HCAL clusters as neutral hadrons.

The particle-flow algorithm was met with a lot of skepticism among physicists when it was proposed, and such a complex reconstruction system was first questioned. The PF approach has, however, proven to be at least as reliable as more straight-forward methods, but with significantly better results. It is now the standard method in CMS analyses and shows the way for future reconstruction algorithms.

Figure 4.1 illustrates how PF converts signals in the detector back into physical objects. When the algorithm has processed all the signals of an event, it gives a list of reconstructed particles with information on their type, direction and momentum. This list can be used in further analysis, such as in jet clustering, as if it was a particle listing from a Monte Carlo event generator.

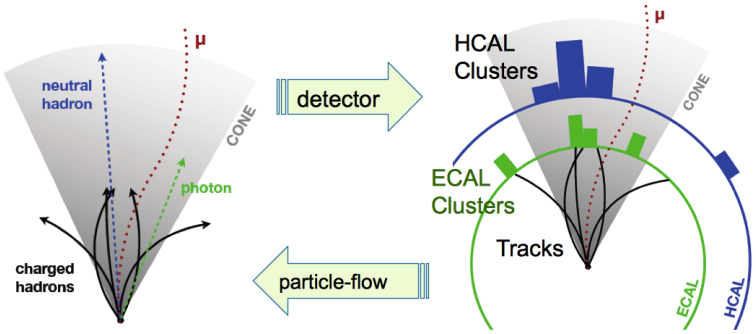

Figure 4.1: Illustration of the PF reconstruction process.[16]

4.2 Anti-k$_t$ clustering

The next phase in the jet reconstruction process is clustering. Clustering is a process where nearby particles are merged together for creating jets, and this is actually how a jet is formally defined, as no unambiguous definition for a jet would otherwise exist.

The anti-k$_t$ algorithm is a successor of the older k$_t$ algorithm with the difference that in anti-k$_t$ hard high-p_{T} particles are clustered first, contrary to the k$_t$ which first clusters soft particles. Mathematically, the anti-k$_t$ algorithm is defined by the metric d_{ij}, that defines the distance of particles i and j, and d_{iB} that is the distance from particle i to the beam:

$$d_{ij} = \min(k_{ti}^{-2}, k_{tj}^{-2}) \frac{\Delta_{ij}^2}{R^2}, \tag{4.1}$$

$$\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2,$$

$$d_{iB} = k_{ti}^{-2}. \tag{4.2}$$

Here k_{ti}, y_i, ϕ_i are respectively the transverse momentum, rapidity and azimuthal angle of particle i and R is the radius parameter that determines the jet size.

The clustering process starts by finding the smallest of the distances d_{ij} and d_{iB}, and then merging particles i and j if $d_{ij} < d_{iB}$. Iteration proceeds until no particle j is found that fulfills $d_{ij} < d_{iB}$. Then i is considered a complete jet and the algorithm continues to process other particles of the event.

The anti-k$_t$ algorithm provides in many ways nearly optimal clustering performance. It produces regular-shaped conical jets and it is both collinear and infrared safe algorithm. Collinear safety means that nearly collinear particles are not clustered as different jets but as one jet whose leading particle decayed or hadronized to two pseudo-collinear particles, as often happens in energetic collisions. Infrared safety is a property that ensures that soft low-p_{T} 'infrared' particles do not change clustering results drastically, as can happen with non-infrared safe algorithms. This is important especially in a hadron collider, where a lot of soft particles unrelated to the hard scatter process are born from pile-up, underlying event and detector noise. [3]

In this work and also generally in nearly all of CMS analyses, anti-k$_t$ with $R = 0.5$ is the chosen algorithm, but in some cases where broader jets are wanted, the radius parameter is chosen to be 0.7.

Figure 4.2 shows the clustering result of a generated event with injected soft particles for illustrating infrared safety and regularity of anti-k$_t$ jets. A complete PF-reconstructed anti-k$_t$ jet is shown in figure 4.3.

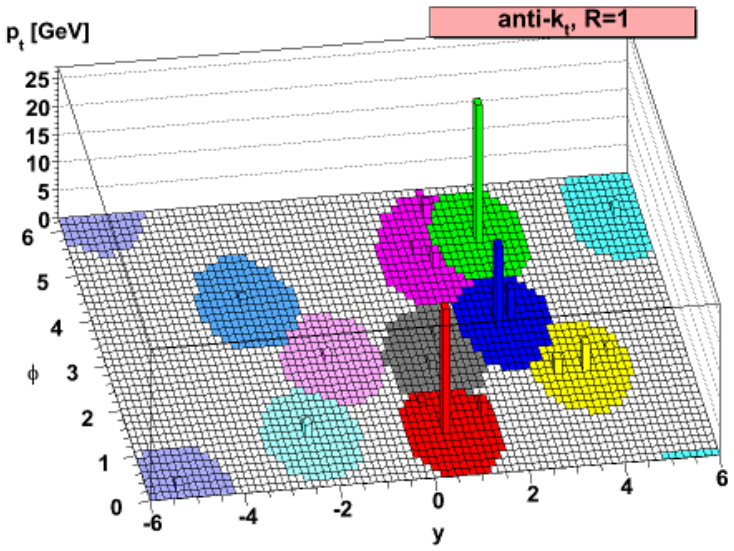

Figure 4.2: Colored areas represent anti-k$_t$ clustered jets at unfolded plane of a cylindrical detector.[3]

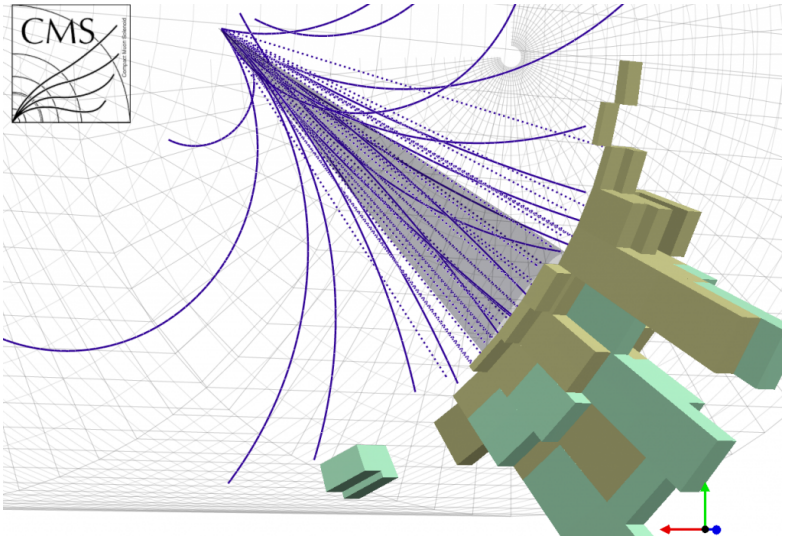

Figure 4.3: A PF reconstructed anti-k_t jet.[8]

4.3 Jet energy composition

Jet energy composition is a concept that measures how the energy of a jet is distributed to different kinds of particles. By virtue of the particle-flow algorithm, jets are built from single particles whose types are classified. Without this in-depth studies of the jet energy composition would not be possible. It is worth noting that this definition of jet energy composition does not consider particle counts at all, but only the sums of energy of each particle categories.

Particle-flow classifies particles to seven categories: electrons, muons, charged and neutral hadrons, photons, forward hadrons and forward electromagnetically interacting particles. On top of that we use event-by-event vertex measurements for separating charged particles originating from pile-up vertices. In the central region of CMS we divide jet energy to six components: charged hadrons, photons, neutral hadrons, electrons, muons and charged hadronic pile-up. Electrons and muons are in this thesis summed together because of their small contribution. In the forward region, roughly $|\eta| > 3$, jet energy is divided only to forward hadrons and forward photons, because due to more rudimentary detector coverage there further particle classification is not possible. The forward region, however, is not at the center of attention for this study, because the majority of CMS physics analyses use mostly jets in the barrel region or inside the tracker's coverage, thus enhanced calibration in $\eta < |1.3|$ and in $\eta < |2.5|$ are the main objectives.

Typical jet energy composition in the barrel region is shown in figure 4.4.

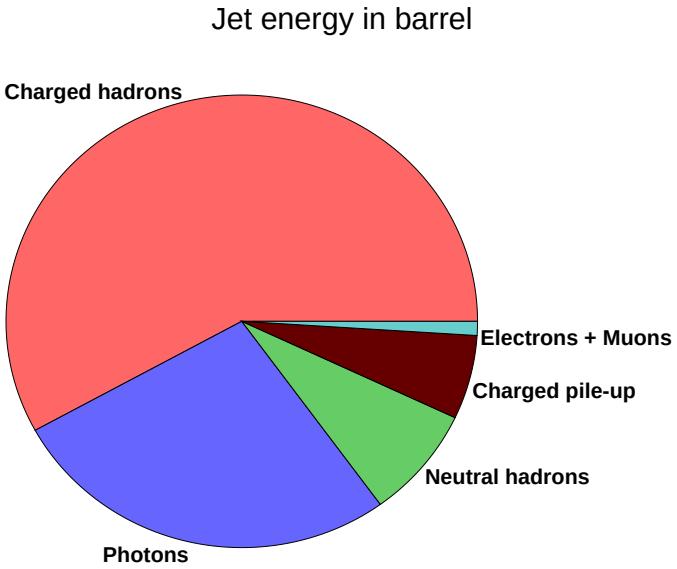

Figure 4.4: Average jet energy composition of jets with $p_{\mathrm{T}} \in [114, 145]$ GeV in barrel region. In numbers the percentages are: charged hadrons 58%, photons 27%, neutral hadrons 8%, charged pile-up 6%, electrons and muons 1%.

Chapter 5

Methods

The analysis chain is presented roughly in the order of execution in the analysis code in section 5.1, and then in section 5.4 the method of using a streamlined detector simulation for studying different miscalibration scenarios is introduced. In the last section 5.5 the visualization techniques used for converting numerical data to meaningful plots are described and example plots are shown for preparing the reader to the final results.

5.1 Event selection and triggering

In this study where features of jets are at the center of attention, event selection is actually mostly jet selection. As the reconstruction process can sometimes falsely identify detector noise as a jet, we need to require certain properties from the jets under investigation for having a representative sample. In the case of Monte Carlo generated jets ('genjets'), we also compare the generator level 'MC truth' information to properties of the reconstructed jet to avoid poorly reconstructed jets from ending up skewing our results.

Jet triggering is another matter that requires special attention. Jets are triggered with multiple jet p_{T} thresholds. A trigger of some energy interval can produce unreliable decisions near its energy threshold due to the rather coarse reconstruction process that can be done with online triggers. For this, the nominal trigger range is different from the efficient range that gives quality trigger decisions, so for each jet trigger we accept only jets whose energy lies in the efficiency interval of the trigger. Triggering is naturally only necessary for measured data and not for Monte Carlo events, as in MC generation one can control which events are wanted.

The jet quality cuts are listed in table 5.1 and triggers with their nominal and efficient triggering thresholds are shown in table 5.2.

Table 5.1: Jet and event selection criteria. If all of the requirements in the 'Value' column are not satisfied, a jet is rejected. Particle fractions refer to fractions of total jet energy.

Quality cut	Value	Purpose	Note
Particles in a jet	> 1	Cut one-particle jets	
Photon fraction	<0.99	Cut unphysical jets	
Neutral hadron fraction	<0.99	Cut unphysical jets	
Photon fraction	<0.9	Cut unphysical jets	In $\|\eta\| < 2.4$
Neutral hadron fraction	<0.9	Cut unphysical jets	In $\|\eta\| < 2.4$
Electron fraction	<0.99	Cut unphysical jets	In $\|\eta\| < 2.4$
Charged hadron fraction	>0	Cut unphysical jets	In $\|\eta\| < 2.4$
Charged hadron multiplicity	>0	Cut unphysical jets	In $\|\eta\| < 2.4$
$E_{\mathrm{T}}^{miss} / \sum E_{\mathrm{T}}$	< 0.3	Cut high E_{T}^{miss} events	
$p_{\mathrm{T}}^{reco} / p_{\mathrm{T}}^{gen}$	< 1.5	Require quality reco.	MC only
$p_{\mathrm{T}}^{reco} / \hat{p}_{\mathrm{T}}$	< 1.5	Require quality reco.	MC only

Table 5.2: Jet triggers with their nominal p_{T} thresholds and effective p_{T} intervals used in this analysis.

Jet trigger	Nominal threshold	Range in analysis
Jet-40	40 GeV/c	56-114 GeV/c
Jet-80	80 GeV/c	114-220 GeV/c
Jet-140	140 GeV/c	220-300 GeV/c
Jet-200	200 GeV/c	300-362 GeV/c
Jet-260	260 GeV/c	362-430 GeV/c
Jet-320	320 GeV/c	430-507 GeV/c
Jet-400	400 GeV/c	507-2000 GeV/c

5.2 Pile-up re-weighting

In each bunch crossing not just one proton pair collides but typically around 20. Out of all the collisions the most energetic is usually chosen as the interesting one and the rest mostly disturb analysis of the event. The extra collisions are called pile-up (PU), and taking its effects into account needs a lot of work.

Particles from pile-up collisions also contribute to jets, and hence there are more particles in jets than actually originate from one energetic parton-parton collision. Contribution of pile-up to jet energies is subtracted from jets in the jet energy correction process. This process, however, only corrects jets' energies without considering energy carried by different types of particles. This changes jet energy composition, and we need to take this effect into account when comparing simulated jets to measured ones.

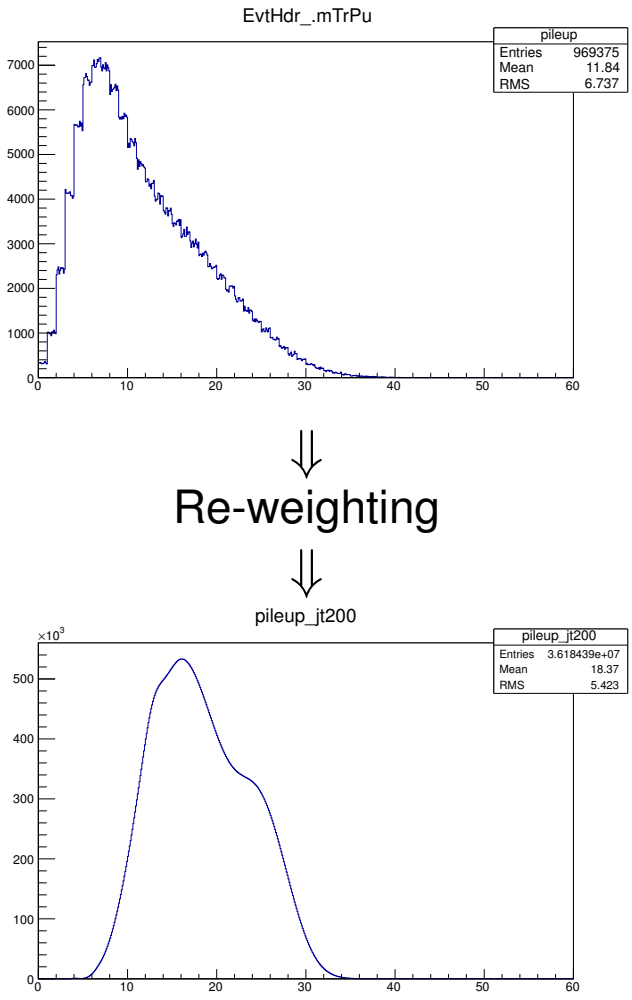

Figure 5.1: Illustration of the pile-up re-weighting process. Number of primary vertices are in the x-axes and number of events in the y-axes.

In order to have comparable simulation and data, the pile-up distributions should be identical. Pile-up profile of data depends on running conditions of the LHC; how well beams are focused, how 'head-on' the bunches collide, how much particles the bunches contain etc. For this it is impossible to generate Monte Carlo simulations exactly reproducing the pile-up distribution of data, as it simply is not known beforehand. The best one can do is to make an educated guess of the average pile-up conditions that will prevail and produce simulations based on that. Of course this procedure gives MC data a pile-up profile roughly similar to what real collisions will have, but not quite. Nevertheless, we

need to have matching profiles, and what we do is we re-weight the Monte Carlo events in the analysis phase so that the distributions will match up to high precision. Re-weighting is done for every simulated event by taking the number of generated pile-up collisions in the event, N_{PU}, and calculating the probability of this N_{PU} in the whole simulation sample. Then, by first finding probability for the same number of pile-up collisions in measured data and then filling jets to histograms with the weight $\frac{P(N_{PU}^{DATA})}{P(N_{PU}^{MC})}$, we obtain similar PU distributions for data and simulation. Figure 5.1 illustrates this procedure, and figure 5.2 shows comparison of reconstructed number of vertices in DATA and MC after re-weighting. The latter is an important cross-check of the PU re-weighting procedure.

5.3 Tag-and-probe jet selection

Tag-and-probe method is a special additional jet selection method made for avoiding any selection bias, or in other words for ensuring that we do not favor jets of some particular composition when binning the histograms in reconstructed jet p_{T} for a steeply falling p_{T} spectrum.

The idea of the method is to look for dijet events with two back-to-back jets, i.e. events with two jets opposite to each other as shown in Figure 5.3. When such a pair is found, we further demand that these are the most energetic, 'leading' jets in the event so that we can assume that they originate from the same initial parton-level process and that their true p_{T} are approximately same. The applied requirements for the events are:

- Dijet angle: $\Delta\phi \equiv \min(\measuredangle(\mathrm{Jet1}, \mathrm{Jet2})) > 2.8$
- Energetic leading jets: $p_T^{Jet3} < 0.3 \times \frac{p_T^{Jet1}+p_T^{Jet2}}{2}$.

When an event fulfills these requirements, we choose one of the jets to be a 'tag jet' and apply the quality criteria described in section 5.1 for for both. The other jet is then a 'probe jet' and if the tag jet passes the trigger p_{T} cuts, we use the probe jet for observing the jet energy composition.

The tag-and-probe jet selection method reduces the statistics by about 33% for real jets and by about 44% for MC jets. The difference is due to poor modeling of the third jet by the Pythia Monte Carlo event generator.

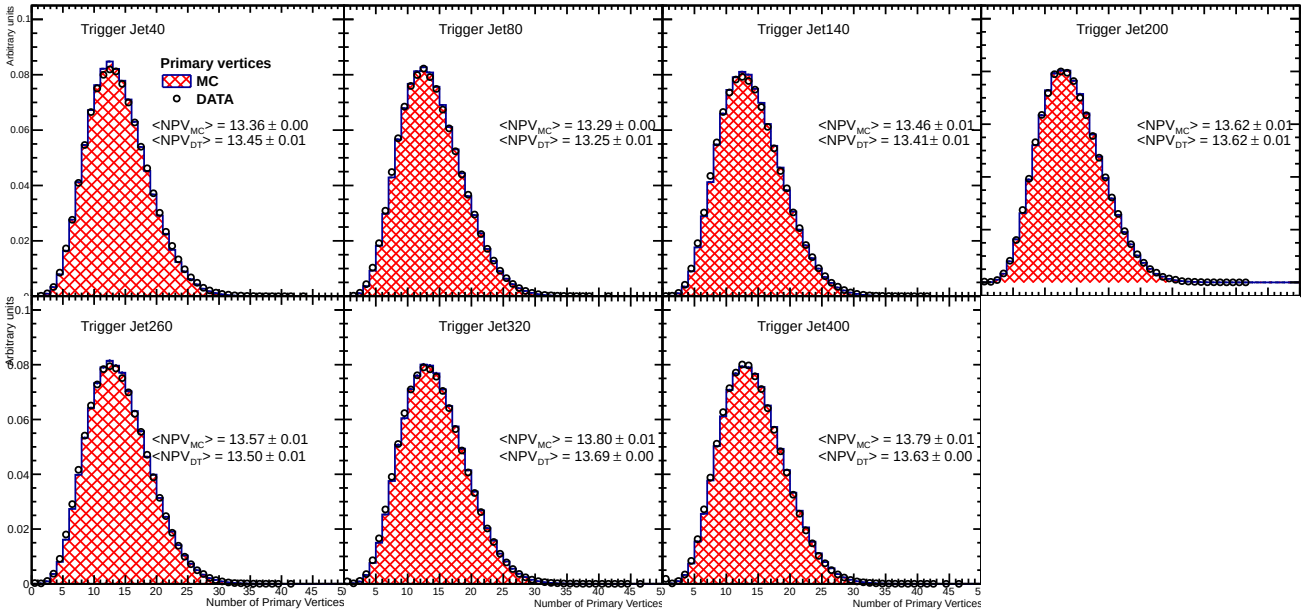

Figure 5.2: Number of primary vertices after re-weighting, monitored for each trigger separately. Simulation is in histograms, data is shown with circles.

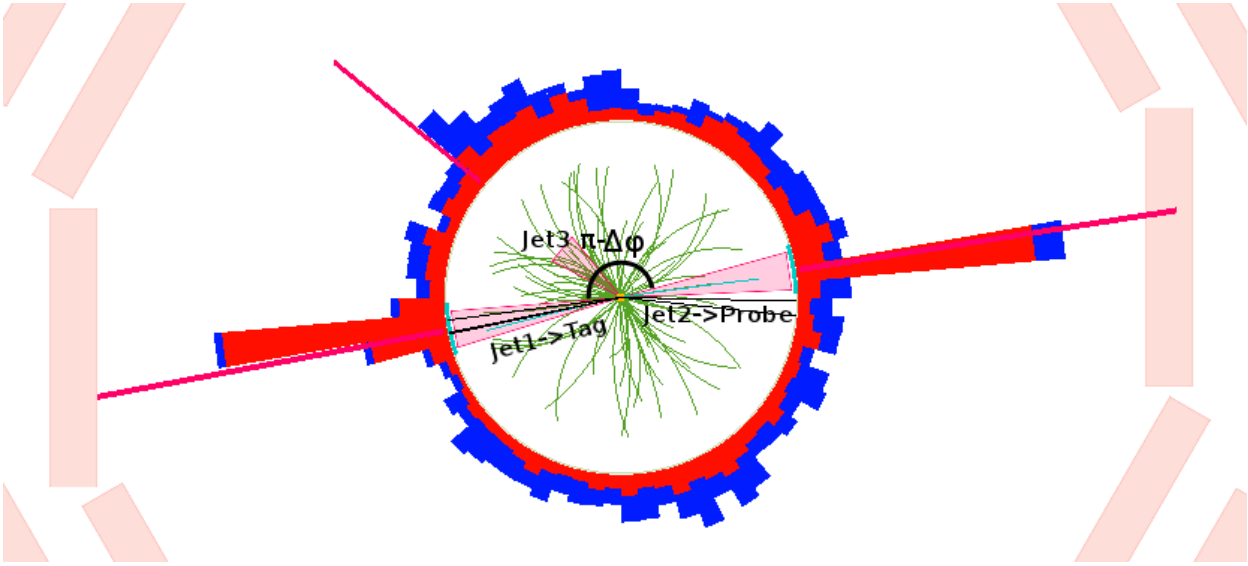

Figure 5.3: An event suitable for the tag-and-probe selection. Green curves are particle tracks and blue and red towers are ECAL and HCAL energy deposits. The $\Delta\phi$ is the smaller angle between the tag and the probe.

5.4 FastSim parameter variation

For testing how sensitivity of different detector elements affect the jet energy composition, we used the lightweight detector simulation framework FastSim as already mentioned in section 2.4. FastSim allows us to perform detector simulation significantly faster than with the full CMSSW framework, which is essential when running simulations with several different detector parameters.

We first varied the hadron response of the whole calorimeter system of the CMS detector and then hadron responses of electromagnetic and hadronic calorimeters separately. For every scenario we changed a response parameter in the FastSim configuration file by three per cent to both directions, i.e. to more responsive (+3%) and to less responsive (-3%). These are then compared to a reference simulation, where all the simulation parameters are at their default values.

In all simulations we generated 100 000 collision events, out of which around 90 000 yielded successful detector simulation and reconstruction and hence were usable in the final analysis. The roughly 10% of the events were rejected in the detector simulation or event reconstruction phase usually due to time limit given for production of an event in the FastSim framework. Detailed information on the performed FastSim runs are listed for each scenario in table 5.3, where also the naming conventions of different scenarios are listed.

Table 5.3: Run parameters of response variation and event counts for each scenario.

Scenario name	FastSim Parameter	Value	Events
CTRL	*all*	*defaults*	94699
SPR-3%	ECAL+HCAL response	0.97	91356
SPR+3%	ECAL+HCAL response	1.03	89925
ECAL-3%	ECAL response	0.97	92192
ECAL+3%	ECAL response	1.03	90600
HCAL-3%	HCAL response	0.97	93112
HCAL+3%	HCAL response	1.03	91488

5.5 Visualization

For observing η-and p_{T}-dependent changes in jet energy composition, it is convenient to visualize composition against these variables. The basic plots we use for jet energy composition analyses are stack plots, where contributions of different particle types are stacked one on the other, the y-axis then being from 0 to 1 (or 0% to 100%) and x-axis either η or p_{T}. We also produce corresponding difference plots, where small changes are easily visible. Example plots are shown in figure 5.4.

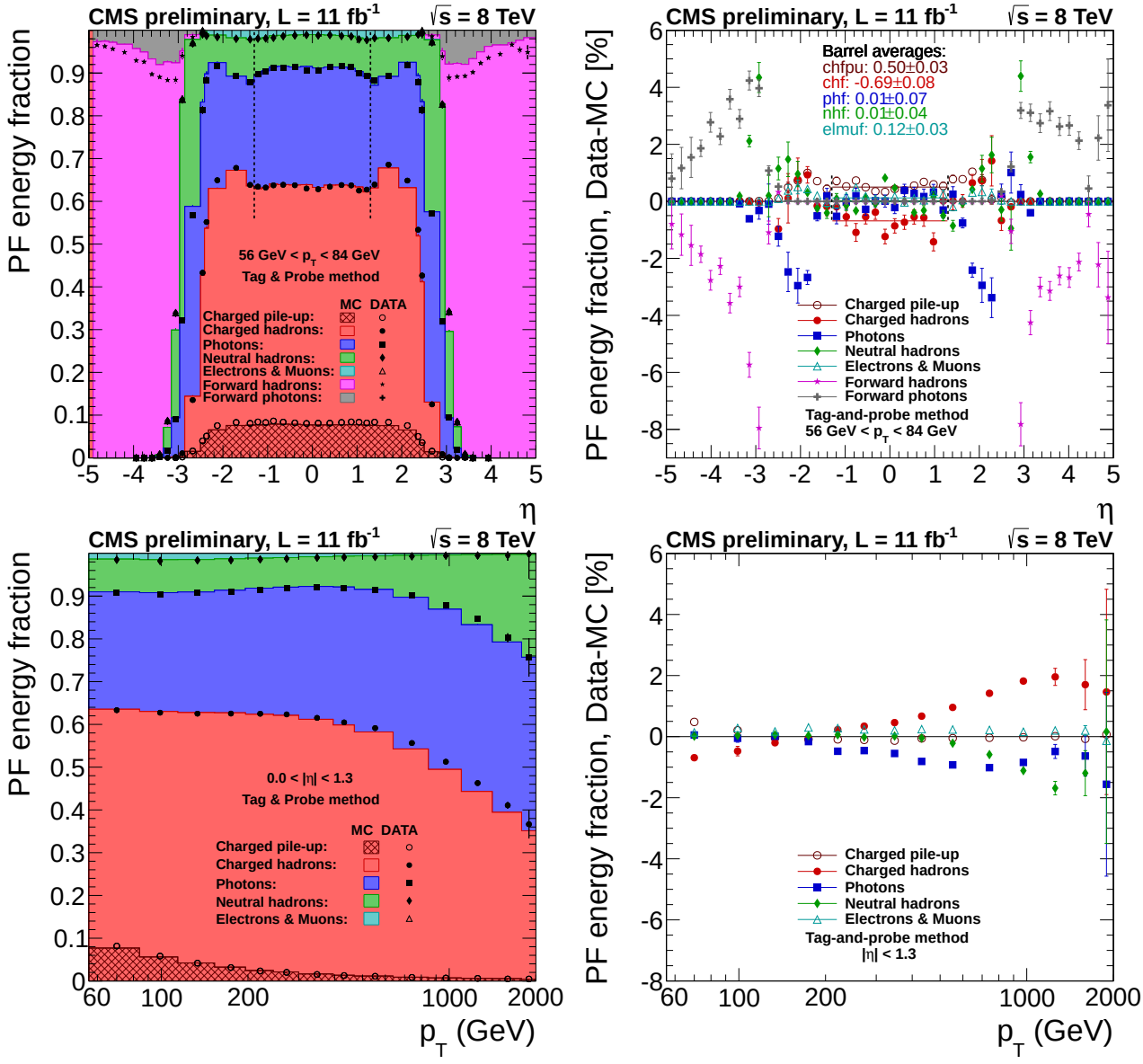

Figure 5.4: Example of energy fraction stack plots (left column) and corresponding difference plots (right column). These results are data versus Pythia MC with $\int Ldt = 11$ fb^{-1}, reconstructed with CMSSW version 5.3.

Chapter 6

Results

In this chapter the results of FastSim parameter variation simulations are shown. The first two plots of each variation scenario show the effect of the variation to the jet response, first plotted as plain response plots where variations to different directions are separate, and then mirrored on top of each other so that the symmetry of the effects is easy to see. Then, the effect of the parameter variation to the jet energy composition is shown in jet energy fraction stack plots, and finally differences between control simulations and varied simulations are highlighted with difference plots. Energy fraction stacks and difference plots are shown against both variables η and p_{T}. For stacks with η in the x-axis also a combination plots including four different jet p_{T} bins are produced and shown in Appendix B.

All the events are generated with the Pythia MC 6 Tune Z2Star event generator and reconstructed with the CMS software version 5.2.5. Jet reconstruction is performed with particle-flow event reconstruction and anti-k_t clustering with radius parameter $R = 0.5$. Also alternative simple calorimeter jet reconstruction method is used for comparing PF jet response to response of straightforwardly reconstructed calorimeter jets, 'calo-jets'. Jet energies are corrected with "START52_V9B_MC" jet energy corrections provided by the CMS jet energy correction group.

Pile-up collisions are generated on top of each event following a measured pile-up distribution from LHC runs of summer 2012. Similar PU is generated to all the scenarios for seeing the effects of PU in our response variation tests.

The simulations were run in the Worldwide LHC Computing Grid (WLCG) and the simulation was divided into ten sub-jobs, so that each sub-job contained simulation 10 000 events. Job configuration was modified so that every parameter variation scenario got the same random generator seeds, which was necessary in order to have comparable results with such a small event count. Simulation jobs were send to the Grid with the CMS Remote Analy-

sis Builder framework CRAB. Each sub-job of 10 000 events takes about 10 hours of processing, which is convenient considering overnight simulations.

6.1 ECAL+HCAL response variation

Effect of single particle response variation in both ECAL and HCAL to jet response is shown in figure 6.1. Figure 6.2 then shows how the ECAL+HCAL response variation affected the jet energy composition.

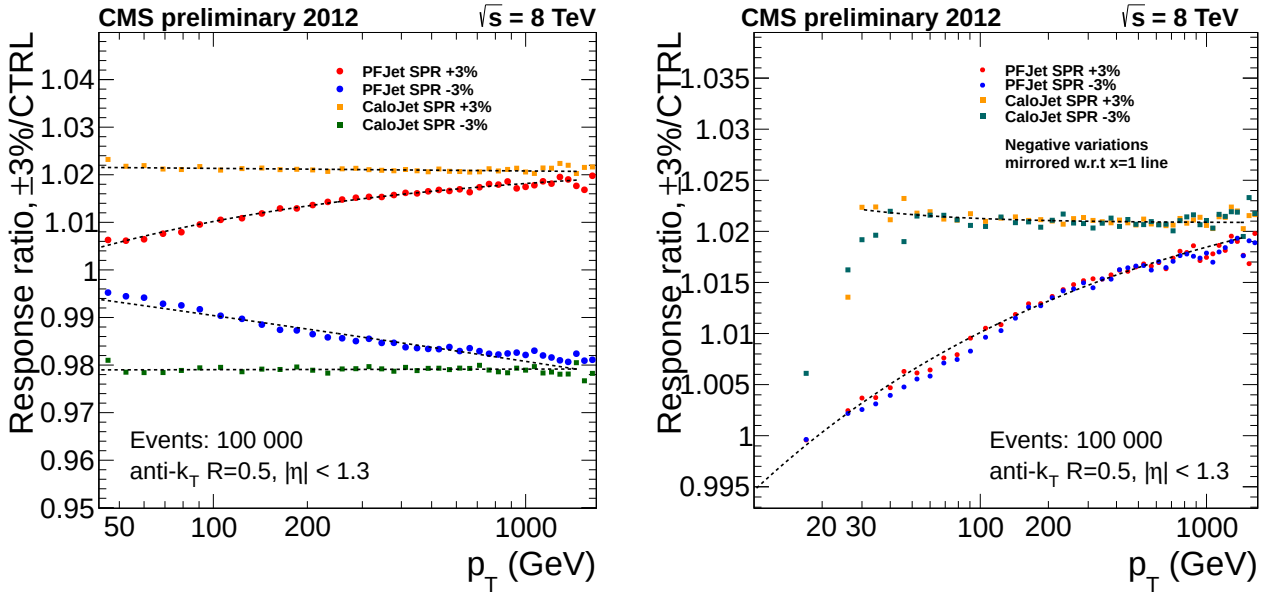

Figure 6.1: Effect of ±3% ECAL+HCAL single particle response variation to jet response. On the left hand side the responses are shown without modifications, and on the right the data points corresponding to negative variation are mirrored so that the symmetry of the effect is clearly visible.

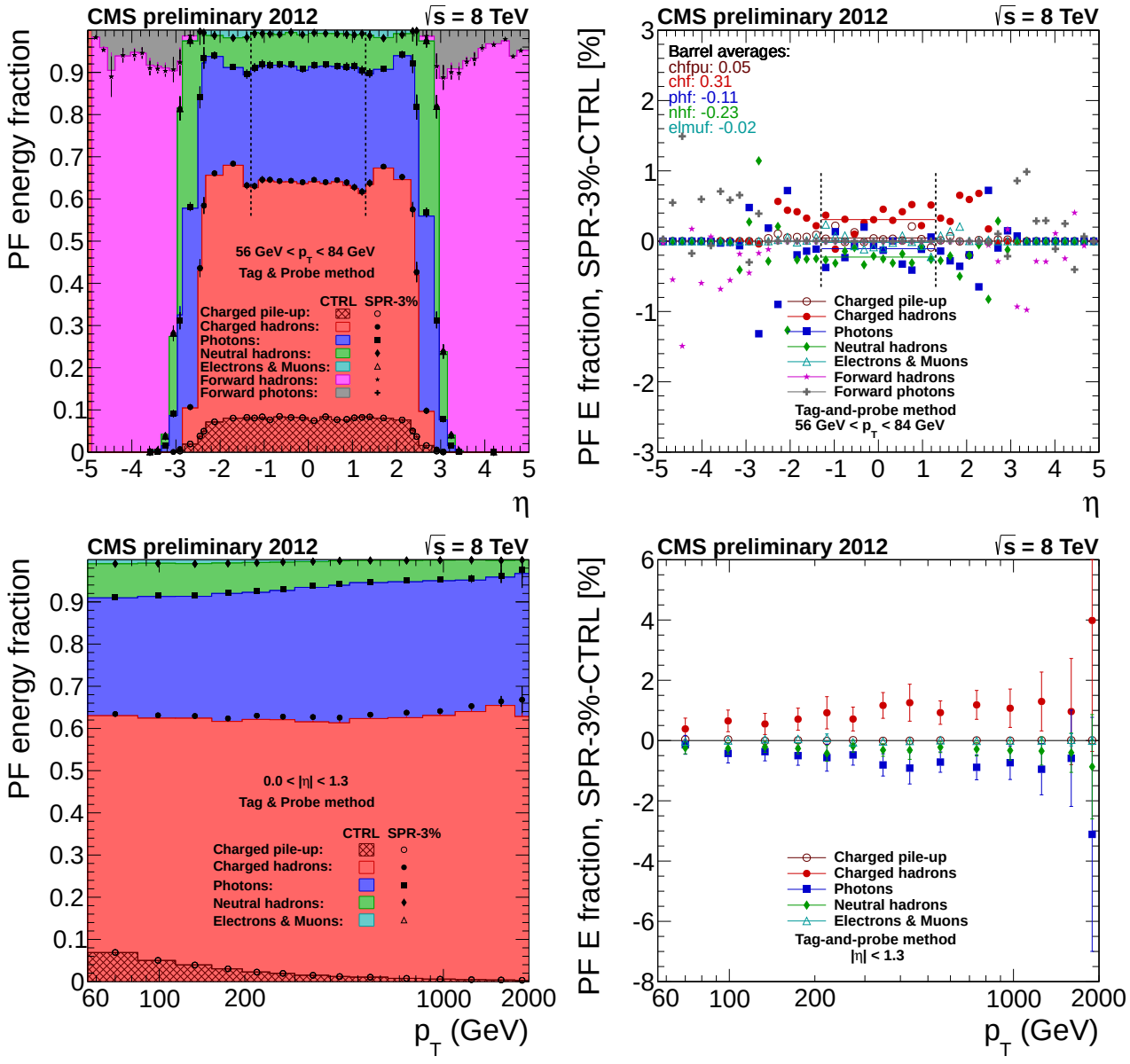

Figure 6.2: Effect of ECAL+HCAL -3% response variation in stack plots (left column) and in corresponding difference plots (right column). Error bars are disabled from the top right figure for clarity. Red stripe at left of the upper left figure is due to insufficient statistics.

6.2 ECAL response variation

Effect of single particle response variation in ECAL to jet response is shown in figure 6.3. Figure 6.4 then shows how the ECAL response variation affected the jet energy composition.

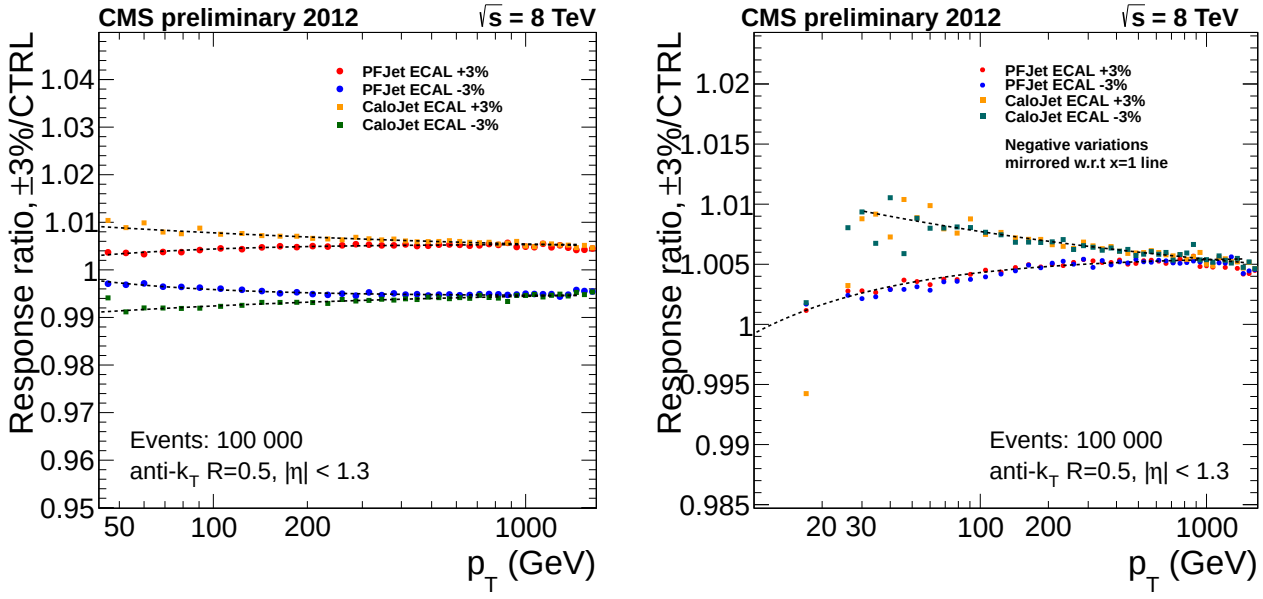

Figure 6.3: Effect of ±3% ECAL single particle response variation to jet response. On the left hand side the responses are shown without modifications, and on the right the data points corresponding to negative variation are mirrored so that the symmetry of the effect is clearly visible.

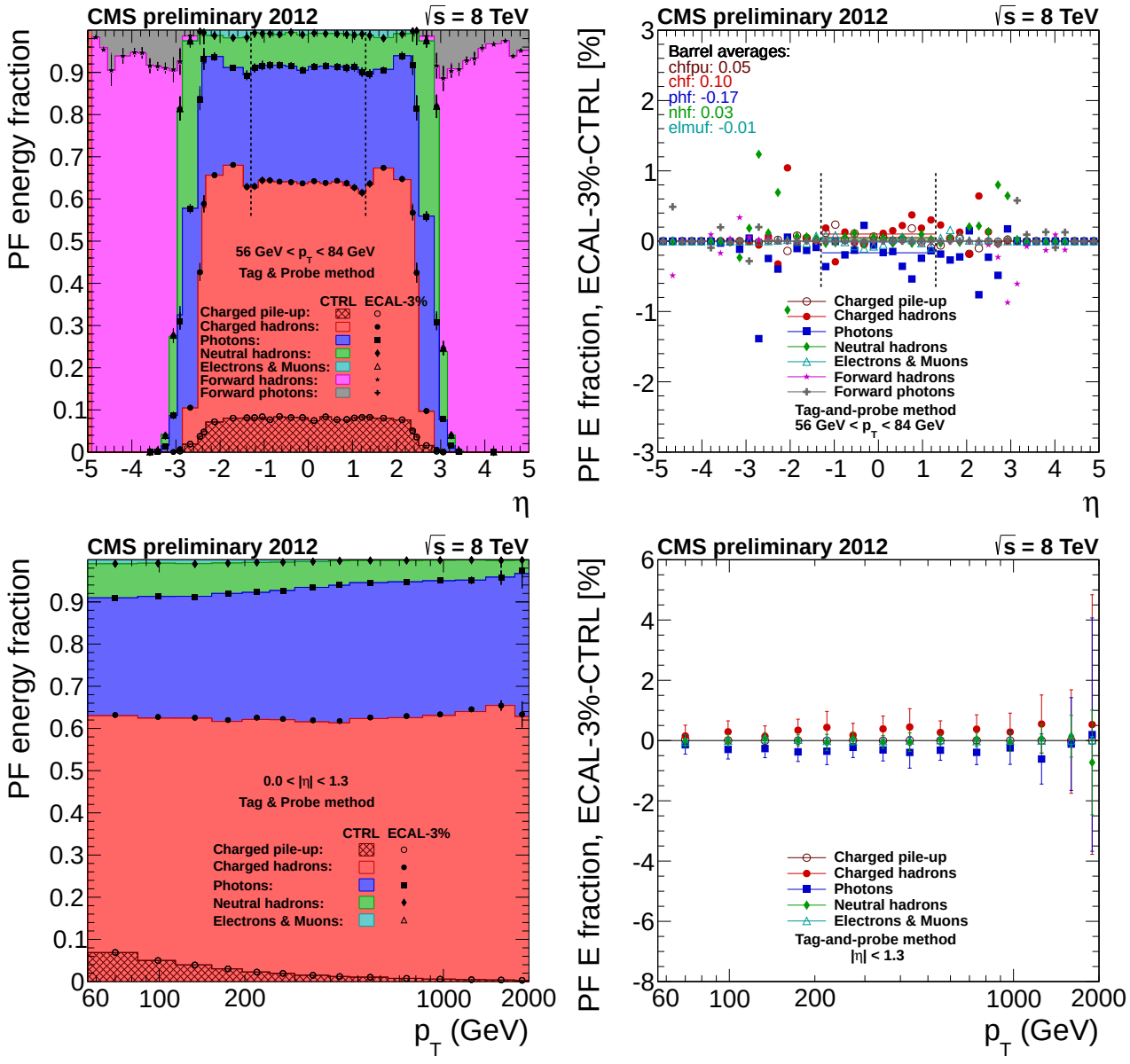

Figure 6.4: Effect of ECAL -3% response variation in stack plots (left column) and in corresponding difference plots (right column). Error bars are disabled from the top right figure for clarity. Red stripe at left of the upper left figure is due to insufficient statistics.

6.3 HCAL response variation

Effect of single particle response variation in HCAL to jet response is shown in figure 6.5. Figure 6.6 then shows how the HCAL response variation affected the jet energy composition.

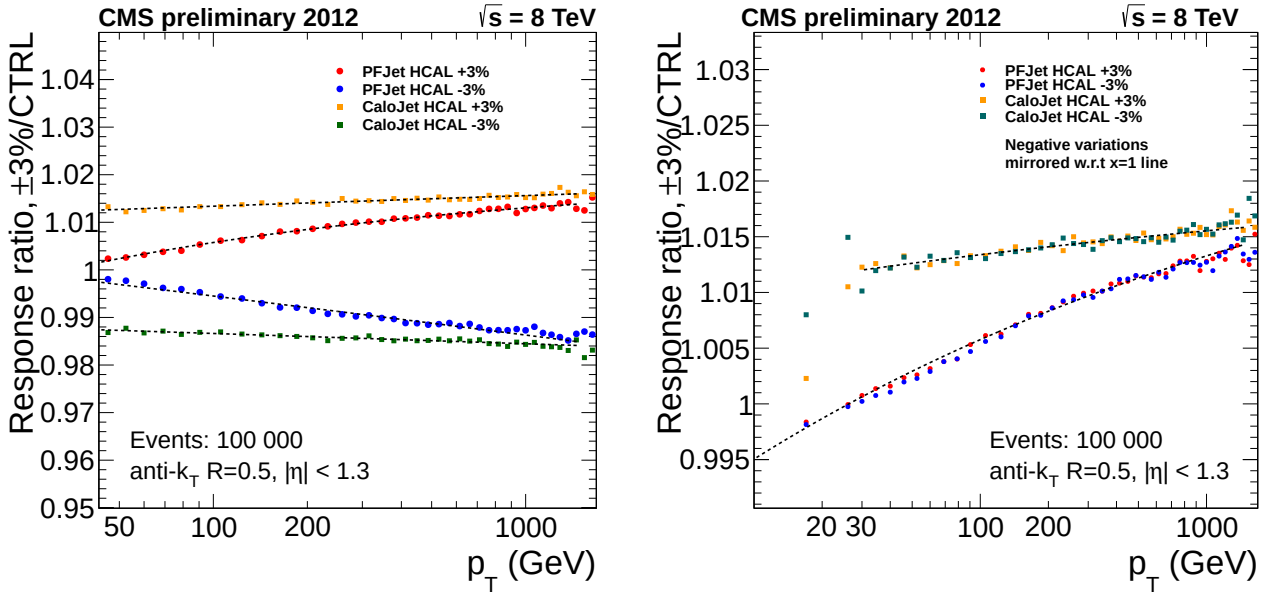

Figure 6.5: Effect of ±3% HCAL single particle response variation to jet response. On the left hand side the responses are shown without modifications, and on the right the data points corresponding to negative variation are mirrored so that the symmetry of the effect is clearly visible.

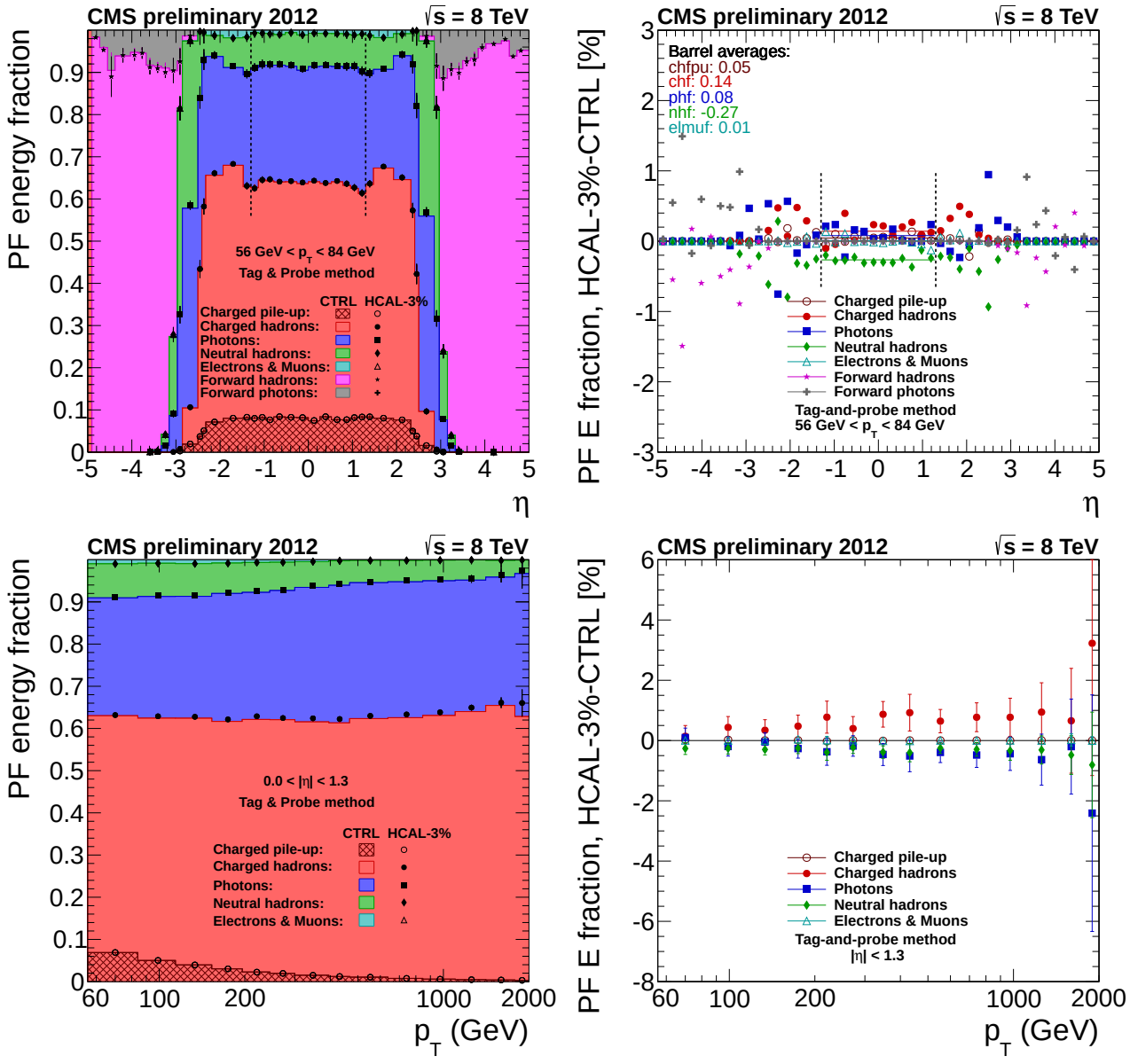

Figure 6.6: Effect of HCAL -3% response variation in stack plots (left column) and in corresponding difference plots (right column). Error bars are disabled from the top right figure for clarity. Red stripe at left of the upper left figure is due to insufficient statistics.

Chapter 7

Discussion

Starting from the effect of detector response variation to jet response, figure 6.1 is an important control plot indicating that the FastSim approach is a feasible method for jet response studies. We have successfully reproduced the jet response behavior observed in a similar study [7], which indicates that we are on the right track. The right hand side plot of Figure 6.1 clearly shows that the response is symmetric at least with variations of $\pm 3\%$, which is a good property for calorimetry.

Figures 6.3 and 6.5 give valuable information on how the change in the jet response is distributed between ECAL and HCAL. As can be seen, the division is far from trivial: the apparent linearity of calo-jet response is a coincidence where nonlinearities of ECAL and HCAL cancel out, and in the case of PF jets the nonlinearity arises mostly from the ECAL variation and effect to HCAL response is nearly linear. At high p_{T} both jet reconstruction methods approach each other, which can be explained by noticing that at high energies the tracks are nearly straight, and calorimeter resolution becomes the dominating factor to jet response. This causes PF jets to behave like calo-jets, because exploiting momentum information from tracks becomes impossible for the PF method. Relative jet energy deposition goes down in ECAL and rises in HCAL as a function of transverse momentum. This results from increasing length of the hadronic showers; shower length scales with $\log(p_{\mathrm{T}})$ and as a shower gets longer, bigger proportion of its energy is deposited in HCAL. These effects are new results that had not been derived before with detector simulation. Results are not completely unexpected, but again fortify our understanding of the behavior of different jet reconstruction approaches.

One serious drawback in the jet response study is that we are lacking estimation of uncertainty altogether. This is due to the fact that error estimation methods of conventional statistical analysis do not apply for our special case,

where we produce statistically correlated events with just slightly modified simulation conditions. The fact that the observed points in figures 6.1, 6.3 and 6.5 seem to lie nicely along the fitted curve without large oscillations is reassuring, but still analytical error estimation is needed. One possible approach for this is simulating response variation in smaller, say 10 000 event simulations and then calculation standard deviation of the observed points. This straightforward approach is simple to implement and also statistically satisfying, but just has not yet been implemented. Another idea for deriving uncertainty estimation is to compare directly jets that are identical in the generator level but have been reconstructed with different detector parameters.

Now considering the effects on jet energy composition, the findings are very encouraging and indicate the potential of the proposed jet composition analysis approach. As shown in the stack plots of figures 6.2, 6.4 and 6.6, the changes of $\pm 3\%$ in response leads to changes barely observable at 0%-100% scale. Still the simulation process with only 100 000 events makes a clear difference in compositions when composition difference is looked at a scale of a few per cent. The important result here is that the effects of the FastSim parameter variation does not get overwhelmed by statistical fluctuations at least in the low energy regime, as shown in the right hand side plots of figures 6.2, 6.4 and 6.6. Jet composition comparisons in Appendix B show that the current statistics is not sufficient for high-p_{T} regime. In future work this can be fixed either by simply generating more events or configuring the event generator to favor harder jets so that the overall jet spectrum is nearly flat.

Perhaps the most important result of this analysis can be seen by comparing the p_{T} fraction stacks in figures 5.4 and 6.2, where in the first one in the comparison are real data and full MC simulation. There for example the fraction of charged hadrons is higher in data than in MC at the high energy region. We reproduced the same behavior with detector simulations as figure 6.2 indicates. We can investigate this result even further by comparing how the effect divides between ECAL and HCAL. When comparing figures 6.4 and 6.6 we see that the main part of the difference arises from HCAL response variation. This can be interpreted as evidence for a hypothesis stating that lowering sensitivity of the CMS hadronic calorimeter in MC would yield better agreement between data and simulation. If the Monte Carlo model is considered trustworthy, this would enhance the accuracy of jet measurements.

Another clear discrepancy between data and simulation (Fig. 5.4, bottom left) is that energy carried by photons in jets is low in data at around $p_{\mathrm{T}} = 200 - 500$ GeV. This effect is not reproduced by our simulations, or then it is hidden behind statistical uncertainties. This is something that could

arise from miscalibration not in the single particle response but in ECAL photon response or in the alignment of the tracking system. We work on applying a similar strategy for simulating photon response variation and tracker misalignment, but some technical difficulties have to be overcome before seeing results from these scenarios.

Results of the FastSim parameter variation simulations give encouraging support to our hypothesis that the jet energy scale uncertainty could be reduced by first pinpointing and then correcting small differences in energy composition between measured and simulated jets. However reproducing the observed discrepancies in simulations does not assure that the sources of the differences are really those that we looked at and quantitative calculations are required for any strong statements. Thus, the next phase in this research is to find a way for calculating how much a miscalibration in some detector element would affect the jet energy composition. If such a calculation would give similar results that our simulations yield, we would have solid evidence for our hypotheses. Should it so happen, we could make a proposition to the CMS collaboration for fine-tuning sensitivities of the detector parameters we have found to be off their optimal values. This would decrease uncertainties in the jet energy calibration and thus have a significant positive impact on all of the CMS physics analyses that make use of jets.

Chapter 8

Conclusions

Jet energy calibration and jet energy scale uncertainties are more and more important concepts in the field of high energy physics. Especially considering the forthcoming LHC upgrade and proton-proton collisions with center-of-mass energy of 13.5 TeV, jet physics is right at the center of attention.

Jet reconstruction is an involved process where the the experimental set-up has to be understood to the finest detail and every possible effect needs to be corrected for. Considering the complexity of the task, accuracy of the current reconstruction process is downright astonishing, as figure 5.4 shows. Still there is room for improvements, especially in decreasing jet energy scale uncertainty for high transverse momentum jets which are potential sources of new physics. We attack the problem with a sort of bottom-up method in which we investigate the contributions of different particle types to jet energy separately. Analyzing jet composition to this extent would not be possible without the state-of-the-art event reconstruction algorithm particle-flow, that is able to reconstruct individual particles and hence provide information on jets' inner structure.

We simulated detector miscalibration in three different miscalibration scenarios using the lightweight FastSim detector simulation framework. Feasibility of the FastSim approach was verified by comparing its results to similar simulation made with the full CMS software. We were able to reproduce the discrepancy in charged hadron fraction that is seen in data-MC comparison in section 5.5. Other observed discrepancies were not reproduced or are hidden behind statistical fluctuations. For this we will continue the FastSim parameter variation research to other aspects of the CMS experiment, such as the ECAL photon sensitivity and tracker alignment.

Apart from applying the technique to other miscalibration scenarios, the research will go on with development of a method for estimating analytically the effect of miscalibrations to jet energy composition. This could fortify

the hypotheses that are supported by findings in the simulations and lead to re-adjusting parts of the CMS detector.

Outlook

The first phase of proton-proton runs at the LHC came to and end in December 2012 and the Long Shutdown 1 will keep the LHC calm and quiet for two years, at least from the perspective of the physicists. Although the physics community has a rare opportunity for a well deserved sit-down after years of persistent diligence and discoveries, we cannot rest on our laurels for long. Protons are stripped from hydrogen again in 2015 and collided with energies nearly double to the current world record. Pile-up conditions will be harsher than ever, energies of single jets will rise to several TeV and we will see the structure of the proton closer than ever before. We need to be prepared. After a couple of deep breaths, a bit of traveling perhaps, the community will continue the persistent work, adapt the analyses for the new energy frontier and develop new methods for exploring the unknown.

Faithful to our employer, playground and muse, we will maintain our integrity and stay open to whatever might be the next surprises planned for us by her, Mother Nature.

Bibliography

[1] Bahr, M., Gieseke, S., Gigg, M., Grellscheid, D., Hamilton, K., et al. Herwig++ Physics and Manual. *Eur.Phys.J. C58* (2008), 639–707.

[2] Beringer, J., et al. Review of particle physics. *Phys. Rev. D 86* (Jul 2012), 010001. (Particle Data Group).

[3] Cacciari, M., Salam, G. P., and Soyez, G. The anti-k_t jet clustering algorithm. *Journal of High Energy Physics 2008*, 04 (2008), 063.

[4] CERN press release. CERN experiments observe particle consistent with long-sought Higgs boson, July 4 2012.

[5] CMS Collaboration. "The CMS experiment at the CERN LHC". *JINST 03* (2008), S08004. doi:10.1088/1748-0221/3/08/S08004.

[6] CMS Collaboration. Precise mapping of the magnetic field in the cms barrel yoke using cosmic rays. *Journal of Instrumentation 5*, 03 (2010), T03021.

[7] CMS Collaboration. Determination of jet energy calibration and transverse momentum resolution in cms. *JINST 6 11002* (2011). arXiv:1107.4277.

[8] Dorney, B. Anatomy of a jet in cms. http://www.quantumdiaries.org/2011/06/01/anatomy-of-a-jet-in-cms/, Accessed 11.8.2012.

[9] Dyson, F. J. The Radiation Theories of Tomonaga, Schwinger, and Feynman. *Phys. Rev. 75* (Feb 1949), 486–502.

[10] Englert, F., and Brout, R. Broken Symmetry and the Mass of Gauge Vector Mesons. *Physical Review Letters 13* (Aug. 1964), 321–323.

[11] European Organisation for Nuclear Research. CERN Document Server. `http://cdsweb.cern.ch`, Accessed 2.8.2012.

[12] Greiner, W., and Müller, B. *Gauge Theory of Weak Interactions*. Physics and Astronomy Online Library. Springer, 2000.

[13] Guralnik, G. S., Hagen, C. R., and Kibble, T. W. Global Conservation Laws and Massless Particles. *Physical Review Letters 13* (Nov. 1964), 585–587.

[14] Halzen, F., and Martin, A. *Quarks and leptons: an introductory course in modern particle physics*. Wiley, 1984.

[15] Higgs, P. W. Broken Symmetries and the Masses of Gauge Bosons. *Phys.Rev.Lett. 13* (1964), 508–509.

[16] Incandela, J. Results from the CMS experiment at the LHC. Talk in the 521. WE Heraeus seminar: The First Resuls from the Large Hadron Collider, 2012.

[17] Liddell, H. G., and Scott, R. A Greek-English Lexicon. *Perseus Digital Library*. `www.perseus.tufts.edu/`.

[18] Martin, B., and Shaw, G. *Particle Physics*. Manchester Physics Series. Wiley, 2008.

[19] S. Agostinelli *et al.* Geant4 – a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 506*, 3 (2003), 250 – 303.

[20] Schaelicke, A., Gleisberg, T., Hoeche, S., Schumann, S., Winter, J., et al. Event generator for particle production in high-energy collisions. *Prog.Part.Nucl.Phys. 53* (2004), 329–338.

[21] Sjöstrand, T., Ask, S., Corke, R., Mrenna, S., Prestel, S., and Skands, P. Home page of PYTHIA event generator. Lund University, `http://home.thep.lu.se/~torbjorn/Pythia.html`. Accessed 2.8.2012.

[22] Sjostrand, T., Mrenna, S., and Skands, P. Z. A Brief Introduction to PYTHIA 8.1. *Comput.Phys.Commun. 178* (2008), 852–867.

[23] University of Groeningen. Hadron Structure with Antiprotons. `http://www.rug.nl/kvi/research/hnp/research/panda`, Accessed 8.8.2012.

[24] Vidal, X. C., and Manzano, R. C. Taking a closer look at LHC. `http://www.lhc-closer.es/` (referred 1.11.2012).

[25] Voutilainen, M. "Measurement of the inclusive jet cross section in $p\bar{p}$ collisions at $\sqrt{s}$=1.96 TeV". *Helsinki University of Technology (Espoo, Finland)* (2008). Ph.D Thesis.

[26] Wikimedia Commons. `http://commons.wikimedia.org`. Accessed 10.8.2012.

Appendix A

Configurations for FastSim simulations

This appendix contains examples of CMSSW (fastsim.py) and WLCG (crab.cfg) configuration files for reproducing the FastSim parameter variation simulations. When these configuration files are in a CMSSW 5.2.5 environment and necessary WLCG and CRAB initializations are done, the task can be sent to the WLCG with the command 'crab -create -submit'.

```
########## fastsim.py #########
import FWCore.ParameterSet.Config as cms
process = cms.Process('JETSIM')
# import of standard configurations
process.load('Configuration.StandardSequences.Services_cff')
process.load('SimGeneral.HepPDTESSource.pythiapdt_cfi')
process.load('FWCore.MessageService.MessageLogger_cfi')
process.load('FastSimulation.Configuration.EventContent_cff')
process.load('FastSimulation.PileUpProducer.
    PileUpSimulator_NoPileUp_cff')

# Include pile-up cff file, where custom PU is defined
process.load('FastSimulation.PileUpProducer.
    PileUpSimulator_8TeV_cff')
process.load('FastSimulation.PileUpProducer.PileUpProducer_cff')

process.load('FastSimulation.Configuration.Geometries_START_cff
    ')
process.load('Configuration.StandardSequences.
    MagneticField_38T_cff')
process.load('Configuration.StandardSequences.Generator_cff')
process.load('GeneratorInterface.Core.genFilterSummary_cff')
process.load('FastSimulation.Configuration.FamosSequences_cff')
process.load('IOMC.EventVertexGenerators.
    VtxSmearedParameters_cfi')
```

```
process.load('FastSimulation.Configuration.HLT_GRun_cff')
process.load('Configuration.StandardSequences.
    FrontierConditions_GlobalTag_cff')

nevt = 10000
ecaltweak = 1.03 # Choose variation of hadron response here
hcaltweak = 1.03
outputfile = 'SPR103_100k.root'

process.maxEvents = cms.untracked.PSet(
    input = cms.untracked.int32(nevt))
# Input source
process.source = cms.Source("EmptySource")
process.options = cms.untracked.PSet(
)
# Production Info
process.configurationMetadata = cms.untracked.PSet(
    version = cms.untracked.string('$Revision: 1.303.2.7 $'),
    #annotation = cms.untracked.string('
        QCD_Pt_30_TuneZ2_7TeV_pythia6_cff.py nevts:10'),
    annotation = cms.untracked.string('
        QCD_Pt_30_TuneZ2Star_8TeV_pythia6_cff.py nevts:%i' % nevt
        ),
    name = cms.untracked.string('PyReleaseValidation')
)
# Output definition
process.RECOSIMoutput = cms.OutputModule("PoolOutputModule",
    splitLevel = cms.untracked.int32(0),
    eventAutoFlushCompressedSize = cms.untracked.int32(5242880),
    outputCommands = process.RECOSIMEventContent.outputCommands,
    #fileName = cms.untracked.string('424QCD15_plus3SPRecal.root
        '),
    fileName = cms.untracked.string( outputfile ),
    dataset = cms.untracked.PSet(
        filterName = cms.untracked.string(''),
        dataTier = cms.untracked.string('GEN-SIM-DIGI-RECO')
    ),
    SelectEvents = cms.untracked.PSet(
        SelectEvents = cms.vstring('generation_step')
    )
)
# Other statements
process.famosSimHits.SimulateCalorimetry = True
process.famosSimHits.SimulateTracking = True
process.simulation = cms.Sequence(process.simulationWithFamos)
process.HLTEndSequence = cms.Sequence(process.
    reconstructionWithFamos)
```

```
process.Realistic8TeV2012CollisionVtxSmearingParameters.type =
    cms.string("BetaFunc")
process.famosSimHits.VertexGenerator = process.
    Realistic8TeV2012CollisionVtxSmearingParameters
process.famosPileUp.VertexGenerator = process.
    Realistic8TeV2012CollisionVtxSmearingParameters

process.GlobalTag.globaltag = 'START52_V9B::All'

process.generator = cms.EDFilter("Pythia6GeneratorFilter",
    pythiaPylistVerbosity = cms.untracked.int32(0),
    filterEfficiency = cms.untracked.double(1.0),
    pythiaHepMCVerbosity = cms.untracked.bool(False),
    comEnergy = cms.double(8000.0),
    crossSection = cms.untracked.double(60411000.0),
    maxEventsToPrint = cms.untracked.int32(0),
    PythiaParameters = cms.PSet(
        pythiaUESettings = cms.vstring('MSTU(21)=1      ! Check
            on possible errors during program execution',
            'MSTJ(22)=2      ! Decay those unstable particles',
            'PARJ(71)=10 .   ! for which ctau  10 mm',
            'MSTP(33)=0      ! no K factors in hard cross
                sections',
            'MSTP(2)=1       ! which order running alphaS',
            'MSTP(51)=10042 ! structure function chosen (
                external PDF CTEQ6L1)',
            'MSTP(52)=2      ! work with LHAPDF',
            'PARP(82)=1.832 ! pt cutoff for multiparton
                interactions',
            'PARP(89)=1800. ! sqrts for which PARP82 is set',
            'PARP(90)=0.275 ! Multiple interactions: rescaling
                power',
            'MSTP(95)=6      ! CR (color reconnection parameters)
                ',
            'PARP(77)=1.016 ! CR',
            'PARP(78)=0.538 ! CR',
            'PARP(80)=0.1    ! Prob. colored parton from BBR',
            'PARP(83)=0.356 ! Multiple interactions: matter
                distribution parameter',
            'PARP(84)=0.651 ! Multiple interactions: matter
                distribution parameter',
            'PARP(62)=1.025 ! ISR cutoff',
            'MSTP(91)=1      ! Gaussian primordial kT',
            'PARP(93)=10.0   ! primordial kT-max',
            'MSTP(81)=21     ! multiple parton interactions 1 is
                Pythia default',
            'MSTP(82)=4      ! Defines the multi-parton model'),
        processParameters = cms.vstring(
            'MSEL=1          ! QCD hight pT processes',
```

```
                'CKIN(3)=15      ! minimum pt hat for hard
                    interactions',
                'CKIN(4)=4000    ! maximum pt hat for hard
                    interactions',
                'MSTP(142) = 2   ! Turns on the PYWEVT Pt
                    reweighting routine'
                ),
            CSAParameters = cms.vstring(
                'CSAMODE = 7     ! towards a flat QCD spectrum',
                'PTPOWER = 4.5   ! reweighting of the pt spectrum'
                ),
            parameterSets = cms.vstring(
                'pythiaUESettings',
                'processParameters',
                'CSAParameters'
                )
        )
)
process.ProductionFilterSequence = cms.Sequence(process.
    generator)

# Adding the part for +/- 3% Single Pion Response
process.famosSimHits.Calorimetry.CalorimeterProperties.RespCorrP
    =(
            1.0,        ecaltweak,  hcaltweak,
            10.0,       ecaltweak,  hcaltweak,
            100.0,      ecaltweak,  hcaltweak,
            1000.0,     ecaltweak,  hcaltweak
            )
# Path and EndPath definitions
process.generation_step = cms.Path(process.pgen_genonly)
process.reconstruction = cms.Path(process.
    reconstructionWithFamos)
process.genfiltersummary_step = cms.EndPath(process.
    genFilterSummary)
process.RECOSIMoutput_step = cms.EndPath(process.RECOSIMoutput)

# Schedule definition
process.schedule = cms.Schedule(process.generation_step,process.
    genfiltersummary_step)
process.schedule.extend(process.HLTSchedule)
process.schedule.extend([process.reconstruction,process.
    RECOSIMoutput_step])
# filter all path with the production filter sequence
for path in process.paths:
        getattr(process,path)._seq = process.
            ProductionFilterSequence * getattr(process,path)._seq
```

```
########## crab.cfg ##########

[CRAB]
jobtype                 = cmssw
scheduler = glite

[CMSSW]

datasetpath             = None
pset                    = fastsim.py
output_file             = SPR103_100k.root
total_number_of_events  = 100000
events_per_job          = 10000
# Try to preserve identical seed initialization for each
    submission
increment_seeds                 = externalLHEProducer,generator,
    VtxSmeared,LHCTransport,hiSignalLHCTransport,g4SimHits,mix,
    mixData,simSiStripDigis,simSiPixelDigis,
    simEcalUnsuppressedDigis,simHcalUnsuppressedDigis,
    simMuonDTDigis,simMuonCSCDigis,simMuonRPCDigis,simCastorDigis
    ,hiSignal,hiSignalG4SimHits,famosPileUp,mixGenPU,famosSimHits
    ,siTrackerGaussianSmearingRecHits,ecalRecHit,
    ecalPreshowerRecHit,hbhereco,horeco,hfreco,paramMuons,
    l1ParamMuons,MuonSimHits,simBeamSpotFilter

[USER]
copy_data               = 1
return_data             = 0
email                   = juska@cern.ch
storage_element         = madhatter.csc.fi
storage_path            = /srm/managerv2?SFN=/pnfs/csc.fi/data/
    cms/store/user/juska/
user_remote_dir         = gridruns/SPR103

[GRID]
ce_white_list           = jade-cms.hip.fi
```

Appendix B

Jet composition in four p_{T} bins

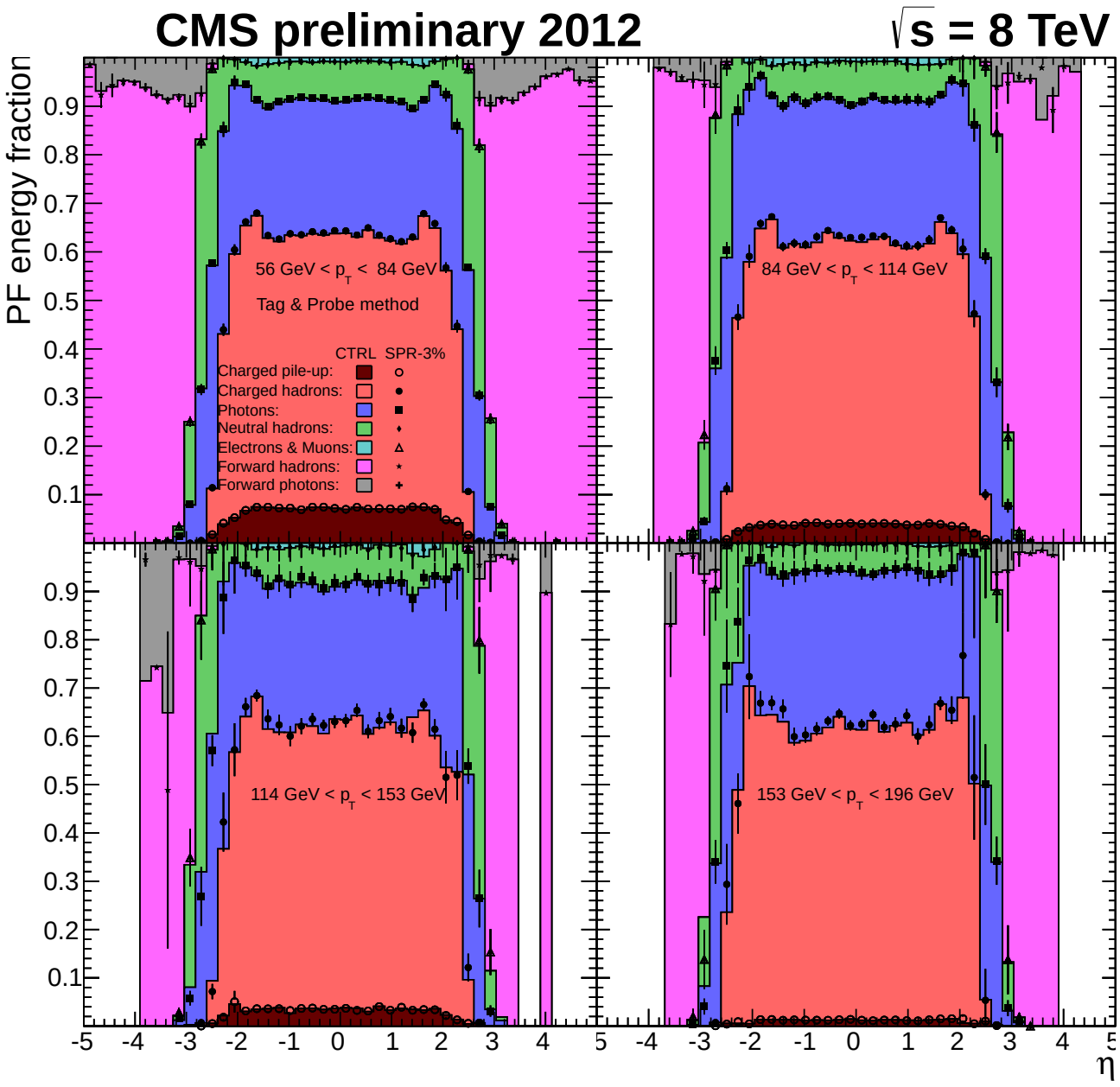

Figure B.1: Effect of ECAL+HCAL -3% response variation to jet composition in four jet transverse momentum bins. White areas near the edges indicate insufficient statistics.

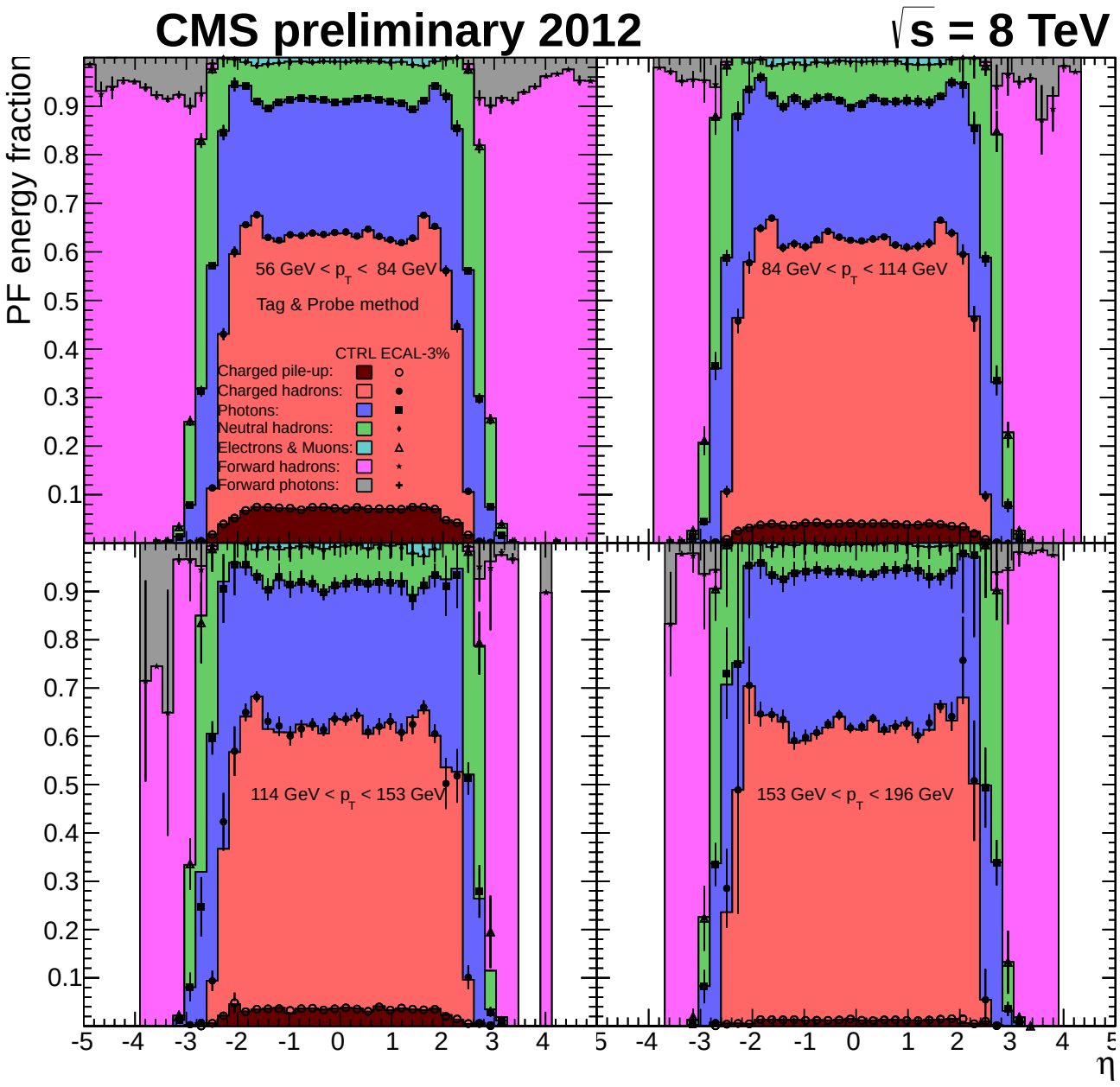

Figure B.2: Effect of ECAL -3% response variation to jet composition in four jet transverse momentum bins. White areas near the edges indicate insufficient statistics.

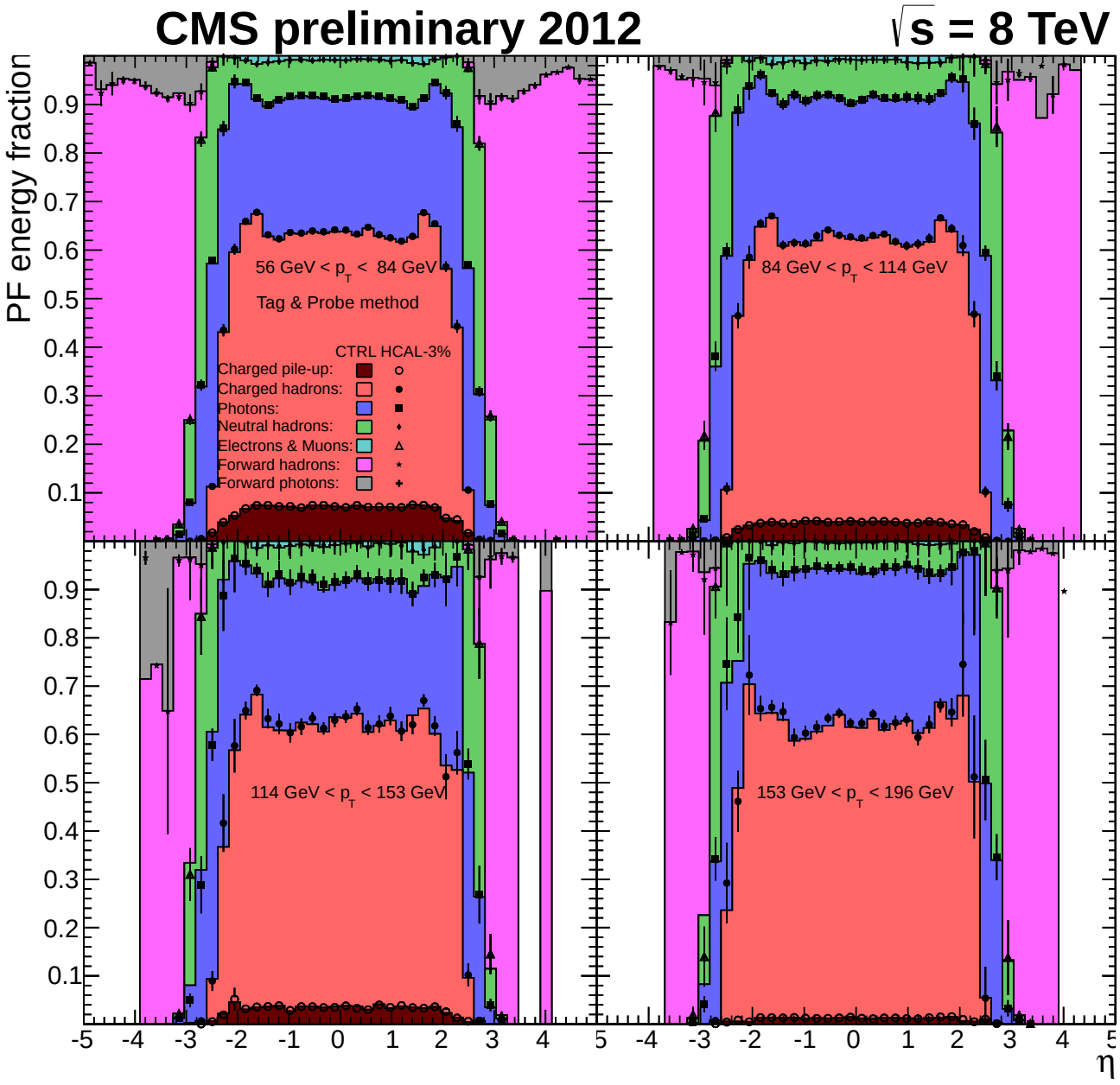

Figure B.3: Effect of HCAL -3% response variation to jet composition in four jet transverse momentum bins. White areas near the edges indicate insufficient statistics.