

Applied physics

Jet Particology & Search for New Massive Particles

Jet composition studies and searches for new physics in dijet events using 13 TeV pp collisions with the CMS detector at the CERN LHC

Juska Pekkanen



Aalto University

DOCTORAL
DISSERTATIONS

Jet Particology & Search for New Massive Particles

Jet composition studies and searches for new physics in dijet events using 13 TeV pp collisions with the CMS detector at the CERN LHC

Juska Pekkanen

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Science, at a public examination held at the auditorium D101 of the Physicum building, Gustaf Hållströmin katu 2 A, Helsinki, on Tuesday December 5th 2017, at 12 o'clock noon.

Aalto University
School of Science
Applied physics

Supervising professor

Professor Filip Tuomisto, Aalto University School of Science, Finland

Thesis advisor

Assistant Professor Mikko Voutilainen, University of Helsinki, Finland

Preliminary examiners

Professor Ian Brock, University of Bonn, Germany

Doctor Caterina Doglioni, Lund University, Sweden

Opponent

Doctor Christopher Young, CERN, Switzerland

Aalto University publication series

DOCTORAL DISSERTATIONS 198/2017

© 2017 Juska Pekkanen

ISBN 978-952-60-7663-8 (printed)

ISBN 978-952-60-7662-1 (pdf)

ISSN-L 1799-4934

ISSN 1799-4934 (printed)

ISSN 1799-4942 (pdf)

<http://urn.fi/URN:ISBN:978-952-60-7662-1>

Images: The CMS Collaboration

Unigrafia Oy

Helsinki 2017

Finland

Publication orders (printed book):

juska@cern.ch



Author

Juska Pekkanen

Name of the doctoral dissertation

Jet Particology & Search for New Massive Particles

Publisher School of Science

Unit Applied physics

Series Aalto University publication series DOCTORAL DISSERTATIONS 198/2017

Field of research Experimental particle physics

Manuscript submitted 12 September 2017

Date of the defence 5 December 2017

Permission to publish granted (date) 13 November 2017

Language English

☒ **Monograph**

☐ **Article dissertation**

☐ **Essay dissertation**

Abstract

When we collide protons at the Large Hadron Collider (LHC) of CERN, the European Organization for Nuclear Research, what actually collides are the quarks and gluons, collectively *partons*, that make up the protons. The behavior of partons is governed by the rules of Quantum Chromodynamics (QCD), the theory of the strong force.

QCD has a peculiar property called *confinement*, which prevents partons from moving freely and forces the formation of composite particles called *hadrons*. Confinement has dramatic consequences for particle physics. Whenever we try to separate a quark from a proton, what results is a burst of hadrons that we call a *jet*. As the LHC collides only hadrons, jets are present in virtually every collision event. Jets are thus building blocks of the majority of the LHC physics analyses, and a thorough understanding of their behavior is very important for achieving precise physics results. By virtue of the particle-flow event reconstruction used in the Compact Muon Solenoid (CMS) experiment, for the first time we can actually take a good look inside the jets and directly measure what the jets are made of, what kind of particles carry most of the energy etc. We call these in-depth jet studies *jet particology*. In the first half of this thesis I present a series of pioneering studies of jets at the particle level, including measurements of jet energy composition, jet response and mitigation of the effects of *pileup*, the unwanted noise caused by simultaneous proton-proton collisions. The studies on the one hand demonstrate the excellent performance of the CMS jet reconstruction, detector simulation and the particle-flow algorithm, but on the other hand show where further studies are needed and open a series of questions and ideas for future jet particology research.

In the second half of the thesis we apply our understanding of jets to searches for new massive elementary particles, which are predicted by theoretical models that try to explain the shortcoming of the current theoretical understanding. For maximal discovery potential, our strategy is to look for *everything* that interacts with quarks and gluons in an analysis called the *dijet resonance search*. We study 13 *teraelectronvolt* (TeV) proton-proton collisions of the LHC and look for events with back-to-back jets, which is the experimental signature of a heavy particle decaying to partons. This search reaches to the highest energies ever achieved in a collider and we observe events with up to 8 TeV dijet masses. Unfortunately we do not find evidence of production of new massive particles, and thus we exclude a series of theory predictions up to 8 TeV resonance masses.

Keywords jets, jet composition, particle-flow, global event reconstruction, QCD, beyond standard model, new physics searches, dijet resonance search, CMS, LHC, CERN

ISBN (printed) 978-952-60-7663-8

ISBN (pdf) 978-952-60-7662-1

ISSN-L 1799-4934

ISSN (printed) 1799-4934

ISSN (pdf) 1799-4942

Location of publisher Helsinki

Location of printing Helsinki

Year 2017

Pages 145

urn <http://urn.fi/URN:ISBN:978-952-60-7662-1>

Tekijä

Juska Pekkanen

Väitöskirjan nimi

Hiukkasryöppyjen tutkimus ja uusien raskaiden alkeishiukkasten etsintä

Julkaisija Perustieteiden korkeakoulu**Yksikkö** Teknillinen fysiikka**Sarja** Aalto University publication series DOCTORAL DISSERTATIONS 198/2017**Tutkimusala** Kokeellinen alkeishiukkasfysiikka**Käsitteilyajon pvm** 12.09.2017**Väitöspäivä** 05.12.2017**Julkaisuluvan myöntämispäivä** 13.11.2017**Kieli** Englanti☒ **Monografia**☐ **Artikkeliväitöskirja**☐ **Esseeväitöskirja****Tiivistelmä**

Kun törmäytämme protoneita toisiinsa Euroopan hiukkastutkimuskeskus CERN:n Suuressa Hadronitörmäytinissä (Large Hadron Collider, LHC), todellisuudessa vuorovaikuttavat protonien rakenneosat eli kvarkit ja gluonit, kollektiivisesti *partonit*. Partonien käyttäytymistä kuvaava teoria on kvanttiväridynamiikka (Quantum Chromodynamics, QCD). QCD:lla on erikoinen ominaisuus nimeltään *partonien vankeus*, joka estää kvarkkeja ja gluoneita liikkumasta vapaasti. Tästä syystä partoneita havaitaan vain niiden muodostamissa komposiittihiukkasissa, *hadroneissa*.

Yrittäessämme erottaa partonia protonista seuraa reaktio, jossa syntyy hiukkasryöppy eli *jetti*.

Jettejä syntyy käytännössä jokaisessa LHC:n törmäystapahtumassa ja ne ovat osa lähes jokaista kokeellisen hiukkasfysiikan analyysiä. Tästä syystä hiukkasryöppyjen perusteellinen tuntemus on ensisijaisen tärkeää, kun haluamme mitata hiukkasmaailman ilmiöitä yhä tarkemmin.

CMS-kokeella (Compact Muon Solenoid) on käytössä particle-flow algoritmi, jolla törmäystapahtumat rekonstruoidaan yksittäisten hiukkasten tasolla. Tämän menetelmän ansiosta voimme ensi kertaa tutkia jettejä hiukkastasolla ja esimerkiksi määrittää, miten ryöpyyn energia jakautuu eri hiukkastyypeille. Kutsumme näitä perusteellisia hiukkastason ryöpytutkimuksia englanninkielisellä termillä *jet particology*. Väitöskirjani ensimmäisessä osassa esittelen joukon urauurtavia tutkimuksia jettien koostumuksesta. Tutkin muun muassa niiden energiakoostumusta ja mittausvastetta ilmaisimissamme sekä sitä, miten useiden samanaikaisten hiukkastörmäysten aiheuttamaa taustakohinaa voidaan vähentää. Tutkimustulokset osoittavat toisaalta, että CMS-kokeen particle-flow rekonstruktio ja ilmaisimen tietokonemallinnus toimivat erinomaisesti, mutta toisaalta myös herättävät uusia kysymyksiä ja näyttävät, missä lisäkehitys hiukkasryöppyjen tutkimuksessa on tarpeen.

Väitöskirjan toisessa osassa sovellamme ymmärrystämme ryöpyistä ja etsimme niiden avulla uusia raskaita alkeishiukkasia, joita teorianmallit ennustavat ratkaistakseen nykyiseen ymmärrykseemme liittyviä puutteita. Maksimoidaksemme analyysin mahdollisuudet uuteen löytöön etsimme vastakkaisten hiukkasryöppyjen analyysissä kaikkia mahdollisia hiukkasia, jotka vuorovaikuttavat kvarkkien ja gluonien kanssa. Analysoimme LHC:n tuottamia 13 *teraelektronivoltin* (TeV) protonitörmäyksiä ja etsimme niistä vastakkaisiin suuntiin syntyneitä hiukkasryöppyjä. Tämä on partoneiksi hajoavan raskaan alkeishiukkasen kokeellinen jalanjälki. Tällä analyysistrategialla pystymme tutkimaan korkeampia energia-alueita kuin koskaan ennen hiukkastörmäytinissä on tutkittu. Valitettavasti emme löydä viitteitä uusien raskaiden hiukkasten olemassaolosta, joten suljemme pois teorioiden ennustamia hiukkasia aina 8 TeV:n massa asti.

Avainsanat kokeellinen hiukkasfysiikka, hiukkasryöpyt, kvanttiväridynamiikka**ISBN (painettu)** 978-952-60-7663-8**ISBN (pdf)** 978-952-60-7662-1**ISSN-L** 1799-4934**ISSN (painettu)** 1799-4934**ISSN (pdf)** 1799-4942**Julkaisupaikka** Helsinki**Painopaikka** Helsinki**Vuosi** 2017**Sivumäärä** 145**urn** <http://urn.fi/URN:ISBN:978-952-60-7662-1>

Mummille

Preface

While I frequently find myself in intimidating situations and seemingly life-threatening locations, for me the project of writing a doctoral thesis on experimental particle physics is the most courageous and frightening one that I have ever challenged myself with. While climbing mountains and searching for elementary particles involve very different challenges, for me the key principles are the same. It's best to proceed step by step and in a good company. The bigger the project, the more important it is to take one step at a time, stubbornly refusing to be overwhelmed by the grandeur of the challenge at hand. At least as important for a successful execution of a great endeavour is to have the right people in the expedition. This is where I have been lucky and what makes me deeply indebted for a big group of friends and colleagues.

The unlimited enthusiasm, the never-ending patience, and the unparalleled expertise of my advisor Mikko is the single most important reason why I've made it this far. Moreover, I am grateful for all the people in the HIP CMS project and other colleagues in Finland and at CERN who always give me unconditional support and comfort, regardless of whether I'm facing technical difficulties or all I need is a moment of distraction from all the things that make my brain hurt.

I also wish to thank the Waldemar von Frenckell and Magnus Ehrnrooth foundations and the Alfred Kordelin General Progress and Education Fund for the support that made my research possible financially, and allowed me to gain invaluable international experience in the couple of summer schools and conferences that I have attended.

With the help of my friends and family I manage to sustain a healthy balance between work and spare time. Be it chasing waves or climbing up a cliff, sitting in a cozy bar or in the heat of a sauna, without the friends around me I would not be there. And I would not be me. Dear mom and

dad, brothers and friends, while I maybe use these words sparingly, you mean everything to me and I would be nothing without you.

I dedicate this work to my dear grandmother, who recently passed away, and to my beautiful goddaughter, for whom the journey has just begun. Mummi, I promise you to do my best to stay as the grandson that you were always so proud of. Ninni, I promise you to do my best as a scientist to make the world a better place for you to grow old in.

Helsinki, November 15, 2017,

Juska Pekkanen

Contents

Preface	3
Contents	5
List of symbols and acronyms	7
Author's contributions	11
1. Introduction	15
2. Theory	17
2.1 Standard Model of particle physics	17
2.2 Quantum Chromodynamics	19
2.2.1 Color confinement	20
2.2.2 Asymptotic freedom and parton shower	21
2.2.3 Hadronization	22
2.3 Jet definition	23
2.4 Physics beyond Standard Model	26
2.4.1 Shortcomings in Standard Model	27
2.4.2 String theories	30
2.4.3 Extra dimensions	30
2.4.4 W' and Z' bosons	31
2.4.5 Compositeness models	31
2.4.6 Dark matter	31
3. Experimental setup	33
3.1 Large Hadron Collider	33
3.1.1 Construction	33
3.1.2 Accelerator chain & LHC operation	36
3.2 Compact Muon Solenoid experiment	40

3.2.1	Construction	40
3.2.2	Coordinate system	42
3.2.3	Tracker	42
3.2.4	Electromagnetic calorimeter	43
3.2.5	Hadron calorimeter	45
3.2.6	Superconductive solenoid	48
3.2.7	Muon system	50
3.2.8	Trigger	52
4.	Global event reconstruction	55
4.1	Particle-flow algorithm	56
4.1.1	Prerequisites for particle-flow	56
4.1.2	Iterative tracking	57
4.1.3	Calorimeter calibration	58
4.1.4	Link algorithm and particle classification	59
4.2	Vertex reconstruction and charged hadron subtraction	62
4.3	Jet clustering	63
4.4	Jet energy corrections	64
4.4.1	Pileup offset correction	65
4.4.2	Simulated response corrections in η and p_T	66
4.4.3	Residual corrections in η and p_T	67
4.4.4	Jet energy scale uncertainties	68
5.	Jet particology	71
5.1	Jet energy composition	72
5.1.1	Composition in simulation	72
5.1.2	Composition with respect to jet p_T	75
5.1.3	Composition with respect to jet η	79
5.1.4	Composition with respect to pileup	81
5.2	Sensitivity of jet response to single pion response	84
5.2.1	ECAL+HCAL response variation	85
5.2.2	ECAL scale variation	87
5.2.3	HCAL scale variation	88
5.3	Charged hadron subtraction	90
6.	Dijet resonance search	95
6.1	Previous searches	96
6.2	Analysis strategy	97
6.2.1	Low-mass calo-scouting analysis	99

6.3	Event selection	100
6.3.1	Jet selection	102
6.3.2	Wide-jet clustering	103
6.4	Triggering	104
6.4.1	Efficiency calculation	104
6.4.2	High-mass analysis	105
6.4.3	Calo-scouting analysis	106
6.5	Background fit	106
6.6	Measured dijet mass spectra	109
6.6.1	Results with 2.4 fb^{-1} of integrated luminosity	109
6.6.2	Results with 12.9 fb^{-1} of integrated luminosity	109
6.6.3	Results with 36 fb^{-1} of integrated luminosity	109
6.7	Signal models	116
6.8	Exclusion limits	116
6.8.1	Limit calculation	117
6.8.2	Limits for qq, qg and gg resonances	118
6.8.3	Dark matter interpretation	123
7.	Summary	125
7.1	Jet particology	125
7.2	Dijet resonance search	126
7.3	Future of particle physics	127
	References	131

List of acronyms and symbols

η	Pseudorapidity ($\eta = -\ln \tan \frac{\theta}{2}$)
ϕ	Azimuthal angle in the $x - y$ plane
fb^{-1}	Inverse femtobarn
$\cancel{E}_T$	Missing transverse energy
θ	Polar angle in the $y - z$ plane
$\vec{\cancel{E}}_T$	Missing transverse momentum
p_T	Transverse momentum
x	Horizontal coordinate transverse to the beam axis
y	Vertical coordinate transverse to the beam axis
z	Coordinate along the beam axis
ATLAS	A Toroidal LHC Apparatus
BSM	Beyond Standard Model
CERN	European Organization for Nuclear Research
CHS	Charged hadron subtraction
CMS	Compact Muon Solenoid
ECAL	Electromagnetic calorimeter
EM	Electromagnetic
eV	Electronvolt
FSR	Final state radiation
GeV	Gigaelectronvolt
HCAL	Hadron calorimeter
HLT	High-level trigger
ID	Identification
JEC	Jet energy correction
JES	Jet energy scale
L1	Level-1 trigger
LEP	Large Electron-Positron Collider
LHC	Large Hadron Collider

MC	Monte Carlo simulation method
MET	Missing transverse energy
PF	Particle-flow
PS	Proton Synchrotron
QCD	Quantum Chromodynamics
SM	Standard Model
SPR	Single pion response
SPS	Super Proton Synchrotron
TeV	Teraelectronvolt

Author's contributions

This monograph thesis is based on results which are presented in the six publications that are listed below, to all of which the author has significant personal contributions that are summarized under each title. Publications **I** – **IV** are published in peer-reviewed high-energy physics journals and publications **V** and **VI** are published in the CERN Document Server [70] after an internal review of the CMS Collaboration.

The jet particology studies presented in Chapter 5 can be also found from the publications **I**, **II** and **VI** and are for the most part performed by the author. The main results of the dijet resonance search presented in Chapter 6 are published in **III**, **IV** and **V**. The author acted as a lead analyzer in the dijet working group during the analysis for these publications, and thus has large and vital contributions to each analysis.

I. Particle-flow reconstruction and global event description with the CMS detector.

The CMS Collaboration. JINST 12 (2017) no.10, P10003.

For the particle-flow paper I studied jet energy composition with respect to jet energy, jet pseudorapidity and event pileup. I was also responsible for studying the pileup jet mitigation properties of the charged hadron subtraction algorithm.

II. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV.

The CMS Collaboration. JINST 12 (2017) no.02, P02014.

For the LHC Run 1 legacy jet energy correction paper I studied jet composition as a function of jet energy and direction and performed a simulation study for determining the sensitivity of jet reconstruction to calorimeter energy scale miscalibrations.

III. Search for narrow resonances decaying to dijets in proton-proton collisions at $\sqrt{s} = 13$ TeV.

The CMS Collaboration. Phys. Rev. Lett., 116(7):071801, 2016.

For the first 13 TeV dijet resonance search I acted as a lead analyzer and was responsible for the entire basic analysis chain from data file processing to producing the final dijet mass histogram. I also carried out luminosity calculations, studied the time stability of our results and took care of applying and updating jet energy corrections.

IV. Search for dijet resonances in proton-proton collisions at $\sqrt{s} = 13$ TeV and constraints on dark matter and other models.

The CMS Collaboration. Phys. Lett., B769:520–542, 2017.

For the second 13 TeV dijet resonance search I continued as a lead analyzer and my tasks were extended to studying trigger efficiencies and determining the analysis lower mass limit.

V. Searches for dijet resonances in pp collisions at $\sqrt{s} = 13$ TeV using data collected in 2016.

The CMS Collaboration. <http://cds.cern.ch/record/2256873>, *Paper for peer-reviewed publication in preparation*, 2017.

I continued as a lead analyzer for the search with the full 2016 with the same responsibilities as for the earlier 13 TeV analyses.

VI. Jet Particology – Studying the Structure of Jets with the CMS Particle Flow Algorithm.

Juska Pekkanen for the CMS Collaboration, <https://cds.cern.ch/record/2264517>, *Moriond EW Session 2017 Proceedings Contribution*, 2017.

I performed both of the detector-level studies that are presented in my Moriond 2017 conference proceedings contribution. These results are also published in publication **II**.

1. Introduction

Our understanding of the subatomic world and the behavior of indivisible elementary particles is formulated as a the mathematically rigorous and self-consistent theoretical framework called the Standard Model (SM) of particle physics. The triumphant success story of the SM culminated in July 2012 when two independent experiments measuring the proton-proton collisions of the CERN Large Hadron Collider (LHC) announced the discovery of the Higgs boson. The Higgs boson is the last missing piece of the SM and the Higgs mechanism explains how other SM particles acquire their masses. The Standard Model is now complete in the sense that all the particles that it predicts are discovered, but mother nature seems to have more to tell and the discovery of a perfectly SM-like Higgs boson opens more questions than it answers.

A compact introduction to the Standard Model is given in the beginning of Chapter 2. For understanding the theoretical concepts behind the protagonists of this thesis, *jets*, we examine more closely the theory of Quantum Chromodynamics (QCD) that governs the phenomena of the strong force and quarks and gluons. As we shall learn in Section 2.2, jets are the experimental footprint of quarks and gluons, which we collectively call *partons*. In the latter part of Chapter 2 we review the main reasons why we believe that the SM is not the ultimate theory of nature and also touch on some alternative models that are being tested later in this thesis.

Where we experimentalists focus on studying the signals in our detectors, we would not have much to measure without the superior performance of the LHC in delivering proton-proton collisions at world record energies. Thus, in Chapter 3, where the experimental setup is described, we first take a walk through the basic accelerator techniques and different machines that form the LHC and make our research possible. Later in Chapter 3 the Compact Muon Solenoid (CMS) experiment is introduced

with a brief description of the overall structure and the used coordinate system, before a more detailed look into the different detector technologies and into the all-important features of *triggering*.

Once we are familiar with the hardware of the experimental setup, in Chapter 4 we continue with an introduction to the software side and algorithms that are used for reconstructing the collision events from separate signals in some of the millions of channels in the CMS detector. Here we pay extra attention to the particle-flow (PF) algorithm and reconstruction and calibration of jets, as we need to learn the basics of the central techniques and challenges in PF jet measurements for understanding the studies and results of Chapters 5 and 6.

In Chapter 5 we introduce the research branch of *jet particology* that we define as the study of the composition and behavior of jets at the particle level. We study how the energy of jets is distributed to different particles, how the sensitivity of certain CMS detector elements affect jet measurements and also how jets originating from the parasitic simultaneous *pileup* collisions can be suppressed with the help of the particle-flow algorithm.

Jets are put into use in Chapter 6, where we search for new massive particles from proton-proton collision events containing back-to-back jets with high energies. These *exotic resonances* are predicted by theories that attempt to solve the problems of the SM. In the *dijet resonance search* we probe the subatomic world at unprecedented energies with dijet events with masses up to 8 *teraelectronvolts* (TeV). As we see no evidence of new physics signatures, we put stringent constraints on the production of various exotic particles predicted by beyond Standard Model (BSM) theories. We are able to set limits for production of dijet resonances at the energy interval from 0.5 up to 8 TeV.

In Chapter 7 the central findings and lessons of jet particology and dijet resonance search are summarized, followed by an outlook and propositions for additional studies. A few words are also spent on discussing the future of experimental particle physics.

2. Theory

In this chapter an introduction to the Standard Model is given with the focus on Quantum Chromodynamics and jet formation. This is followed by a review of some of the key issues of the SM and a brief introduction to a selection of new physics models that try to solve the problems in the current understanding.

2.1 Standard Model of particle physics

The Standard Model of particle physics is a theoretical framework that incorporates all the knowledge of the subatomic world that we have gathered since the dawn of science, with only a handful of anomalous results challenging it and keeping us physicists on our toes. The Standard Model has been extremely successful in predicting and describing nature since the early 1960s when the first parts of it were established. A critical mind could say, though, that this is hardly surprising as the SM has always been extended to include the latest experimentally verified theoretical developments. A concise introduction to the SM can be found from Reference [5].

In short the Standard Model combines under one umbrella model the two theoretical frameworks that describe three of the four known fundamental forces of nature: Quantum Chromodynamics (QCD) that describes the behavior of quarks and gluons and Electroweak theory that combines together electromagnetic (EM) and weak interactions. Electromagnetism tells us how electrically charged particles behave in electric and magnetic fields and weak interactions govern the phenomena of radioactivity. We call the forces arising from these interactions the color force or the strong force, the electromagnetic force and the weak force, respectively. The mediator or *gauge particles* of these forces are the gluon, the photon and the

$W^\pm$ and Z bosons respectively, each of which to the best of our knowledge is an indivisible point-like elementary particle.

The particle spectrum of the SM is presented in Figure 2.1. Leptons and quarks, which are all spin- $\frac{1}{2}$ fermions, are the matter particles of our universe and come in three *generations* or *families* in the SM, although the number of families is merely an experimental result and not predicted by the theory. The only stable matter particles are electrons and up and down quarks; all the heavier fermions decay to these lightest particles of the families. Neutrinos are stable and massless in the SM, although experiments irrefutably show that they have minuscule yet non-zero masses which to date are unknown [21].

The Higgs boson is a special particle in the SM. It has spin 0 and it is not a matter particle nor a force mediator, but its existence is crucial for explaining the origin of mass of other elementary particles, as described in more detail below.

All of these particles are discovered and re-discovered in various particle physics experiments, most recently the Higgs boson by the ATLAS and CMS collaborations in 2012. [8, 24]

The charges of the three fundamental interactions are called the color charge in QCD, the electric charge in EM and the weak isospin charge in weak interaction. Quarks carry all these charges *i.e.* they feel all the three forces. Electrons, muons and taus carry electric charge and weak isospin charge so they feel EM and weak force. Neutrinos carry only weak isospin charge and thus interact only via weak interaction.

The greatest shortcoming of the SM was that in its simple form it predicted the gauge bosons all to be massless, but the short range of the weak interaction requires massive mediator particles. This assumption got experimental proof when the $W^\pm$ and Z were discovered in 1983 and masses were measured to be close to $100 \text{ GeV}/c^2$ for the $W^\pm$ and Z bosons. A theoretical idea now known as the Brout-Englert-Higgs (BEH) mechanism, or the Higgs mechanism in short, that would fix the issue was proposed in 1964 by three independent groups [67, 77, 75]. The proposed mechanism postulated a new field - a Higgs field - that is present everywhere in the universe. Under suitable circumstances the field experiences a phenomenon called spontaneous symmetry breaking, which in turn causes the Higgs mechanism to give mass to elementary particles such as the $W^\pm$ and Z. Since the discovery of a new boson in 2012, which later proved consistent with the hypothesis of the SM Higgs boson, all the new results

support the idea that the Higgs mechanism is the origin of mass of otherwise massless elementary particles. The only exception are neutrinos, the tiny masses of which are not explained at all by the SM. [97]

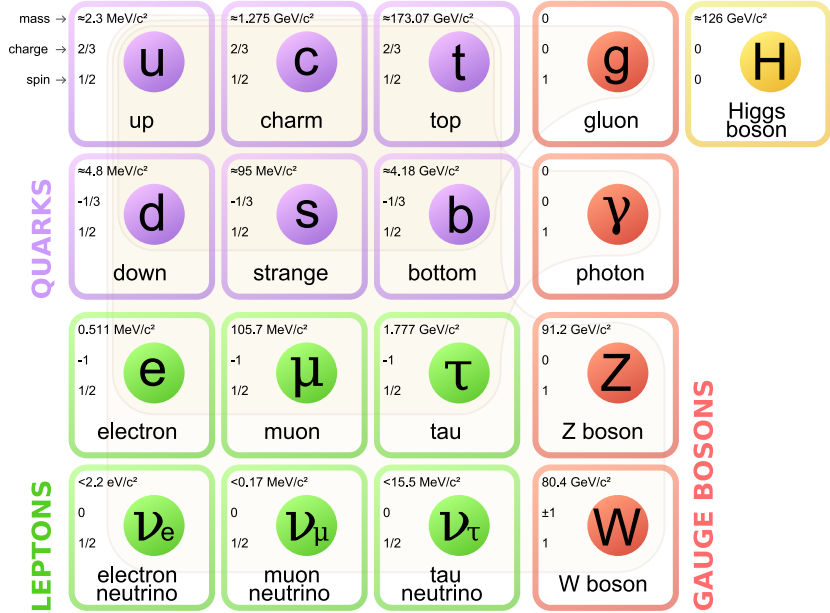


Figure 2.1. Particles of the Standard Model. [98]

2.2 Quantum Chromodynamics

In the realm of jet physics the color force is the dominant and most interesting of the SM interactions. The gauge group of the Standard Model is $SU(3) \times SU(2) \times U(1)$ and these three symmetry groups dictate the behavior of the color force, the weak force and the electromagnetic force, respectively. The unitary $U(1)$ group is an *Abelian* group where the generators of the group commute with each other. For this reason the force carrier of electromagnetism, the photon, does not itself feel the force that it mediates. An equivalent statement is that photon does not carry electric charge. This is why electric and magnetic fields weaken with distance. The special unitary $SU(3)$ group behind QCD (and $SU(2)$ behind weak interaction) on the contrary is a non-Abelian group where generators do not commute. Because of this the gluon carries color charge and feels the force that it mediates. This fact has drastic consequences for experimental particle physics. For understanding how jets are created in particle collisions we discuss three distinctive properties of QCD, all of which result from the

fact that gluons carry color charge: *color confinement*, *asymptotic freedom* and *hadronization*.

2.2.1 Color confinement

Quarks are believed to be confined so that they can only appear in "colorless" quark–anti-quark doublets or triplets of three quarks with each carrying different color charge. These configurations are called mesons and baryons, respectively, and are called colorless by the analogue that the anti-color of an anti-quark cancels the color of a quark and a mixture of the three main colors red, green and blue form a colorless state, although this is only a vague analogy and colors do not exist in the subatomic world. Although a rigorous mathematical proof showing that QCD is confining waits to be formulated, it is experimentally a well justified assumption; in experiments we only observe mesons and baryons. It is also shown that a non-Abelian gauge field theory such as the QCD *can* be confining. [84, 99]

Where electric and gravitational fields dilute with distance, the color force does not. When for example a quark in a meson is pulled away from the anti-quark, a *color flux tube* between the colored particles is formed and a constant strong force between the particles remains, until a process called *hadronization* takes place. In this process a quark–anti–quark pair materializes from the energy of the field between the two original partons before they get far apart from each other. The hadronization property of partons, which results from quark confinement, ensures that no quark or gluon can be observed without one or two other partons in the immediate vicinity of it.

In order to further clarify this fairly complicated but extremely important phenomenon for this thesis, I quote Sheldon Glashow, who attempts to explain the process in detail in his book [72]:

The strong force between two colored quarks... ...does not diminish with distance but remains constant, independent of the separation of the quarks, and an enormous amount of energy would be required to isolate a quark. ...the separation of a single quark by just an inch from the proton of which it is the constituent would require enough energy to raise the author from the earth by some 30 feet.

Long before such an energy level could be attained another process would inter-

vene. From the energy supplied in the effort to extract a single quark, a new quark and antiquark would materialize. The new quark would replace the one removed from the proton and would reconstitute that particle. The new antiquark would adhere to the dislodged quark, making a meson. Instead of isolating a colored quark, all that is accomplished is the creation of a colorless ('white') meson.

Also gluons are subject to confinement as they are colored and thus cannot be observed in freedom. According to QCD they can, however, form colorless combinations of gluons called glueballs and some experimental results hint at the possibility of the presence of a glueball, but definitive evidence for such *color singlets* has not been found yet. Distinguishing a glueball from a pure meson or a mixed glueball-meson state is experimentally challenging due to the similar experimental footprints.

2.2.2 Asymptotic freedom and parton shower

Where QCD is seen confining at 'long distances' when we try to separate colored partons from each other, at short distances or high energies the particles are understood to be in asymptotic freedom, meaning that they can move about freely and do not interact with the color force with other partons. This is due to *running* of the strong coupling constant α_s that is seen in Figure 2.2: when the energy scale of the process is high, or equivalently the distance scale is short, the coupling constant approaches zero.

When two hadrons – proton to proton at the LHC – are collided with high energies what actually collide are the quarks and gluons inside the hadrons. When a quark collides and gets momentum in a direction different from the mother hadron it is decelerated by the attractive force from its mother hadron and radiates gluons, just like an accelerated electron radiates photons in the process called *bremssstrahlung*. The emitted gluons can emit more gluons themselves or split into a $q\bar{q}$ pair. This called the parton shower continues until the boosted quark is far enough from the mother hadron so that α_s becomes strong and QCD acts confining again. This threshold is called the hadronization scale and needs to be fixed in the Monte Carlo simulations that are used for modeling parton showers computationally.

The mechanism behind the running of the strong coupling constant can

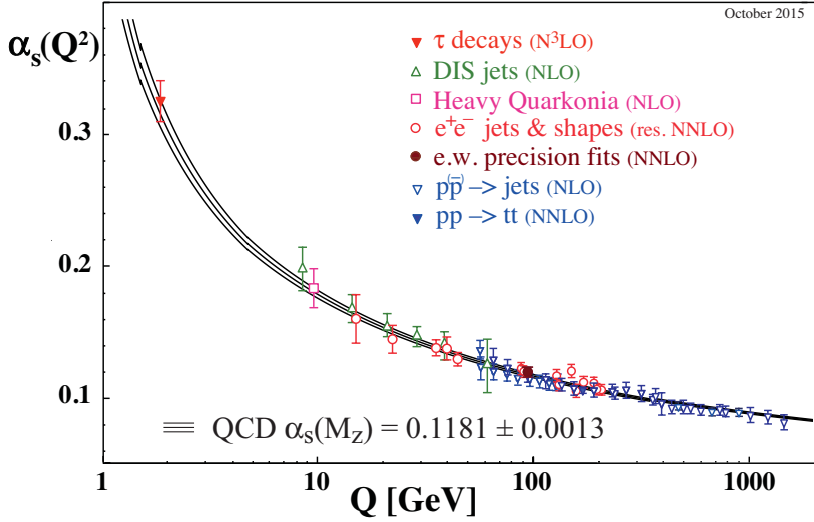


Figure 2.2. Strong coupling constant α_s in various processes as a function of Q , the momentum transfer in the process in question. The value of α_s is most commonly given at it's value at the Z-boson mass, 91.2 GeV/c². [84]

be explained with an effect called color charge antiscreening where the vacuum polarization caused by virtual gluons created in the vicinity of a colored particle effectively make the color field to strengthen with distance. This is opposite to the vacuum polarization and charge screening in electromagnetism, once again due to the color charge of gluons. It is by the virtue of asymptotic freedom that we can use perturbation theory in QCD calculations. Once the hadronization scale is reached, the assumptions of perturbative QCD do not hold anymore and we need to rely on phenomenological models. A particularly useful and widely researched non-perturbative method is *lattice QCD*, where continuous space is approximated by points in a grid, and QCD phenomenology can be studied by solving equations for each lattice point separately. In view of collider physics, the lattice calculations have the serious limitation that they can be mostly applied to static configurations only.

2.2.3 Hadronization

The hadronization process, that was already mentioned above, is the next process in jet formation after parton shower. As the quark (or gluon) gets further away from the mother hadron the strong force field persists and accumulates increasing energy in the color flux tube. Soon there is enough energy in the bond to create a quark–anti-quark pair from vacuum and

a $q\bar{q}$ pair is formed. The hadronization process repeats until there is not enough energy to form new mesons. Also the gluons radiated in the parton shower process hadronize once they are far enough for the confinement to take action.

After the parton shower and hadronization processes a jet with typically tens of hadrons and also some photons and leptons is formed and the only experimental footprint that we have access to are the stable or pseudo-stable final state particles.

A schematic of jet formation in a Monte Carlo simulation is presented in Figure 2.3 and an experimentalist's view is presented in Figure 2.4.

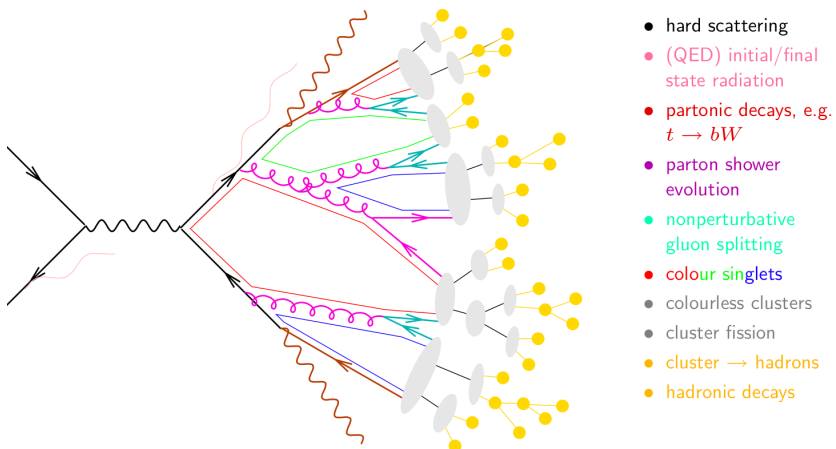


Figure 2.3. A schematic of jet formation in a Monte Carlo simulation in a $\gamma \rightarrow t\bar{t}$ process. [100]

2.3 Jet definition

As jets are somewhat chaotic collections of particles flying roughly to the same direction and it is generally not possible to know on a jet-by-jet level which particles actually originated from the hard parton collision that we are interested in, we need a definition for a jet. A jet definition comes in form of an algorithm that unambiguously decides which particles belong to a jet and how many jets there are in a collision event. As we do not have a complete understanding of the jet creation process, this does not necessarily reflect what actually happened in reality in parton level and different algorithms can give very different outcomes. What is important is that the algorithm is deterministic and it can be applied in different experiments so that we can compare the results of different studies and

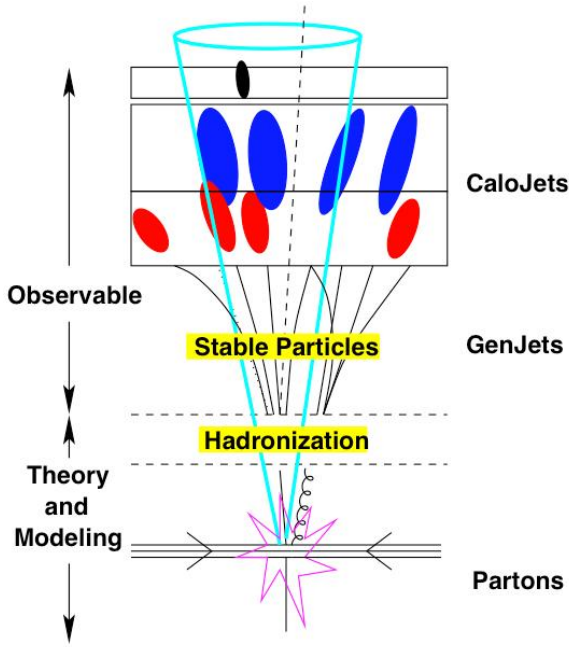


Figure 2.4. A schematic of jet formation from an experimentalist's point of view. [92]

trust that the jets are defined similarly. The so-called Les Houches accord states that a jet definition must include a complete description of the clustering algorithm together with all the used parameters, and a specification of the recombination scheme that defines how four-momenta are recombined during jet clustering. [18]

An important property for a jet finding algorithm is infrared and collinear (IRC) safety, which means that soft 'infrared' gluon radiation or collinear splitting of a parton does not alter result of the clustering process (see Figure 2.5). Non-IRC safe algorithms are strongly disfavored by theorists as they cause problems in perturbative QCD calculations.

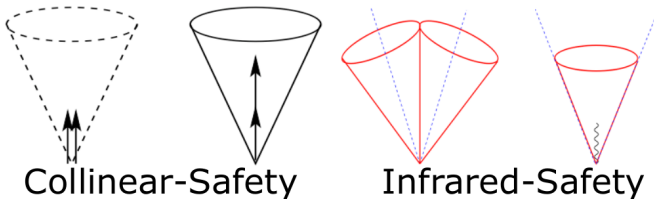


Figure 2.5. An illustration of infrared and collinear safety. The outcome of a good jet finding algorithm should not be altered by soft gluon radiation or collinear splitting. [92]

Numerous jet finding algorithms have been developed to address different challenges in different experimental and theoretical scenarios. The

first jet algorithms were cone algorithms which form jets from particles within a cone with radius R around a seed particle that is chosen with some criteria to be the jet centroid. Cone algorithms, however, are generally not IRC safe and include non-physical parameters. Due to the aforementioned limitations, cone jet finding algorithms were largely replaced by the next generation of IRC safe jet algorithms called sequential recombination algorithms which were first found problematic because of computational complexity that exceeded the capacity of computing infrastructure. This problem was significantly alleviated when an optimized way of performing the clustering was introduced in the FastJet package [20] and nowadays computing capacity is not an issue any more. Sequential clustering algorithms do not have a fixed cone size but merge particles into jets in momentum space iteratively, until certain stopping criterion is fulfilled. Also these algorithms have a radius parameter R that controls the radius of the jet and it acts in the distance parameter d_{ij} between particle i and jet axis j defined in Eq. 2.1–2.3. These equations define a general type of sequential recombination jet finding algorithm.

$$d_{ij} = \min(k_{ti}^a, k_{tj}^a) \frac{\Delta_{ij}^2}{R^2}, \quad (2.1)$$

$$\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2, \quad (2.2)$$

$$d_{iB} = k_{ti}^a. \quad (2.3)$$

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 exponent parameter a determines the behavior with respect to particle momenta and defines the type of algorithm. The Cambridge/Aachen algorithm uses $a = 0$, k_T algorithm uses $a = 2$ and anti- k_T uses $a = -2$.

For the clustering process we introduce distance d_{ij} between entities (which are either particles or unfinished jets called *pseudojets*) i and j , and distance d_{iB} between entity i and the beam (B). The clustering proceeds by finding the smallest of the distances, and if it is a d_{ij} , entities i and j are recombined, while if it is d_{iB} , entity i is called a jet and removed from the list of entities. After this the distances are recalculated and the procedure is repeated. [19] The process is repeated until either all the particles of an event are clustered into jets (inclusive clustering) or when a pre-defined number of jets are found from the event. At CMS we use inclusive clustering and define the number of jets in an event by applying a transverse momentum cut which is usually on the order of 15 GeV.

Currently the most widely used algorithm is the anti- k_T jet clustering algorithm [19] that became the standard clustering method in the LHC experiments soon after its introduction in 2008. Anti- k_T is found to be highly insensitive to particles originating from pileup vertices (collisions of other proton pairs within the same bunch crossing) or underlying event (collisions of other partons in the same pp collision). It also produces circular cone-shaped jets (see Figure 2.6) despite being a sequential recombination algorithm. The radius parameter R is usually chosen between $R = 0.4$ – 0.8 and in LHC Run 2 ATLAS and CMS collaborations have agreed to use $R = 0.4$ as the default distance parameter, so that results from jet measurements are directly comparable. In the results presented in this thesis we use distance parameter 0.5 for the 8 TeV LHC Run 1 results and 0.4 for 13 TeV LHC Run 2 results. [92, 7]

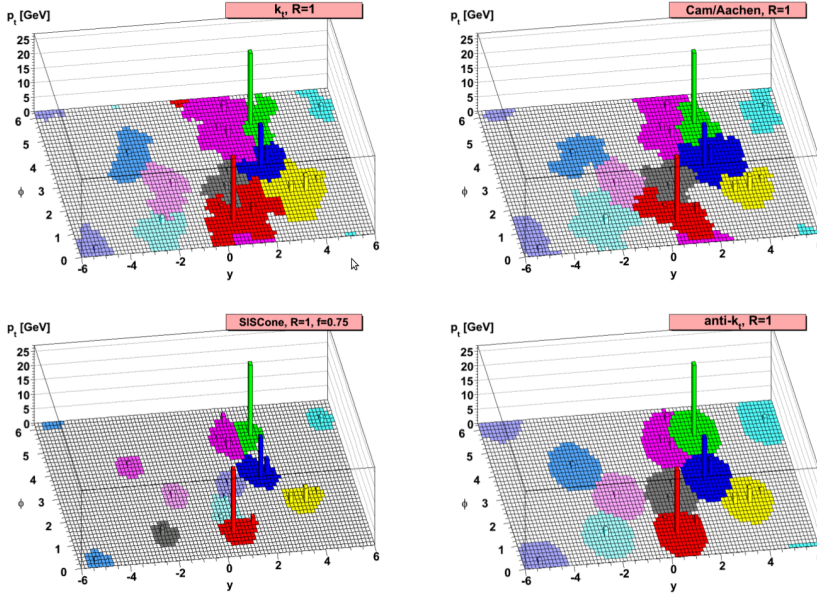


Figure 2.6. Jet shapes from different jet finding algorithms. Anti- k_T and SIScone produce circular jets where Cambridge/Aachen and k_t clustering results to irregular shapes. [19]

2.4 Physics beyond Standard Model

Although the Standard Model is arguably the most successful theory describing the nature ever made and an astonishing triumph of human endeavor, we have several unquestionable experimental results indicating that the SM is an incomplete description of the sub-atomic world. The

most obvious issues are neutrino masses, dark matter (DM), dominance of matter over anti-matter in the universe and the lack of quantum description of gravity. A short review of the problems with the SM is presented, followed by a selection of beyond standard model (BSM) ideas that propose to solve some of the problems and are relevant to the new physics searches with the dijet event topology discussed in this thesis. The basic idea of the new physics searches with dijet events is to reconstruct a mass spectrum from events with back-to-back jets, called dijets, and look for resonances, or bumps, from the smoothly falling spectrum that is predicted by the SM. The dijet resonance search strategy is described in detail in Chapter 6.

2.4.1 Shortcomings in Standard Model

In the following a non-exhaustive list of physical phenomena, which are not described by the SM, is presented together with a short introduction to each of the selected shortcomings.

Neutrino masses

Neutrinos are massless fermions in the SM, but several neutrino experiments show that neutrinos *oscillate*, *i.e.* change their flavor during flight from the source to an experimental apparatus. Historically the most significant result on the subject is the resolution to the solar neutrino problem where the low rate of electron neutrinos observed on Earth turned out to be due to electron neutrinos changing their flavor to muon neutrinos during the flight from Sun to Earth. According to the current theoretical understanding neutrinos can oscillate only if they have non-zero mass. [73]

Although for the other fermions masses are explained by the Higgs mechanism, this is problematic in the case of neutrinos. This is due to the fact that no neutrino with *right-handed chirality* has ever been seen and thus neutrinos are understood to exist only as left-handed, meaning that the spin and momentum vectors of neutrinos always point to opposite directions. For this reason, the SM considers the weak interaction to affect only left-handed particles. In the Higgs mechanism particles acquire mass by successive interactions with the Higgs field, each of which changes the chirality of the particle. For this a massive particle should be equally possible to be found with left or right-handed chirality, but as neutrinos are left-handed, this cannot be the explanation for their small

but non-zero mass.

Dark matter

Tantalizing cosmological evidence shows that the majority of mass in the galaxies is dark, *i.e.* it does not emit or reflect light. Perhaps the most concrete evidence is provided by studying the rotations of galaxies. The velocity of stars circulating around the galaxy center must be the slower the further away it is from the center; otherwise it would be thrown out from its orbit. Observing the rotation velocities of stars in galaxies shows that velocities cannot be explained by the amount of luminous matter that we see. This and a suite of other observations indicate that some 85% of matter is dark and non-baryonic. The Standard Model does not have any explanation for the origin and nature of dark matter. [63, 2]

Dark energy

The universe is measured to expand with an accelerating rate. Whatever phenomenon is behind this is dubbed to be *dark energy*. Possible explanations for this strong shortcoming in our understanding of the physics of the cosmic scale include modified gravity and a new force field driving the expansion, but the understanding of dark energy is still at an early stage and concrete and measurable theories explaining it are not many.

Naturalness, fine-tuning & hierarchy problem

While more an aesthetic disturbance than a direct experimental contradiction, the question of *naturalness* is considered to be one of the most problematic issues in the Standard Model. The assumption behind the naturalness hypothesis is that the constants of nature should have values of the same order of magnitude and thus the dimensionless ratio of constants of nature would be $\mathcal{O}(1)$. This is not true in our current understanding of the universe, as for example the strong interaction is some 10^{40} times stronger than gravity. One particularly disturbing aspect is the *hierarchy problem* or Higgs boson *fine-tuning* problem. When the newly observed boson could have any mass value between zero and the Planck mass $m_p = 1.2 \times 10^{19} \text{ GeV}/c^2$, by all odds the quantum mechanical loop corrections associated with other elementary particles would lead to a much higher mass than the observed $125 \text{ GeV}/c^2$. It thus appears that the effects of the loop corrections are fine-tuned to roughly cancel each other so that the Higgs has a mass so close to the other fundamental particles. Many BSM models introduce mechanisms that solve the hierarchy

problem and restore naturalness.

Gravity

Albert Einstein's theory of general relativity (GR) explains all the cosmological phenomena that mankind have observed with as high an accuracy as we have been able to measure them. However, in GR theory gravity is not described as a renormalizable quantum field theory like the other three fundamental forces of nature but more a recipe on how the curvature of spacetime will affect mass and vice versa. For this it is fundamentally incompatible with the electroweak and strong interactions which leads to problems when one would like to describe what happens in circumstances where both gravitational and quantum mechanical effects are important, for example in black holes and the big bang.

It is possible that also gravity is actually a quantum field theoretical force like the three others, the hypothesized messenger particle being the *graviton*. Gravitons are being sought for in the collision experiments but indications of their existence have not been found.

Matter–anti-matter asymmetry

The only mechanism that creates matter that we know of is a process where the energy of a particle is converted to mass in *pair production* process by obeying the universal exchange rate $E = mc^2$. Let us consider the most simple example where a photon turns into an electron-positron pair. This is the inverse phenomenon of a perhaps more familiar process, the annihilation of an electron-positron pair. In pair production an equal amount of matter and anti-matter is always produced. Still in our visible universe we see only matter and the only place where anti-matter is detected is our collision experiments. Thus the origin of the drastic matter–anti-matter asymmetry in the universe is a great mystery to which the SM does not have any explanation.

So-called *Sakharov conditions* are requirements that must be fulfilled in order for the Universe to have the observed matter–anti-matter imbalance. One of the conditions states that C- and CP-symmetries must be violated in the Universe. C-symmetry means that a physical system stays unaltered if charges of elementary particles are changed to opposite values, and CP-symmetry means invariance when both charge and parity of particles are changed simultaneously. [91]

2.4.2 String theories

Where the SM assumes fundamental particles to be pointlike, string theories predict that all the particles and thus all the matter in the universe consist of vibrating one-dimensional strings. The size of the strings is thought to be so small that for the current experimental technology and for all other practical purposes they appear pointlike. String theories predict the existence of graviton and thus provide an alternative for a *theory of everything*, i.e. a theory that explains all the four known fundamental interaction within one mathematical framework.

The greatest issue of string theories is that the models are not mathematically consistent without the introduction of a large number of extra dimensions, usually 6-7 on top of our 3+1-dimensional space-time. Extra dimensions have been sought for by ATLAS and CMS collaborations, but no evidence of their presence has been found. [79]

String resonances are being sought for in the dijet resonance search presented in this thesis. Strings would produce a very strong signal in the invariant dijet mass spectrum but so far no traces of excess production has been observed, and we have ruled out the existence of certain types of strings up to the energy of 8 TeV. [6, 64]

2.4.3 Extra dimensions

New physics models with extra dimensions explain the weakness of gravity and thus alleviate the hierarchy problem by assuming that we live in a 4-dimensional plane in a higher dimensional universe. According to the models we only see the 3+1 dimensions because electroweak and strong interactions are trapped in three dimensions, but gravity acts in the higher dimensional space and thus 'dilutes' and seems weak to us but is actually of comparable strength with the other three fundamental forces.

The only extra dimension model that predicts resonances in the dijet mass spectrum is the Randall-Sundrum (RS) model of extra dimensions that introduces gravitons in 5-dimensional space. In the 'RS1' model with a warped extra dimension, resonant production of dijet events originating from gravitons is predicted. Assuming certain model parameters we have ruled out these Randall-Sundrum gravitons up to the energy of 1.7 TeV. [88]

2.4.4 W' and Z' bosons

Many BSM theories predict the existence of new symmetries and fields which come together with particles that resemble W and/or Z bosons but are heavier in mass. These are normally referred to as W' ("W-prime") and Z' ("Z-prime"). Both W' and Z' particles can produce a resonance bump in the dijet mass spectrum. In the case of W' the jets would originate from a $q\bar{q}^{(\prime)}$ pair and in the case of Z' from a $q\bar{q}$ pair. Note that Z' cannot decay to gluons as gluons do not carry electroweak charge. As an anti-quark is required in the production of W' and Z' the production cross section in the LHC is small. We have set lower limits on the masses of heavy W and Z bosons to 3.4 and 2.7 TeV respectively. [66]

2.4.5 Compositeness models

What we at the moment consider as the indivisible elementary constituents of matter may well have internal structure. The fact the elementary particles of the SM are thought to be point-like and indivisible is only a corollary from the experimental results that agree best with the hypothesis of these particles lacking substructure and spatial spread. Another way to say this is that we have not yet reached the collision energies that would be needed to probe the length scales of the diameter or size of what we think are elementary particles.

Where different variants of compositeness models predict substructure to different SM elementary particles – composite Higgs being one popular hypothesis – in the dijet resonance search we are sensitive to the substructure of quarks. Just like the other composite objects of the subatomic realm – atoms, nuclei and hadrons – a composite quark can be pushed to an excited state, by an energetic gluon in this case, where the constituents are in non-minimal energy configuration. The release of this excitation produces a back-to-back dijet event with one jet coming from the gluon and the other from the quark as the partons fly to opposite directions in a decay-like process. Excited quarks are thus searched for in the final state with a qg pair and are ruled out up to a mass of 6 TeV. [12]

2.4.6 Dark matter

As discussed in Section 2.4.1 the dark matter mystery is one of the greatest shortcomings of our current understanding of nature and significant

efforts are made for finding a solution in various sectors of experimental physics. The hypothesis that dark matter consists of yet unknown elementary particles is a popular one and numerous experiments look for direct dark matter signals, but also at the LHC a selection of experimental techniques are being used for searching evidence of possible dark matter candidates.

Where the typical signal of dark matter particle is missing energy in events, in certain models a DM particle can interact with partons via a dark matter mediator particle. Such interaction would yield DM mediators decaying to quarks and gluons, which makes also dark matter accessible to the dijet resonance search. The assumption of having an exotic mediator particle additionally to the dark matter particle is motivated by the fact that no evidence of the direct interaction between DM and SM particles have been found by direct detection measurements. [81]

Dark matter mediators with certain properties are excluded up to a mass of 2 TeV, and further limits for mediator masses are set as a function of dark matter mass. [22, 68, 17]

3. Experimental setup

3.1 Large Hadron Collider

The Large Hadron Collider (see Fig. 3.1) is the world's biggest particle accelerator, the largest single machine in the world, the most complex experimental facility ever built, it produces collisions to the largest scientific collaborations in history and it provides the most energetic particle collisions ever achieved, surpassing the previous world record by a factor of seven. The LHC is truly a machine of superlatives. It is also home to one of the best scientific experiment ever made, the CMS experiment. Let us take a look at the central features and operating principles of the forerunner of sub-atomic research in order to understand how the proton-proton collisions that we strive to measure are actually even possible to produce.

3.1.1 Construction

The LHC occupies the 26.7 kilometer tunnel that was excavated in the mid-1980s for the Large Electron-Positron Collider (LEP) that used to fill the ring until its dismantling in 2001, when the construction of the LHC started. The LHC tunnel runs 45 to 140 meters underground in the stable bedrock under France and Switzerland, with some 90% of the ring being under French territory. The plane spanned by the ring is not perpendicular to the vertical direction defined by gravity, but is inclined 1.4% towards lake Geneva. The highest point of the LHC ring is close to the Jura massif. Both the depth and inclination of the tunnel are chosen for reaching stable geological environment for the excavations, for optimizing the cost of digging the service shafts and for placing the tunnel near the Super Proton Synchrotron (SPS) tunnel from where the particles are injected to the LHC and were previously injected to the LEP.

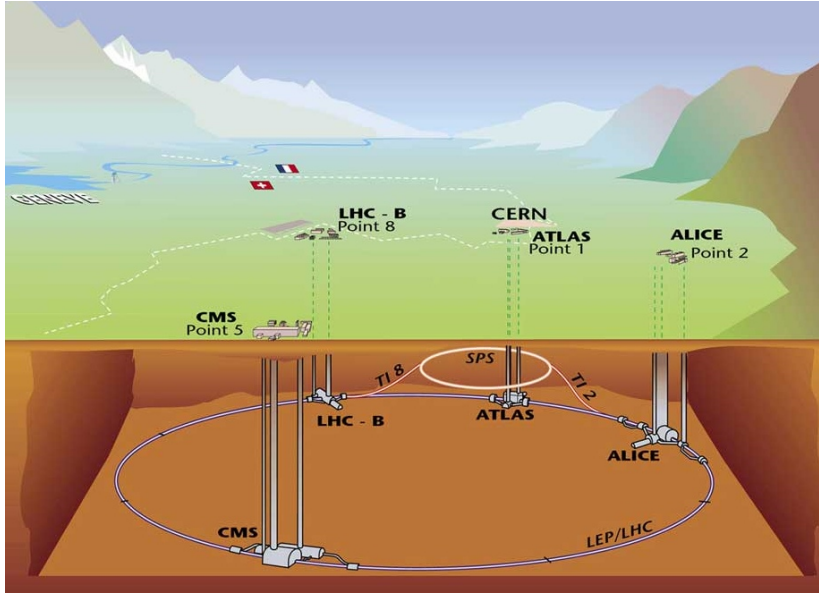


Figure 3.1. A sketch of the Large Hadron Collider and the four experimental sites under the Franco-Swiss border near Geneva. [70]

Although we normally speak about a circular machine, the LHC actually consists of eight arcs where the proton trajectories are bent and eight straight sections where the experiments and beam facilities are placed. As seen in Fig. 3.2 the LHC tunnel is divided into eight octants and in the center of each octant is an access point with a service shaft and elevator. The proton beams cross in four interaction points called Point 1, 2, 5 and 8 which house the four main LHC experiments called ATLAS [27], ALICE [26], CMS [23] and LHCb [62], respectively. The accelerating radio-frequency (RF) cavities are in Point 4 and the beam dump in Point 6. Points 3 and 7 are used for beam cleaning and quality control. Unlike at the LEP, where synchrotron radiation was a major obstacle and several accelerating facilities were needed for compensating for the losses, at the LHC the longitudinal acceleration for the protons is done in only one facility where proton energy is increased inside two cryomodules containing in total eight accelerator cavities for each beam. Synchrotron radiation is not an issue at the LHC because the magnitude of energy losses due to synchrotron radiation is inversely proportional to the fourth power of the particle mass ($\propto 1/M^4$). As protons are nearly two thousand times heavier than electrons, the radiation losses at the LHC are very small compared to the losses at the LEP.

The diameter of the LHC/LEP tunnel is 3.7 meters, which is enough for

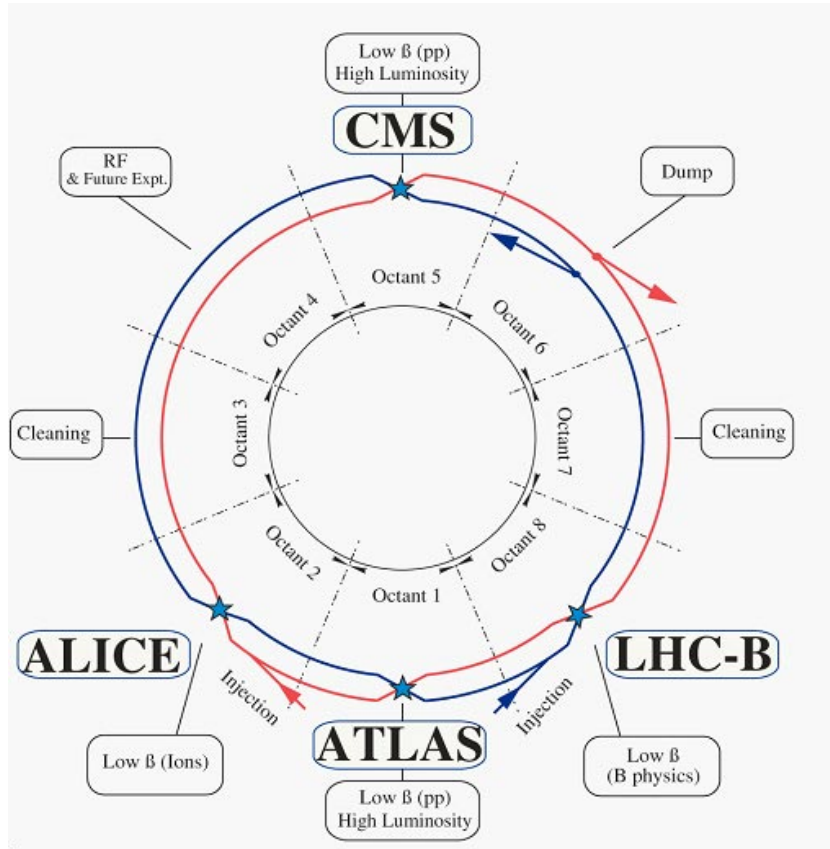


Figure 3.2. The Large Hadron Collider is divided into eight octants and in the center of each octant is either an experiment or beam operations facility. [69]

housing a dipole magnet assembly while leaving enough space for transporting other magnets to their places. The distance from the closest service shaft can be as long as 1.7 kilometers so the magnets had to be designed compact enough for two magnets to fit side-by-side. The restricted space available in the LEP tunnel was one of the main reasons why the magnets were designed so that the particle beams travel inside the same magnet assembly which houses two beam pipes; one for clockwise and one for counter-clockwise circulation direction. For achieving the desired effect of turning both beams towards the center of the ring the superconductive magnets needed to be designed so that the magnetic field points upwards in one beam pipe and downwards in the other. Implementing this field configuration to the very restricted space called for extremely careful design and placement of the superconducting niobium-titanium coils. For maintaining the superconducting state the coils are surrounded by iron *cold mass* that has heat exchange pipes where liquid, superfluid

helium flows at the temperature of 1.7 Kelvin. The LHC dipoles are capable of generating a magnetic field of more than 8 Tesla, which is enough for steering protons with the design energy of 7 TeV. The cross-section of a *cryodipole* module is shown in Fig. 3.3

Besides dipoles that have one magnetic north and one south pole, LHC utilizes also many other multipole magnets for beam quality operations. Most important of these are the four-pole quadrupole magnets, which are used for focusing the beams of protons that naturally repel each other causing the beams to de-focus. Having a well focused beam is important for achieving high collision rates at the interaction points.

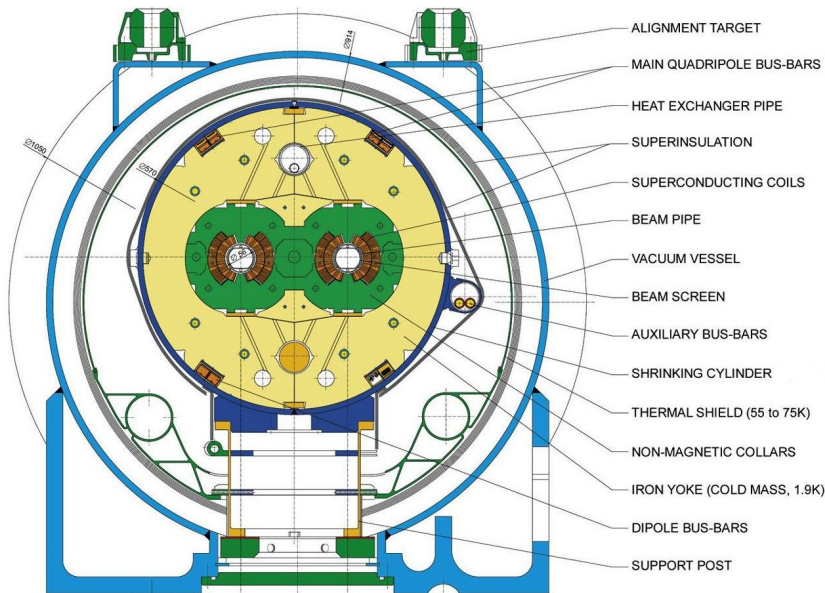


Figure 3.3. Cross-section of an LHC cryodipole, 1232 of which are used for keeping the particle beams in a circular track. [69]

3.1.2 Accelerator chain & LHC operation

Just like CERN does not consist only of the LHC, the LHC consists of many more accelerators than just the one giant ring. As seen in Fig. 3.4, the LHC accelerator complex has a total of seven accelerators and for proton-proton operation the big ring is the fifth in the chain. Two of the four pre-accelerator machines have previously been used for cutting-edge physics research but now after their collision energies are far surpassed by new accelerators, they are used to ramp up the proton energy in a step-by-step pre-acceleration process.

The journey of a proton towards a collision at the center of the CMS

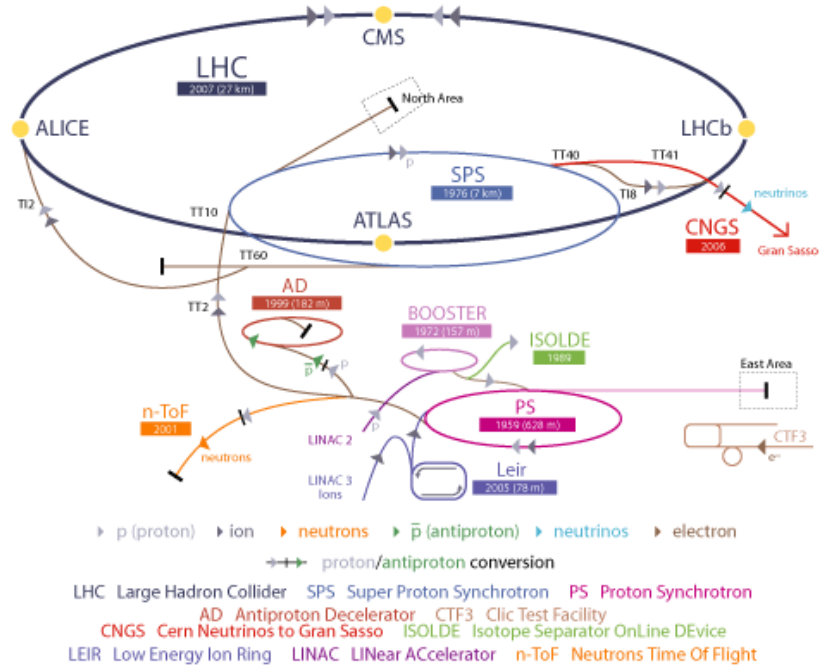


Figure 3.4. The CERN accelerator complex in 2009 when the LHC was first started. [69]

experiment starts from a hydrogen bottle at the beginning of the linear Linac 2-accelerator. The proton inside a hydrogen atom is freed from its electron in a machine called duoplasmatron after which, it now carrying a net charge, it can be accelerated with electric fields and steered with magnets. Linac 2 accelerates the proton to the energy of 50 MeV after which it is injected to Proton Synchrotron Booster that boosts the proton to 1.4 GeV. The next step from Booster is the Proton Synchrotron that ramps up our proton's energy to 25 GeV before injecting it to the 7 kilometer Super Proton Synchrotron (SPS). The SPS, which was formerly used for the discovery of W and Z bosons, ramps up the energy to 450 GeV, which finally is the injection energy of the LHC, where the proton energy is further boosted with a factor of 15 to reach 6.5 TeV before steering it onto a collision course at the four interaction points.

Our proton is not alone but is accompanied by some 10^{11} protons in the same bunch and the LHC can fit in up to 2808 of these bunches to both beams, totalling to $\mathcal{O}(10^{15})$ protons circulating in the machine simultaneously, before the number starts to slowly decrease in collisions. As each proton has the energy of 6.5 TeV, the total energy of the beams is of the order of 1 GJ, corresponding to an asteroid of 1000 kg with the velocity of 5000 km/h or a ten tonne truck speeding at 500 km/h.

The bunch crossing frequency of the LHC is 24.95 nanoseconds, in which a proton with velocity very close to the speed of light covers the distance of 7.48 meters. LHC thus has $\frac{26659 \text{ m}}{7.48 \text{ m}} = 3564$ bunch places or *buckets*. Due to the limitations arising from the technicalities of injecting bunches to and from SPS, not all of the buckets can be filled with a bunch, as some space between *bunch trains* is required so that the *kicker magnets* can push proton bunches from the SPS-LHC transfer line to the big ring. When taking into account the space that is needed for injection and some extra empty buckets that the LHC needs for beam quality operations, the maximum number of bunches is the aforementioned 2808 per beam. For bunches traveling close to the speed of light, this yields a bunch crossing frequency of up to 40 MHz. Thus, in total 756 buckets are left empty in a full LHC beam, as seen in Fig. 3.5 showing the exact bunch structure. The biggest contribution to empty buckets comes from the time needed for the beam dump kicker magnets to ramp up the magnetic field for a controlled beam dump. Bunches passing a kicker magnet during its rise time would result to uncontrolled steering of bunches to the machine structures, and potentially cause significant damage.

In 2016 the LHC was mainly operated with 2208 bunches due to a malfunctioning beam dump module in the SPS. As seen in Fig. 3.6, measuring the structure directly from data with the CMS detector reveals a structure of 46 *trains* with 48 bunches in each. The nominal LHC train has 72 bunches.

For a proton-proton physics run it takes roughly 1-2 hours to complete the injection of the LHC and to ramp up the energy from the injection energy of 450 GeV to 6.5 TeV. Once the desired energy is reached and beams are considered stable, the magnets around the four interaction points start focusing the proton beams so that collisions can start to take place. The beams start then slowly to *burn* as millions of protons are lost every second in collisions. The beams are dumped to graphite targets at Point 6 once the beams are too sparse to produce satisfactory instantaneous luminosity. Instantaneous luminosity is defined as the number of particles passing through a unit of area in a unit of time, and we can calculate the number of proton-proton collisions per time unit by multiplying instantaneous luminosity and the total proton-proton interaction cross-section. One full LHC *fill* can last up to 1-2 days if any safety mechanism does not trigger a premature beam dump before that.

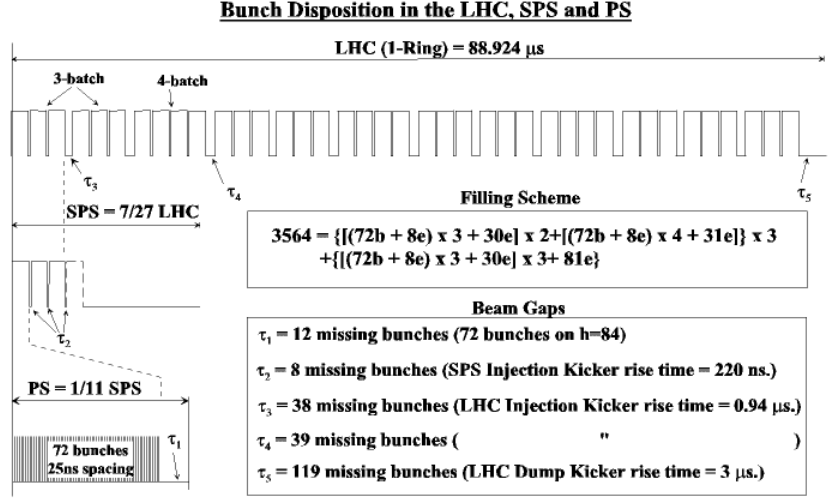


Figure 3.5. Bunch structure and the sources of empty bunch buckets of the LHC in a maximal 2808-bunch proton-proton fill. [69]

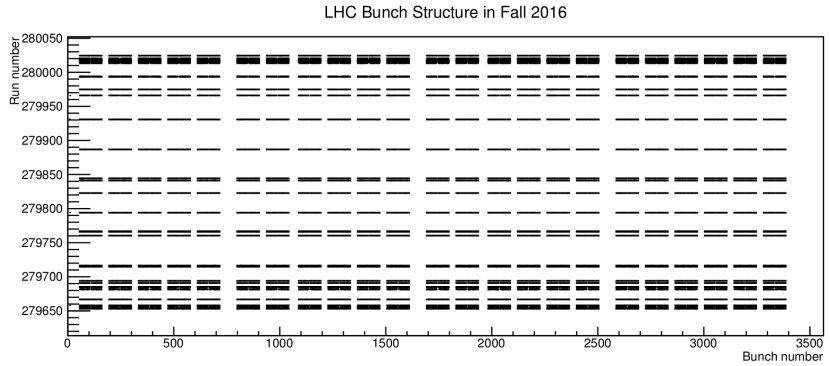


Figure 3.6. Bunch structure of the LHC in Fall 2016 with 2208 bunch operation directly measured from data. Bunches are grouped into four *batches*, each of which contains 10-12 *trains* of 48 bunches. [87]

3.2 Compact Muon Solenoid experiment

The Compact Muon Solenoid experiment [23] (Figure 3.7) is a multi-purpose particle detector at Point 5 of the LHC ring that is capable of measuring a wide range of phenomena, from Standard Model reactions to exotic signatures of new physics beyond the SM. With its cylindrical, nearly hermetic shape and four sub-detector layers, the CMS is designed so that only neutrinos can escape detection. All the other SM particles are either stopped by calorimeters or measured by their track, with the rare exception of a particle hitting for example a small crack in a detector layer or going unmeasured due to a malfunctioning detector element.

The word compact comes to the name from the fact that, despite its humbling size of 15×22 meters, the 14 000 tonne CMS detector is actually quite compact. This is due to the extremely dense build with almost all the space within the detector filled with either particle detection materials or necessary peripheral equipment such as cabling, read-out electronics and cooling systems. Muon made its way to the name for the large and precise muon detection system that constitutes the bulk of the experiment. The third characteristic of the CMS is the huge superconducting solenoid magnet that is capable of producing a magnetic field up to 4 Tesla over the inner detectors and thus enabling precise measurement of the momenta of charged particles.

Let us take a closer look first at the construction and structure of the CMS and then familiarize us with the triggering and data-acquisition techniques that enable successful data-taking despite the harsh data flow of up to 40 Terabytes per second.

3.2.1 Construction

The CMS site at Point 5 of the LHC is located next to the village of Cessy in the French countryside. Despite the original plan to place CMS to Point 1 close to the CERN main site in Meyrin, Switzerland, the locations of ATLAS and CMS were switched due to the better geological conditions in Meyrin for excavating the bigger experimental cavern for ATLAS.

Contrary to the usual method of assembling detectors directly to their final location the CMS was assembled and tested on an assembly hall on ground before lowering it 100m down the shaft to the cavern in slices each weighting thousands of tonnes. Assembling and testing the experiment above ground allowed the needed extra time for cavern excavations

CMS Detector

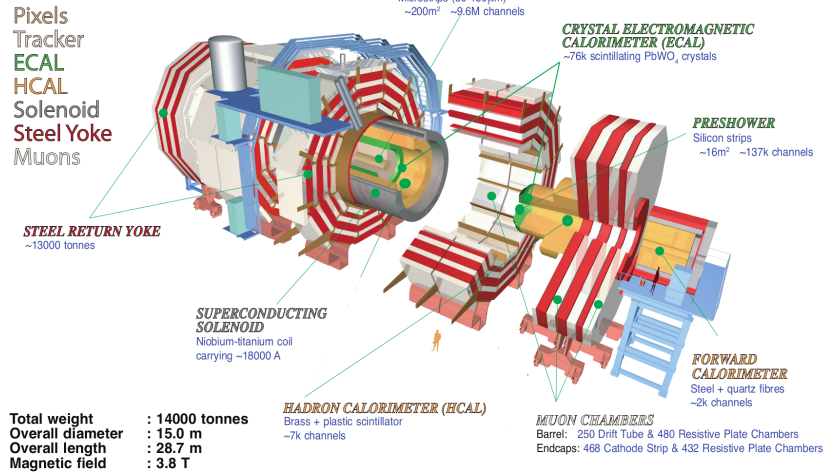


Figure 3.7. An exploded view of the CMS detector. [45]

that were hindered due to the discovery of an ancient Roman villa and unexpected amount of flowing water underneath the experimental site at Point 5. [71]

The cylindrical CMS detector consists of five consecutive layers, each serving a different purpose in the particle detection task. Starting from the interaction point where protons collide the first layer a particle faces is the tracking detector that measures trajectories of charged particles without affecting particles' momentum significantly. The next two layers are electromagnetic and hadronic calorimeters, ECAL and HCAL, which destructively measure energies of electromagnetically (ECAL) and strongly (HCAL) interacting particles. Calorimetry is surrounded by the giant superconducting solenoid magnet that creates a strong magnetic field that bends charged particles' tracks for momentum measurements with the tracker. The fifth and outermost layer of CMS is the muon detection system, which is interleaved with steel return yoke that supports the entire experiment and focuses the magnetic field lines to stay mainly inside the detector volume. In a nutshell the experiment is designed to work so that the tracker measures momenta of all the charged particles, the calorimeters stop and measure the energy of electrons, photons and hadrons, and the muon system measures muon momenta. Thus, the only known particles escaping detection are neutrinos, the total transverse momentum of which can be inferred as by using the law of momentum conservation in the plane transverse to the beams.

3.2.2 Coordinate system

CMS uses a coordinate system with the origin at the nominal interaction point at the center of the detector, the x -axis pointing towards the center of the LHC, the y -axis pointing upwards with respect to the LHC plane and z -axis is defined as the direction of the anti-clockwise proton beam. The polar angle θ is measured from the positive x -axis and the azimuthal angle ϕ is in the $x - y$ plane *i.e.* the plane perpendicular to the beam. In practice instead of the polar angle we use pseudorapidity η , which is defined with the polar angle as $\eta \equiv -\ln [\tan(\theta/2)]$. [23]

3.2.3 Tracker

The CMS tracking detector uses two different technologies both based on silicon as the active material producing electric signals from traversing charged particles. The inner part of the tracking detector is the pixel tracker that consists of roughly 66 million $100 \times 150 \mu\text{m}^2$ pixels surrounding the interaction point. The pixel tracker covers pseudorapidity up to $|\eta| = 2.5$ with three cylindrical layers in the barrel and two layers in both endcap regions, as illustrated in Fig. 3.8. The closest barrel layer is only 4.4 centimeters from the interaction point and the entire pixel detector is subject to high rate of charged particles with up to one billion proton interactions taking place every second so radiation hardness is a key design feature of the pixel detector. Still, the entire pixel detector was changed to a new one in the spring of 2017 to ensure smooth operation for the following years. The support structure where the pixel modules are installed is made in Finland.

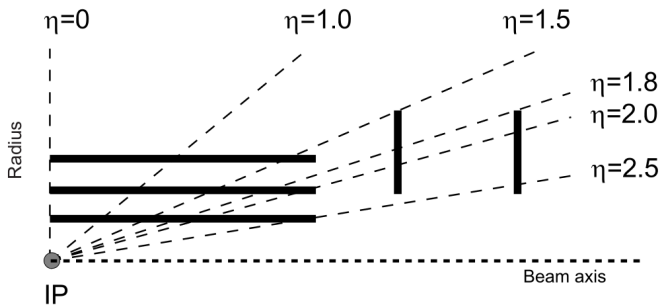


Figure 3.8. Layout of the CMS pixel tracker. The coverage spans the pseudorapidity region $|\eta| < 2.5$ with 2-3 pixel layers covering any straight particle trajectory. IP denotes the interaction point. [23]

The outer tracking detector consists of silicon strips, 9.3 million in total,

in a configuration of four strip layers in the *inner barrel* and three layers in *inner disks*, *i.e.* endcaps. The *tracker outer barrel* has six silicon strip layers and *tracker endcaps* nine layers each. The layer configuration is shown in detail in Fig. 3.9.

With this multilayer configuration and with the help of the 3.8 T axial magnetic field the track reconstruction performance of CMS reaches typically 0.7 % nominal momentum resolution for a 1 GeV/c track and track origin determination accuracy (or *impact parameter resolution*) of 10 μm for high-momentum tracks. The momentum resolution decreases with energy as the tracks get straighter and is around 5 % for a 1 TeV/c track. [39]

The high-performance tracking detector is at a central role enabling the use of the Particle Flow global event reconstruction algorithm in CMS where the track information is combined with calorimeter signals for optimal physics object reconstruction.

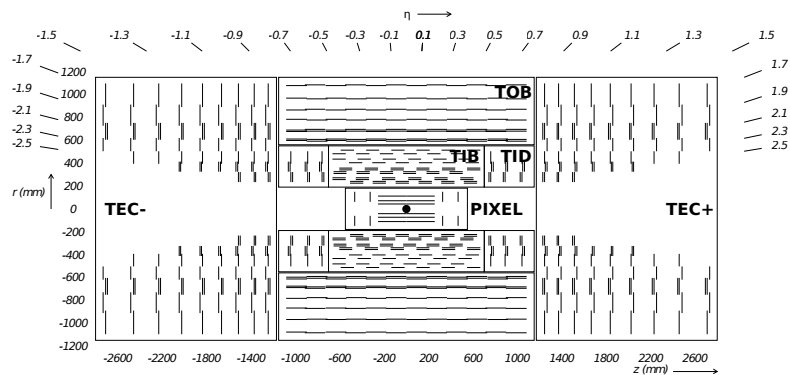


Figure 3.9. Layout of the whole CMS tracker detector. After the pixel layers the barrel region is covered by inner (TIB) and outer barrel (TOB) layers and endcaps by tracker inner disks (TID) and endcaps (TEC-, TEC+). [39]

3.2.4 Electromagnetic calorimeter

The electromagnetic calorimeter that surrounds the tracker is made of PbWO_4 *i.e.* lead tungstate crystals, as presented in Fig. 3.10. Lead tungstate is a special material that is optically as transparent as glass but is denser than steel with density of 8.3 g/cm^3 . Such a material is very suitable for a calorimeter as the same medium acts as absorber and scintillator. The working principle is that an incoming photon or electron initiates an electromagnetic cascade that results to scintillation light inside the lead

tungstate crystals. The created photons are then measured by photodetectors connected to the ends of each tile and the energy of the original particle is proportional to the light collected in a few neighboring crystals. A total of 75 000 crystals is used to achieve a nearly hermetic coverage.

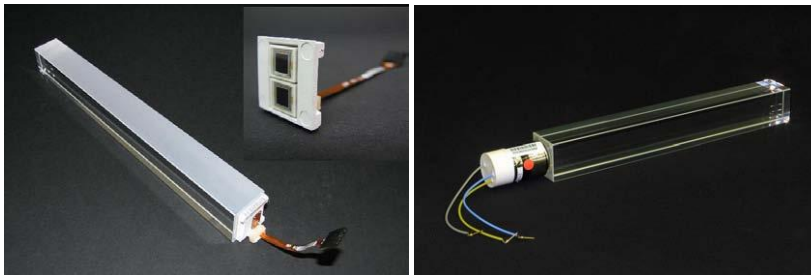


Figure 3.10. A barrel crystal with two avalanche photodiodes is shown on the left and an endcap crystal with a vacuum phototriode on the right. [23]

The barrel part of the ECAL covers $|\eta| < 1.5$ and consists of 60 000 23 cm long crystals, the length corresponding to 26 radiation lengths. Two avalanche photodiodes (APD) are connected to the end of each barrel crystal. The crystals have a tapered wedge-like shape which varies in pseudorapidity for having a gapless construction. The crystals are also adjusted to point slightly off the nominal interaction point for avoiding the situation where a particle travels exactly between two crystals.

The ECAL endcap covers $1.5 < |\eta| < 3.0$ with 22 cm long crystals corresponding to 25 radiation lengths. The 15 000 endcap crystals are all identical with one vacuum phototriode (VPT) photodetector glued to the end of each tile. The tiles are adjusted to point 13 cm off the interaction point to again prevent particles traveling between crystals.

As seen in Fig. 3.11, in the endcap region the ECAL crystal layer is preceded by a high-granularity preshower detector consisting of two layers of lead absorbers connected to silicon strip sensors. The ECAL preshower, covering $1.7 < |\eta| < 2.6$, is an important tool for distinguishing between the nearly collinear photon pairs from the decays of boosted π^0 mesons and the single high-energy photons from, for example, Higgs boson decays. The width of the preshower silicon strips is 2 mm, *i.e.* narrow compared to the 3 cm wide ECAL endcap crystals, so nearby photons can be distinguished much better than with the crystals alone. A preshower detector is not necessary in the barrel, where the photons coming from neutral pions are less likely to be nearly collinear. [23]

The relative energy resolution $\frac{\sigma}{E}$ of the ECAL barrel region was determined with electron test beams and can be parameterized as a function

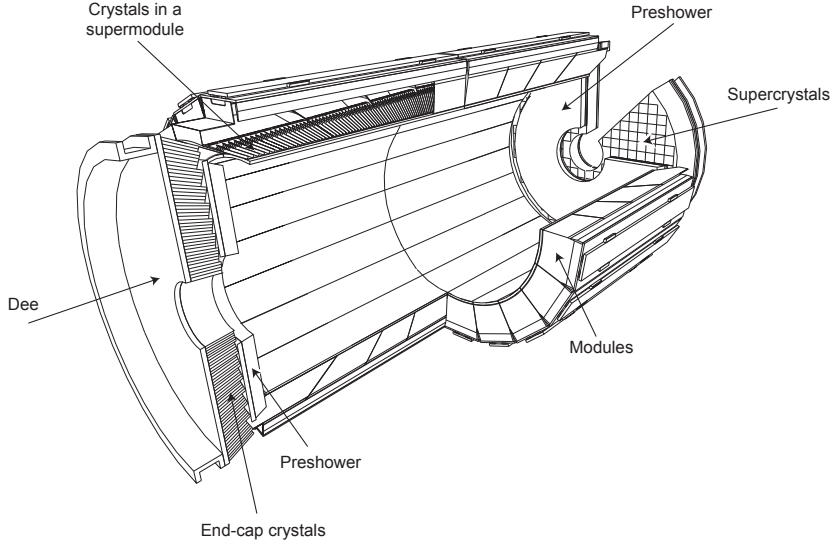


Figure 3.11. Layout of the electromagnetic calorimeter. The lead tungstate crystals are grouped into modules and supermodules and the endcap crystals are preceded by the preshower detector layers. [23]

of the electron energy E as shown in Eq. 3.1, where the first term on the right hand side of the equation is called the stochastic term, the second is the noise term and the last one is called the constant term. [38]

$$\frac{\sigma}{E} = \frac{2.8\%}{\sqrt{E [\text{GeV}]}} \oplus \frac{12\%}{E [\text{GeV}]} \oplus 0.3\% \quad (3.1)$$

The CMS ECAL is a remarkable detector. It is radiation hard and fast, and it is based on very special crystals that double as absorbers and scintillators. The calorimeter plays a crucial role in the accurate measurement of photon energies, which is of crucial importance in the search of the Higgs boson from the $H \rightarrow \gamma\gamma$ decay channel, which is experimentally one of the easiest decay modes.

3.2.5 Hadron calorimeter

Hadrons travel through matter significantly longer than electrons and photons because they are much heavier. Although a traversing charged pion can start a hadronic shower already in the ECAL, the majority of energy originating from hadrons is measured in the hadronic calorimeter (HCAL) surrounding the ECAL. The energy deposited in the ECAL is measured with an efficiency less than one due to the non-compensating nature of the CMS calorimeters, which means that the hadronic component of hadron showers is measured less efficiently than the well-calibrated electromagnetic component. This is the main source of non-unity jet re-

sponse, which has to be accounted for with jet energy corrections.

The thickness of 26 radiation lengths of the ECAL barrel crystals corresponds only to one nuclear interaction length, meaning that 38% of hadrons do not interact strongly in the ECAL and even if a nuclear interaction takes place the bulk of the initiated hadronic shower is stopped only in the HCAL. The CMS HCAL is a sampling calorimeter with brass (70% Co, 30% Zn) as the absorbing material interleaved with wavelength shifting Kuraray SCSN81 plastic as scintillator. The working principle of a sampling calorimeter is that the particles interact and start a shower in the absorber material (brass) and the energy in the shower is measured in the scintillator layer (plastic) that produces light that is collected and converted into electrical signals. The HCAL is divided into hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters as depicted in Fig. 3.12.

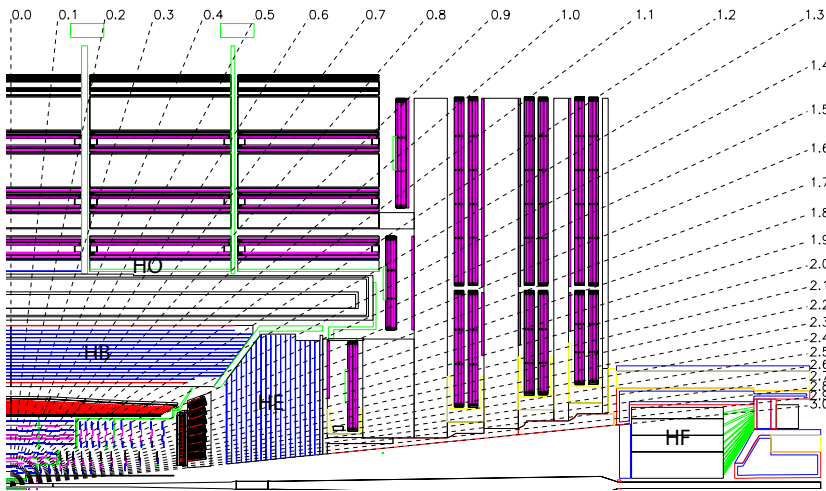


Figure 3.12. Cross sectional view of one fourth of the CMS detector with the HCAL barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters highlighted. [23]

The HCAL barrel covers $|\eta| < 1.3$ and consists of 1 meter thick brass wedges with 12 hollow slots for the plastic scintillators. At right angle from the beam axis the thickness corresponds to 6 nuclear interaction lengths. As the particle path length through the HB increases with pseudorapidity, at $\eta = 1.3$ the thickness is already 11 nuclear interaction lengths.

The HCAL endcaps cover $1.3 < |\eta| < 3$ and consist of roughly 1.5 meter thick brass modules that provide 9 nuclear interaction lengths for stopping the hadronic showers. The scintillators are made of the same wave-



Figure 3.13. A prototype of a hadron endcap brass absorber wedge in the CERN Microcosm exhibition. [86]

length shifting plastic that is used in HB. As can be seen in Fig. 3.12, the HE is constructed so that it fits tightly to the barrel calorimeter in order to avoid any parts of showers going undetected. A prototype of a HE wedge is shown in Fig. 3.13. The pseudorapidity coverage is further extended with the hadron forward detector covering $3 < |\eta| < 5$. In this forward region the particle flux is fierce and thus the HF is constructed with different materials, radiation hardness and longevity being the most important requirements.

As the stopping power for hadrons in the barrel $|\eta| < 1.3$ varies between 7 and 11 nuclear interaction lengths, after including the contribution of ECAL crystals, a up to one per mil of hadrons travel through both the ECAL and the HB. In order to have signals also for these *punch throughs* an extra layer of absorber iron and plastic scintillators are installed to the outside of the solenoid magnet. The hadron outer calorimeter, or *tail catcher*, efficiently recovers the energy of the punch through particles that would otherwise be only partially measured. [23]

The HCAL energy resolution $\frac{\sigma}{E_\pi}$ was measured in the barrel region with a pion test beam as a function of the pion energy E_π and is shown in Eq. 3.2, where the first term on the right hand side of the equation is the stochastic term and the second one the constant term. [61]

$$\frac{\sigma}{E_\pi} = \frac{110\%}{\sqrt{E_\pi [\text{GeV}]}} \oplus 9\% \quad (3.2)$$

The resolution of hadron measurement shown in Eq. 3.2 is significantly

worse than the resolution of the electromagnetic measurement of the ECAL shown in Eq. 3.1. The main reasons for the worse resolution for pions compared to that of electrons are the fact the CMS HCAL is non-compensating, *i.e.* the hadronic component of the showers is measured with a smaller efficiency than the electromagnetic component, and that the HCAL is relatively thin in order to fit inside the solenoid magnet, causing occasional energy leakage for high energy pions. The depth segmentation information of the HCAL is also not yet available for use in the energy reconstruction.

3.2.6 Superconductive solenoid

One of the most distinctive features of the CMS detector is the large and powerful solenoid magnet, which is the most powerful magnet of its kind in the world. Where the calorimeters are placed outside the magnet in most cylindrical detector designs, the CMS solenoid is so big that the ECAL and HCAL fit inside, yielding more precise calorimetry as there is no significant amount of dead material between the interaction point and the calorimeters. An artistic view of the solenoid is shown in Fig. 3.14.

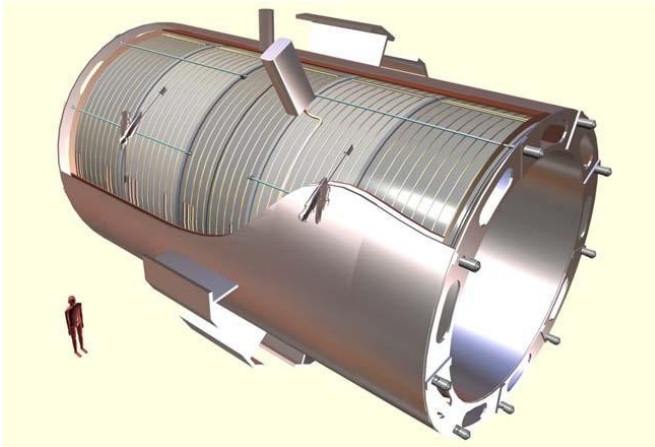


Figure 3.14. An artistic view of the CMS solenoid magnet. A standing human prototype is drawn to the left for a sense of scale. [23]

For bending trajectories of charged particles efficiently, the solenoid is designed to produce a homogenic magnetic field with flux density of 4 Tesla inside the solenoid and 2 Tesla in the parts of the detector outside the solenoid. For achieving this extremely strong field for a volume spanned by a 14 m long 6 m diameter cylinder, a current of 19 000 A is needed. For achieving such a high electric current a superconducting coil

is the only reasonable solution. The CMS solenoid is made of niobium-titanium superconductor filaments submerged inside a sturdy aluminum structure (see Fig. 3.15) that both supports the solenoid and transfers the heat out of the NbTi conductors with the help of liquid helium for keeping it at a temperature of less than 2 Kelvin for maintaining superconductivity. The assembly is also subject to large forces during operation, so a sturdy support structure is a necessity. For ensuring the longevity of the magnet with a minimal impact on the physics performance, the CMS solenoid is operated at 3.8 Tesla. [23]

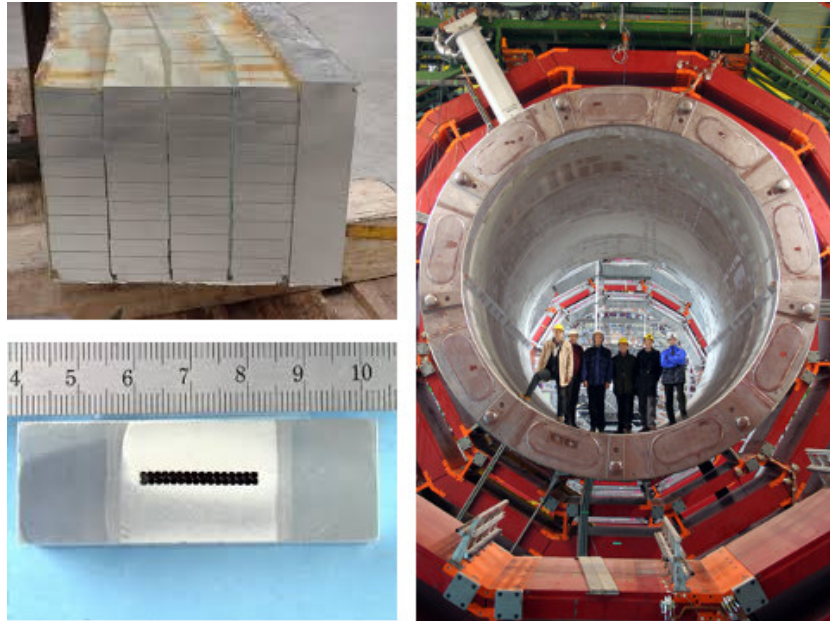


Figure 3.15. A prototype of the aluminum support structure (top left), and aluminum bar with the superconductive cable installed (bottom left) and people standing inside the solenoid during the assembly phase (right). [70]

The superconductive solenoid is the biggest solid part of the CMS detector and lowering it down from the surface assembly hall to the experimental cavern the 100 m underground with only centimeters of space between the solenoid and the concrete shaft was one of the biggest challenges of the assembly project. The solenoid brought more challenges to us in 2015 when an oil leak in the circulation of the liquid helium cooling system caused a lot of precautionary magnet ramp downs and reduced the data taking efficiency. After a substantial amount of inspection and massive efforts from the CMS and CERN technicians the cryogenics problems have been fixed and the operation in 2016 and 2017 has been smooth.

The superconducting cable consisting of 20 000 filaments as thin as a

human hair was entirely produced in the Finnish town of Pori. A total of 2000 km – or 18 tons – of such niobium-titanium cable was needed, which is also the distance between Pori and CERN as the crow flies.

3.2.7 Muon system

Outside the solenoid magnet all the known elementary particles are either decayed or stopped by the calorimetry, with the exception of neutrinos and muons. Neutrinos escape detection entirely and muons also fly out of the experimental cavern. Muons, however, can be measured for they carry electric charge and ionize the measuring medium. We are able to measure the momentum of muons precisely with the help of the tracker, the strong magnetic field both inside and outside of the solenoid and the colossal muon chambers and iron return yoke. Those make the bulk of the detector both in terms of mass and size, as is evident from Fig. 3.16 where the muon system and yoke make the entire outer solenoidal structure. Combining the information of the tracker and the signals in the muon chamber we can reconstruct the muon flight trajectory and calculate its momentum. When traversing the magnet the curvature of the muon flight path changes direction together with the magnetic field, producing S-shaped tracks that are also present in the CMS logo.

The muon system uses three different detector technologies called drift tubes (DT), resistive plate chambers (RPC) and cathode strip chambers (CSC), all of which are based on ionization of gas atoms by a passing charged particle. Drift tubes are simple detector units where a positively charged wire is spanned to the center of a tube filled with ionizable gas. When a charged particle knocks electrons off the gas atoms, they are likely to trigger an avalanche that drifts to the positively charged wire and yield an electric signal.

Resistive plate chambers have a sandwich-like structure with a volume of gas between anode and cathode plates. A traversing charged particle initiates an avalanche of electrons that drifts to the metallic strips attached to the positively charged anode plate. The RPCs have an extremely fast response of only 1 ns and they are used and installed mainly for triggering events with high-momentum muons.

Cathode strip chambers consist of interleaved negatively charged cathode strip layers and positively charged anode wire layers at right angle with respect to each other. The cathodes and anodes are inside a gas volume, and a traversing charged particle causes an electrical signal on

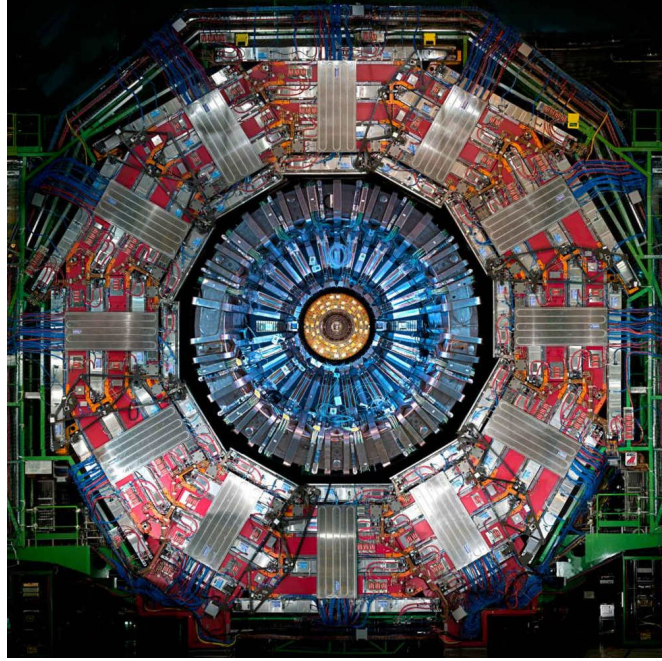


Figure 3.16. Cross section of the inside of the CMS detector with endcap layers removed. The innermost dark part is the tracker surrounded by the gold-yellow electromagnetic and electric blue hadron calorimeter. The silver ring is the solenoid magnet surrounded by the red and silver muon system interleaved with the iron return yoke. [70]

nearby strips and wires. The design is fast, radiation hard and provides good spatial resolution. The CSCs also perform well in non-uniform magnetic field such as in the CMS endcap region, which is one of the key reasons why this technology is used in the endcaps.

As shown in Fig. 3.17, the barrel region $|\eta| < 1.2$ consists of four layers of DTs and six of RPCs. Resistive plate chambers are also present in the endcaps up to $|\eta| < 1.6$ and the whole endcap region $1.2 < |\eta| < 2.4$ is covered with three layers of CSCs. [23]

The colossal muon detection system paired up with the strong magnetic field is a key design feature of CMS. The Higgs boson is known to decay to muons and other leptons via Z-bosons in the $H \rightarrow ZZ^* \rightarrow \mu\mu ll$ process, where Z^* denotes an off-shell Z and l refers to e , μ , or τ . Thus, a powerful muon detection system is advantageous in the search of the origin of the electroweak symmetry breaking.

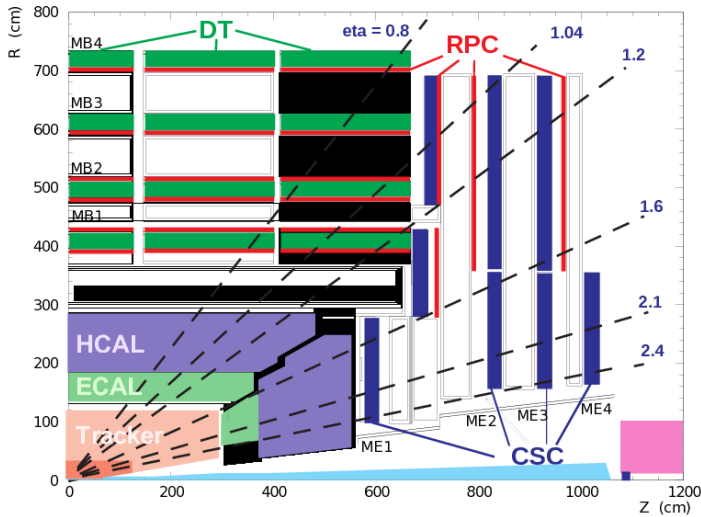


Figure 3.17. One fourth of the longitudinal cross section of CMS highlighting the muon system structure. Drift tubes (DT) are utilized in the barrel and cathode strip chambers (CSC) in the endcaps. Resistive plate chambers are used both in barrel and endcaps. [48]

3.2.8 Trigger

With a 25 nanosecond bunch spacing, the LHC is capable of producing up to 40 million bunch crossings and more than 1 billion proton-proton interactions per second. In practice, the average bunch crossing frequency is around 30 MHz and depends on the bunch filling scheme, but this does not change the fact that it is effectively impossible to record the detector signals from every bunch crossing, as this would require recording rate of tens of terabytes per second. Fortunately only a tiny fraction of the collision events contain features that we physicists are interested in – in the vast majority of bunch crossings protons only slightly scatter from each other and no hard parton-parton collision is present. It is the task of *triggers* to carefully select the events that have the best potential to produce new ground breaking observations. The CMS data acquisition system is capable of recording roughly 1000 events per second, so the trigger system has to reduce the rate of events with at least a factor of 30 000, *i.e.* less than one event in 30 000 can be accepted.

Triggers are not physical detector components but logical rules that use information from the detector for deciding if an event should be recorded or not. The trigger rules or *paths* are executed in real time during data taking either in field-programmable gate arrays (FPGA) connected directly to detector modules or in a computing farm right next to the ex-

perimental cavern for minimizing time used for data transfer. CMS uses a two-level trigger system where the Level-1 (L1) trigger makes extremely fast decisions based on a very coarse picture of the collision event in hand and forwards the potentially interesting events to the high-level trigger (HLT) for a more complete but still approximate fast event reconstruction for a final decision on whether the event is worth recording or not. As track reconstruction is too time-consuming and resource intensive operation, only calorimeters and muon systems are used in triggering at the moment. Whenever L1 or HLT decides to reject an event the information about it is lost for good. [23]

Level-1 trigger

The maximal design throughput frequency of the L1 trigger system is 100 kHz, corresponding to a filtering factor of 300. Due to the limitations of temporary signal storage in buffers, the L1 trigger has only $3\ \mu\text{s}$ of time for the decision, out of which a significant fraction is needed for transferring the electrical signals. Buffers are memory units that can store signals of multiple events for giving extra time to the triggers. Thus, the L1 is implemented entirely with FPGAs and custom electronics that can process the simple L1 trigger algorithms very fast. The L1 trigger algorithms are fairly simple and use only the immediately available information of an event. A L1 algorithm can for example check whether there are signals in several muon stations, indicating the presence of high-energy muons, or if there is a lot of energy in some calorimeter clusters hinting at hard jets, and thus making the event worth forwarding to the HLT for a closer inspection. [23]

High-level trigger

The high-level trigger makes a streamlined version of the full CMS offline event reconstruction, and much more elaborate trigger algorithms can be given to the HLT than to the L1. The HLT does a full event reconstruction with calorimetry, and an on-request regional particle-flow reconstruction with fewer steps of iterative tracking than used offline. In the dijet search presented in Chapter 6, the calo-scouting analysis uses calorimeter-only HLT trigger, while the high-mass analysis uses triggers with regional PF reconstruction.

The HLT performs the reconstruction and triggering process using a computer farm with about one thousand commercial processors. This second step reduces the event rate further by a factor of 100, and roughly

one thousand events per second are accepted for full reconstruction and permanent storage. [23]

Pre-scaling and data scouting

Some interesting yet frequent physics processes happen so often that recording every triggered event would result in too high event rates for the data acquisition system. This is why some triggers are *pre-scaled*, meaning that only a part of the triggered events are actually recorded. Pre-scaled triggers are very often used in the kind of analyses discussed in this thesis, as jets are present in almost every LHC collision and thus jet triggers have very high rates. Triggers with low-energy thresholds typically have higher pre-scale factors as nature favors processes with lower energy – a trigger that is set to fire for every event containing a jet with transverse momentum of 50 GeV or more can have a pre-scale factor of 100 000, so only ten such events out of a million are actually recorded. For reconstructing distributions that depend on jet transverse momentum, the histograms must be multiplied by the used pre-scale factors for restoring physical behavior.

For certain HLT triggers that are normally heavily pre-scaled, the compact event information of the streamlined event reconstruction is recorded in order to be able to have access to full statistics also for analyses normally subject to heavy pre-scaling. This technique is called *data scouting* and it is utilized in the calo-scouting dijet analysis presented in this thesis. This allows for measuring jet spectra at lower momentum regions than usually achievable, with the added complication of having to calibrate HLT calorimeter jets to offline PF jets. The energy resolution of HLT calorimeter jets is worse compared to offline PF jets, but sufficient for the dijet resonance search. [46]

4. Global event reconstruction

Once an event is triggered and recorded, it ends up on disk or tape storage in raw event format, which is a disordered collection of millions of signals from thousands of sensors from all around the detector. It is the demanding task of reconstruction algorithms to combine the signals of each bunch crossing into events containing only signatures of traversing particles, so that analysis groups can rely on the reconstruction, leave the complications of the detector aside and concentrate on the physics with a ready picture of the details of the event conveniently at hand. Figure 4.1 shows how the signals from different detector layers can be connected for finding and categorizing the different particles present in the collision events.

Although in the big picture finding jets from events is part of the complete event reconstruction procedure, I will dedicate a separate section for jet reconstruction and calibration as they are central topics in the research discussed in this thesis.

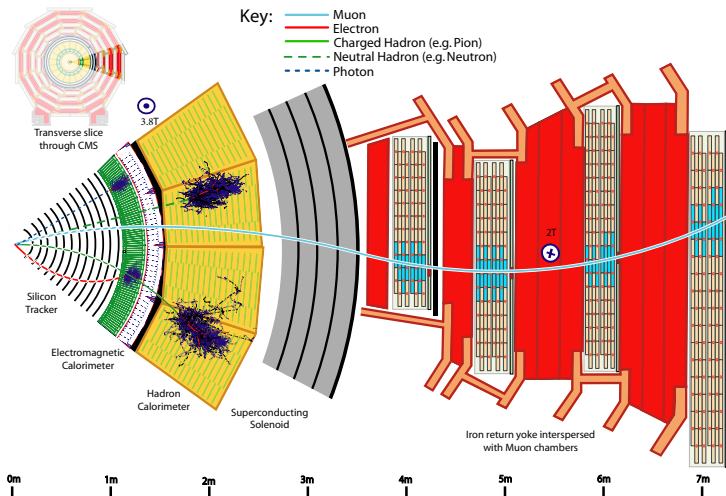


Figure 4.1. A slice of the CMS experiment illustrating the different detector layers and the signals that are used in the event reconstruction process. [58]

4.1 Particle-flow algorithm

In most of the particle physics experiments the collision events are reconstructed roughly by first finding photons, electrons and perhaps muons with the help of tracker and electromagnetic calorimeter, and then by looking for energy in the hadron calorimeters and associating deposits to jets. The particle-flow (PF) event reconstruction approach used at CMS is quite different. The PF algorithm uses all the information in the different detector layers for reconstructing individual particles. As shown in Figure 4.1, PF can find and classify electrons, muons, photons and charged and neutral hadrons from the events. The ambitious objective, illustrated in Figure 4.2, is to convert the complicated image of energy deposits in the detector back to how the event actually took place in nature. The global particle-flow event reconstruction is used for all the collision event recorded by CMS, and it is not used only for jet reconstruction, but for all the detected physics objects in the events.

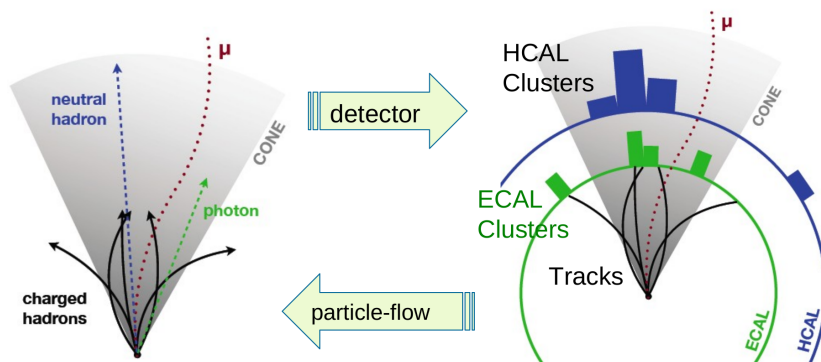


Figure 4.2. The CMS particle-flow algorithm combines sub-detector information and reconstructs individual particles in collision events for converting detector signals back to physical objects. Image: Francesco Pandolfi

4.1.1 Prerequisites for particle-flow

The PF paradigm was first developed in the ALEPH experiment at LEP [25], where electrons were collided with positrons. The approach worked well in the reconstruction of the much more simple collisions and it was not clear if the same method could ever be applied in hadron colliders where there is a much larger number of particles in the final state. The PF was first tested cautiously at CMS alongside traditional event reconstruction approaches, but with time it proved out that PF provides superior reconstruction performance compared to more conventional reconstruction

methods.

For the global event reconstruction approach to be successful, a number of features must be present in the experimental setup. In CMS the key features making PF successful are a highly segmented tracker, a high-granularity ECAL, a HCAL with good spatial coverage, a strong magnetic field separating charged particles and a very powerful muon spectrometer. [58]

One of the main reasons why PF is adopted as the reconstruction scheme in CMS is the fact that the energy measurement of hadrons using CMS calorimeters yields substantial uncertainties, and by exploiting tracking information the energy resolution of charged hadrons can be efficiently restored. After the particle-flow has proven to perform extremely well in CMS, also ATLAS collaboration has started to use it in some of the analyses, although the superior liquid-argon hadron calorimeter of ATLAS initially did not pose such a big need for enhancing hadron response as in CMS. [37]

4.1.2 Iterative tracking

A downside of a high-performance tracking detector is that the numerous layers of silicon on the particles' flight path unavoidably affect the particles that are being measured. Roughly 10% of charged hadrons undergo first nuclear interaction already in the tracker before entering ECAL. When using conventional tracking algorithms which require a hit in almost all of the tracker layers, each of these charged hadrons would get reconstructed as neutral hadrons by the PF algorithm and would cause a significant amount of misreconstructed energy. For this and other more technical reasons CMS has adopted an iterative tracking approach, where a total of 10 iterations over the tracker signals are done, starting from easily reconstructible clear signals to successively more obscure combinations of hits in the pixel and/or strip layers. As after each of the iteration the tracker hits assigned to a charged track are deleted from further consideration, the combinatorics and thus computational complexity stays manageable.

As shown in Figure 4.3 where different tracking approaches are studied using multijet events, the iterative tracking method yields similar misreconstruction rate with a conventional combinatorial tracking, but with significantly better efficiency for tracks with transverse momentum between 1 and 50 GeV. For both approaches the efficiency goes down to

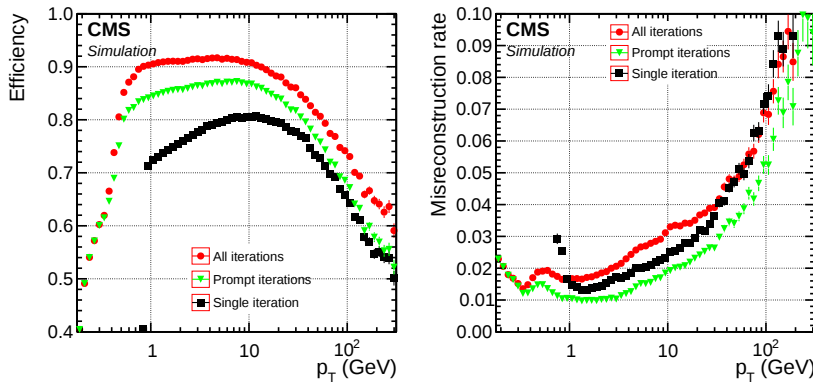


Figure 4.3. Comparison of conventional combinatorial iterative tracking (black squares) and iterative tracking with 6 (green triangles) and all 10 (red circles) iterations. Tracking efficiency is shown on the left and track misreconstruction rate on the right. Iterative tracking yields much higher efficiency with comparative iteration rate as the conventional tracking approach. [58]

less than 70% for tracks of 100 GeV and more, but at these energies the calorimeter measurement is already efficient and compensates for the decreasing tracking efficiency.

4.1.3 Calorimeter calibration

The electromagnetic showers initiated by photons and electrons in the ECAL spread out to more than one calorimeter crystal and the same is true for the hadronic showers caused by charged and neutral hadrons in the HCAL modules. Thus, the calibration of an energy deposit in calorimeters is started by clustering together nearby *calorimeter cells* (ECAL crystals or HCAL modules) starting from the calorimeter cell with highest energy and including cells with energy deposit above certain minimum threshold to the cluster. The clustering is done separately for ECAL and HCAL, and also separately for barrel regions and endcaps.

Although the single particle responses of the calorimeters are initially calibrated using test beams, radioactive sources and cosmic ray measurements, further calibration is necessary. In the case of ECAL the minimum energy thresholds in the clustering procedure unavoidably bias the energy measurement and for the HCAL, as an addition to the clustering threshold bias, the hadronic showers primarily start already in the ECAL and are not well calibrated with the pion test beam measurements where the ECAL was not present.

The residual calibration for the calorimeter clusters is done with simu-

lated photons for ECAL and neutral hadrons (K_L^0) for HCAL, heavily relying on the GEANT4 detector simulation which models the shower behavior inside the detector material. [90] Also, in the case of HCAL separate calibration coefficients have to be derived for hadrons that start showering already in ECAL and for showers that are contained entirely within HCAL, as the energy response of ECAL to hadrons is very different than for electrons and photons that it is calibrated to.

The details of this somewhat involved process are beyond the scope of this thesis and are well presented in Reference [58]. The central benchmark results are shown in Figure 4.4, where on the left the π^0 mass peak is reconstructed from the simulated and measured photons and shows agreement within 1%. In Figure 4.4 (right) the calorimeter response and resolution is shown before and after the calibration procedure. The calibrated relative response, defined as the mean relative difference of measured and true energy, is flat and compatible with zero for the entire energy range of 0 to 350 GeV and also the resolution is smaller (*i.e.* better) after the calibration, especially so for hadrons with energy below 50 GeV.

4.1.4 Link algorithm and particle classification

Once the calorimeter deposits are calibrated and charged particle tracks for both those inside the tracker and those containing hits also in the muon chambers are found by iterative tracking, the decisive step of the PF algorithm can be executed. The particle-flow link algorithm makes links between tracks and calorimeter deposits both in ECAL and HCAL. By analyzing signals in roughly the same direction in different detector layers, we can fine-tune calorimeter measurements with tracking information and *vice versa*, and end up exploiting all the information available for as accurate physics object reconstruction as possible.

After the link algorithm has grouped the collision event signals to *PF blocks*, each of which contain connected signals in one direction or neighborhood in the detector. From PF blocks the algorithm starts to classify particles by determining in which detector layers a block in investigation has signals, much like an educated human being can easily classify different particle types from Figure 4.1 representing the classification task at the CMS experiment, or Figure 4.5 illustrating the signals left by different particle types to different layers of a generalized particle collider experiment.

First come muons, whose experimental footprint are hits both in the

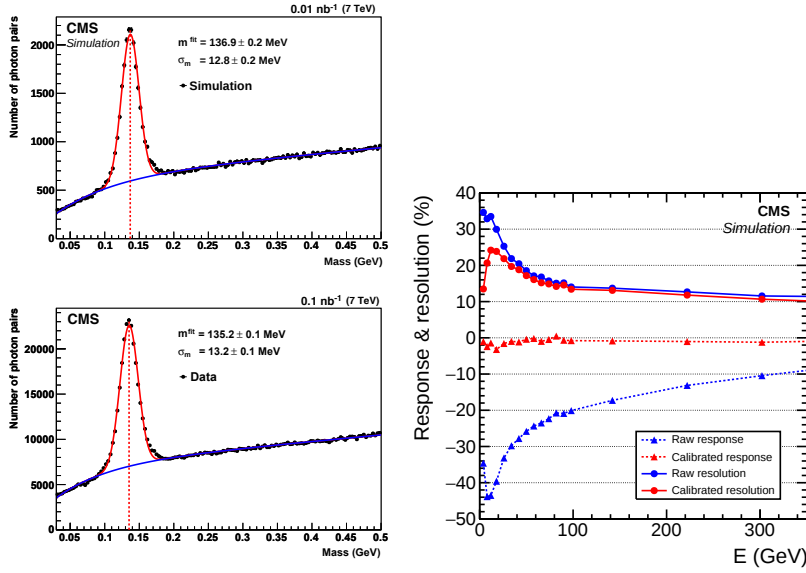


Figure 4.4. On the left the π^0 mass peak is reconstructed with the simulated photons used for ECAL calibration (top) and with actual neutral pion decays from a 0.01 nb^{-1} sample of 7 TeV collision data. The true π^0 mass being 135 MeV the simulation and data agree well with both each other and the true mass value, justifying the use of simulated photons in the ECAL calibration. On the right the response and resolution is shown before and after HCAL calibration process. The relative response is shown to be flat and compatible with zero, meaning that calibrated energy is close to true energy, and the energy resolution is shown to get better after the calibration, especially at the low energy regime. [58, 84]

tracker and in the muon chambers, but with no significant amount of energy in the calorimeters in between. The muon momentum and charge can be measured from a long S-shaped track that starts from the pixel and strip trackers and coincides with signals in the muon chambers at locations matching the extrapolation of the trajectory when taking the magnetic field into account. After the muon candidates in the event are found, the corresponding signals in the detector are removed from the PF block so that they do not cause complications or ambiguities in the next steps.

In the second step of particle classification we find electrons and isolated photons from the PF block. Electrons are identified from a charged track followed by an energy cluster in the ECAL but not in the HCAL. Also the bremsstrahlung photons that electrons emit in the strong magnetic field are used for identification. Isolated photons deposit nearly all of their energy in the ECAL, leave no signals in the tracker and little to no energy to the HCAL, and are thus identified by isolated ECAL clusters with at

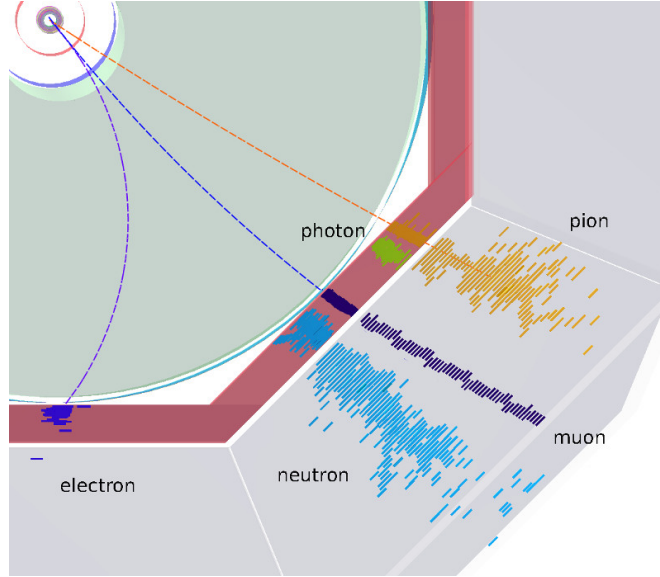


Figure 4.5. The particle classification task shown in a simplified model of a particle collider experiment. The outer two layers represent the calorimeters surrounding the tracking detectors. Pion is used as an example of a charged hadron and neutron as a neutral hadron. The energy deposited by the muon to the calorimeters seem to be somewhat exaggerated. [70]

least 10 GeV of energy. The signals that make up the PF electrons and isolated PF photons are masked from further consideration once this step is finished.

After muons, electrons and isolated photons, the PF algorithm looks for hadrons and non-isolated photons. Calorimeter deposits within tracker coverage, $|\eta| < 2.5$, which are not linked to any tracks, are reconstructed as non-isolated photons for ECAL clusters and as neutral hadrons for HCAL clusters. Where the true identity of the origin of the ECAL energy cannot be resolved, this choice is justified by simulation results indicating that 25% of the jet energy in hadronic jets is carried by photons, and neutral hadrons leave only 3% of the jet energy in the ECAL. Outside the tracker coverage charged and neutral hadrons cannot be distinguished, and linked ECAL+HCAL deposits are reconstructed as hadronic energy without specifying the charge and isolated ECAL deposits are reconstructed as electromagnetic energy.

At this point all the stable or pseudo-stable Standard Model particles are reconstructed and categorized, with the exception of neutrinos. The presence of neutrinos is indirectly inferred by exploiting the energy conservation in the transverse plane, and the vector sum of transverse momenta of neutrinos (or BSM particles escaping detection) is reconstructed

as missing transverse momentum, denoted by $\vec{E}_T$. The magnitude of $\vec{E}_T$ is called missing transverse energy, and it is denoted by E_T or called MET. Where the determination of the amount and direction of MET is of paramount importance for many CMS analyses, it is of limited significance for the topics discussed in this thesis.

A tangible demonstration of the result of the above classification procedure is shown in Ch. 5, where the energy composition of jets is examined at particle level. This study is possible only by the virtue of successful application of the particle-flow algorithm.

4.2 Vertex reconstruction and charged hadron subtraction

The objective of the LHC is to run with high instantaneous luminosities in order to produce as many high-energy collision as possible. This, however, leads also to a large number of simultaneous proton collision in each bunch crossing, *i.e.* pileup collisions, and we are generally interested only in the most energetic interaction. The pileup is a great challenge in the physics analyses, where the particles from vertices other than the most interesting one only disturb the study of the chosen interaction. Fortunately, the pileup contamination of the reconstructed collision events can be drastically reduced with advanced algorithms, such as the charged hadron subtraction (CHS) algorithm.

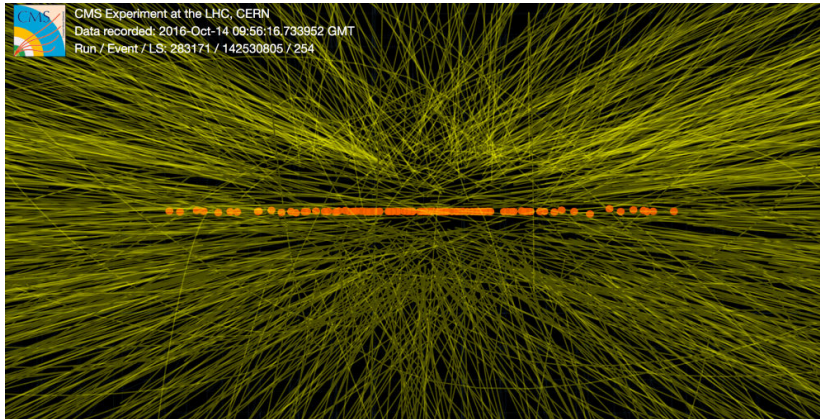


Figure 4.6. A real collision event from a 13 TeV high-pileup proton-proton run in 2016 with roughly 100 reconstructed primary vertices. [70]

During the event reconstruction all the simultaneous interaction vertices are located by looking for the origin of the charged tracks. This naturally requires extrapolation from the first layer of the pixel detector to

the beam line, which are 4 centimeters apart. As shown in Figure 4.6, the multiplicity of the tracks is high, and sophisticated algorithms are needed for solving the ambiguities and computing the most probable vertex configuration. The problem of tracking and vertexing is solved with excellent results in CMS as described in Reference [51], with roughly 10 micron position resolution for the locations of primary interaction vertices. The primary vertex that is chosen as the most interesting, typically the origin of the most energetic tracks, is called the *lead vertex* and all the other are pileup vertices.

As the particle-flow associates part of the tracks to charged hadrons, we can efficiently identify and remove from further consideration all the charged hadrons that originate from pileup vertices with the CHS algorithm. Roughly two thirds of the pileup contribution consists of charged hadrons so the reduction is significant. CHS is, however, limited to the angle $|\eta| < 2.5$ that is covered by the tracker, beyond which we cannot distinguish charged and neutral hadrons. The effectiveness of this method in subtracting pileup from jets is examined in Chapter 5.

4.3 Jet clustering

Jets that are being studied in this thesis are clustered with the anti- k_t algorithm that is introduced in Section 2.3. With the exception of the calo-scouting analysis described in Section 6.2.1, the jets are clustered from individual particles, or *PF candidates* to be exact, reconstructed with the particle-flow algorithm. This bottom-up jet reconstruction approach is what enables the in-depth jet particle substructure studies and pileup mitigation methods presented in Chapter 5. A PF-jet with the associated tracks and calorimeter deposits is shown in Figure 4.7. The traditional way of simply clustering calorimeter deposits to jets naturally does not give any handle for closer inspection of the jet particle substructure.

The anti- k_T distance parameter $R = 0.5$ is the default for the CMS analyses using the 7 and 8 TeV LHC Run 1 data and is used in the studies of Chapter 5. The default radius parameter was changed to $R = 0.4$ for the 13 TeV LHC Run2, so in the dijet resonance search presented in Chapter 6 the used jets are anti- k_T $R = 0.4$ PF- or Calo-jets, often referred to as AK4PF- and AK4Calo-jets in short, respectively. In this search the charged hadronic pileup component of the AK4PF-jets is subtracted by the CHS-algorithm. For optimizing the discovery potential of this anal-

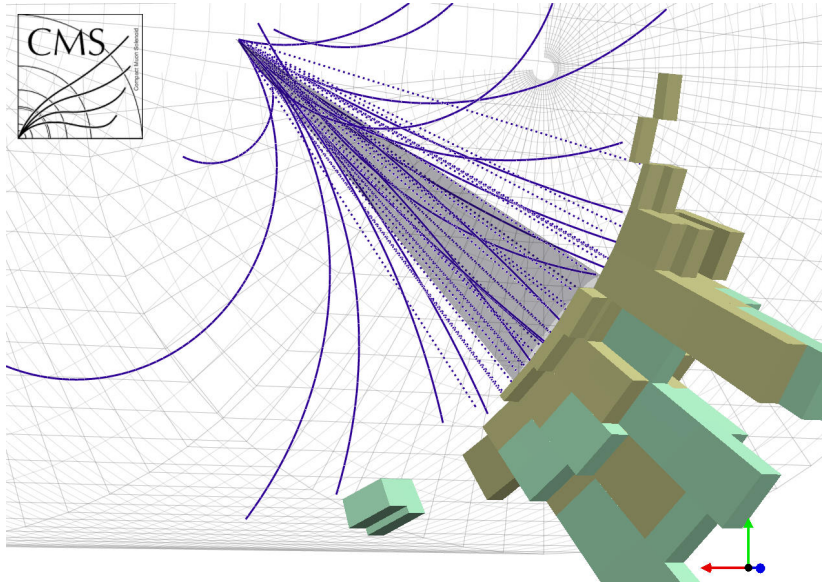


Figure 4.7. A jet reconstructed with the PF-algorithm with charged particles drawn with solid lines and neutral particles with dashed lines. The jet is marked by the dark cone and some particles bend out of the jet cone due to the strong magnetic field. [65]

ysis, the nearby jets are re-clustered into $R = 1.1$ Wide-jets, as will be elaborated in Chapter 6.

4.4 Jet energy corrections

Although we calibrate the calorimeters as well as possible using the cluster calibration introduced in Section 4.1.3, and use precise tracking information for calculating the momenta of the individual particles that end up inside a jet in the clustering process, a series of corrections still needs to be made for each jet for having as good a match as possible between the energies of the measured jet and the particle-level jet. In simulation, the true particle-level jets are defined and built by clustering all stable (decay length $c\tau > 1$ cm) particles, excluding neutrinos. For hadrons with $c\tau < 1$ cm, the decay products are clustered into the jets.

The CMS jet energy correction procedure consists of a sequence of corrections, each accounting for different sources of difference from the true particle-level jet energy. The four levels that are used by default for every measured jet are pileup offset correction, response correction in pseudo-rapidity, response correction in transverse momentum and residual correction in η and p_T , as depicted in Figure 4.8. Residual corrections are not

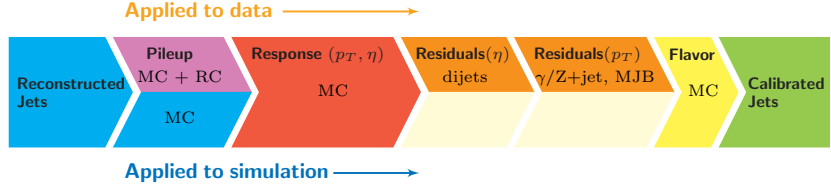


Figure 4.8. The jet energy corrections are applied as a sequence of calibration coefficients multiplied in succession to the jet four momentum. Some of the correction are applied only for measured jets and some only for simulated jets. RC refers to the random cone pileup offset calculation method and MJB is a short-hand for the multi-jet balance methods used for residual corrections alongside γ/Z + jets events. [57]

needed for simulated jets. The correction factors derived for each level are multiplied to jet four momentum vectors and vary depending on the jet energy and direction in η . Separate corrections need to be derived for jets clustered with different clustering distance parameter, for jets clustered after charged hadron subtraction, and for calo-jets that are clustered only based on calorimeter information. Let us briefly introduce the working principle and motivation behind each of the four correction steps, leaving the often convoluted details to Reference [57].

4.4.1 Pileup offset correction

The pileup offset correction is designed to subtract the energy contribution of particles from pileup vertices ending up inside the jet cone. The first approximation of the offset correction is calculated using simulated collision events with and without pileup contamination for determining the average amount of uniformly distributed offset energy as a function of number of pileup vertices N_{PV} . The correction is derived as a function of event offset energy density ρ and jet area A , jet centroid direction in η and jet transverse momentum p_T .

As relying entirely on simulations is an ill-advised approach, a residual offset correction as a function of η is derived using the random cone (RC) method, which calculates the pure pileup contribution on average as a function of N_{PV} using real events that are triggered randomly and do not contain hard scatterings, thus giving a reliable environment for estimating the pileup background. These kind of events are called *zero-bias* events in the often horrible particle physics jargon.

As charged hadron subtraction drastically reduces the pileup contamination in jets, separate correction factors are derived for jets with and without the CHS algorithm applied in the event reconstruction.

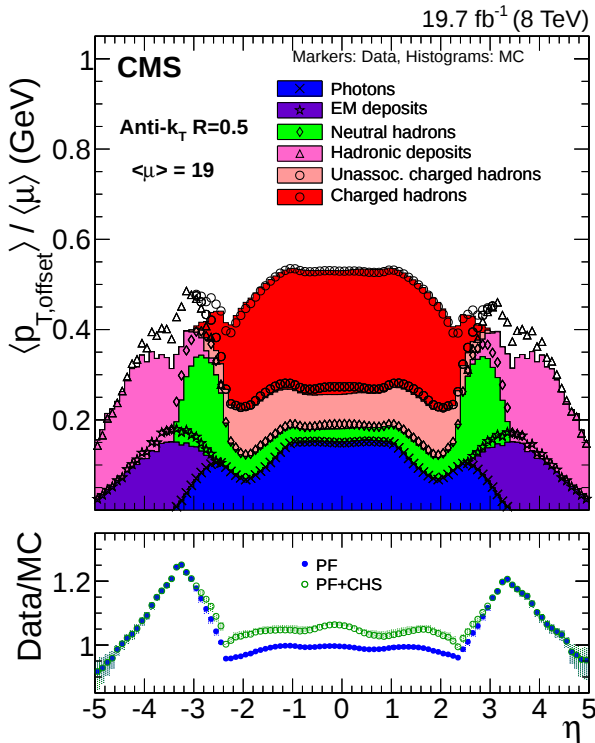


Figure 4.9. Offset energy contribution and particle composition of pileup as a function of η and normalized by average number of pileup μ in data (markers) and simulation (histogram) in anti- k_T $R = 0.5$ jets. The ratio of data and simulation, which is also the applied correction factor in the residual correction, is shown in the bottom panel for PF-jets (blue full circles) and PF+CHS-jets (green open circles). [57]

The amount and particle composition of offset pileup energy in jets as calculated by the random cone method is shown in Figure 4.9, where also the residual data-to-simulation correction is shown in the bottom panel.

4.4.2 Simulated response corrections in η and p_T

Once the pileup offset correction is made the jet particle response R_{ptcl} , defined in Equation 4.1, is studied using multi-jet events simulated by PYTHIA 6.4 tune Z2* and exploiting the highly detailed GEANT4 detector simulation that accurately models the particle showers inside different detector layers.

$$R_{\text{ptcl}}(\langle p_T \rangle, \eta) = \frac{\langle p_T \rangle}{\langle p_{T, \text{ptcl}} \rangle} \quad (4.1)$$

In Equation 4.1 the mean transverse momentum of reconstructed jets in a p_T bin is denoted by $\langle p_T \rangle$, and $\langle p_{T, \text{ptcl}} \rangle$ is the true mean of sum of the transverse momenta of the particles that at generator level make up the

jet in consideration. To clarify the previous, simply put R_{ptcl} is the ratio of reconstructed and true jet transverse momentum in simulation. The reconstructed jets are matched to the particle-level true jets by requiring the centroid of the reconstructed jet to be within $\Delta R = 0.25$ with respect to the particle-level jet in the case of $R = 0.5$ jet distance parameter.

When the response is determined as a function of η and p_T using a sample of 10 million simulated multijet events we get a fairly smooth two-dimensional response function that can be directly used for inverting the differences between reconstructed and true jet energies. The required operation is to multiply the reconstructed jet four momentum vector with the inverse of the determined response in the $[p_T, \eta]$ bin where the jet belongs to. The simulated raw jet response of anti- k_T $R = 0.5$ PFCHS-jets in five different energy regions is shown in Figure 4.10 (left) and the corrected response in four different pseudorapidity regions is shown on the right. The latter is called a closure study, which are routinely used for confirming the effectiveness of the correction. As Figure 4.10 (right) shows, the response is corrected to unity within one per cent for the entire energy region, and above 20 GeV is within a few per mil.

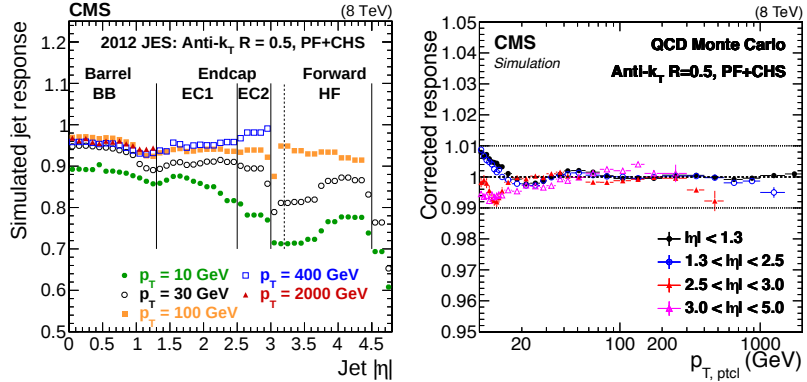


Figure 4.10. Left: simulated jet response before the correction in five energy regions as a function of absolute pseudorapidity $|\eta|$. Right: response after the correction is applied in four pseudorapidity regions as a function of true jet p_T . [57]

4.4.3 Residual corrections in η and p_T

The remaining small differences in jet response between data and simulation are corrected using data-driven residual corrections, which typically correct jet energies by $\mathcal{O}(1\%)$.

The residual differences with respect to jet direction η are studied and corrected using dijet events where one jet is in the barrel control region

$|\eta| < 1.3$ and other jet outside it. As the jets of a dijet system originating from the same parton-level process must have close to equal transverse momentum due to the conservation of momentum in the xy -plane, the jet in the control region can be used to measure the energy calibration of the other jet. With a sufficiently large data sample of dijet events this correction is able to restore the response of jets with $|\eta| > 1.3$ to have similar response with the jets in the control region.

The final step of the default jet energy correction sequence is the residual correction with respect to jet transverse momentum. This correction is derived using a combination of different measurements for covering the entire energy range. Events with a $Z(\rightarrow ee \text{ or } \mu\mu) + \text{jet}$ or $\gamma + \text{jet}$ process are used for correcting jet energies at the 30 to 700 GeV region by exploiting the precise electron, muon and photon measurements together with momentum conservation in the transverse plane. Using the Z decay products or photons as reference objects, miscalibration of the recoiling jet can be measured and correction factors derived.

As using Z -decays and photons as standard candles is not possible at jet p_T higher than roughly 700 GeV, the extrapolation of JEC to the TeV range has to rely on a more clever method. With a bit of creativity we can use the well calibrated lower- p_T jets to access the response of the TeV-scale jets. The trick is to exploit multijet events where a high- p_T jet is recoiled with a system of two or more softer jets. As the energy scale of the softer jets is well understood, we can once again use the transverse momentum conservation in the multijet system for calibrating the high- p_T jet.

The results of the above methods for the residual correction in p_T are combined by producing a *global fit* that then represents our best understanding of the true jet response in the energy range of 30 to 1000 GeV and which can be also extrapolated to higher jet energies. The global fit for the 8 TeV Run 1 data is shown in Figure 4.11 (left).

4.4.4 Jet energy scale uncertainties

Systematic uncertainties in the jet energy scale are estimated from various uncertainty sources and each layer of JEC is assigned with an uncertainty, as explained in detail in Reference [57]. After the combined uncertainty of the roughly twenty sources is calculated, the total uncertainty at $\eta = 0$ goes from 2% at 20 GeV down to 1% for a jet with 200 GeV of transverse momentum, and then increases slowly towards higher energies, as shown in Figure 4.11 (right). This shows that we have a historically pre-

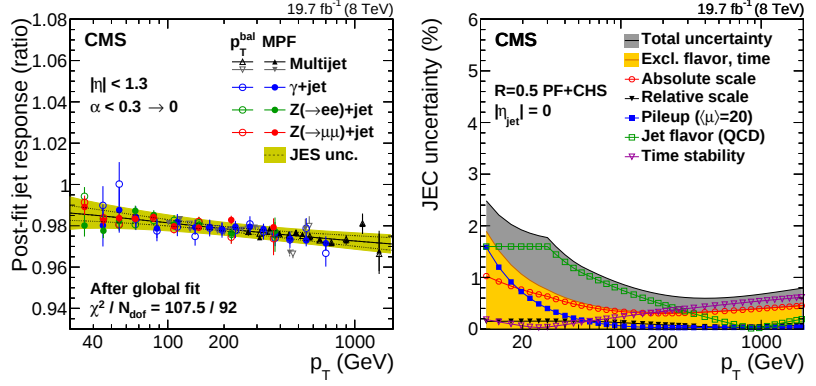


Figure 4.11. Left: global χ^2 fit to the four methods used for the p_T -dependent residual corrections. p_T^{bal} and MPF (E_T^{miss} projection fraction) refer to the two different methods for calculating the response from the recoil systems. Right: jet energy correction (JEC) uncertainty from different sources with respect to jet transverse momentum. [57]

cise jet energy scale, which is the result of great and strenuous efforts from the multi-national jet correction group of CMS, which for years has been convened from Helsinki.

The JEC uncertainty is one of the most significant sources of systematic uncertainty for analyses with jets in the final state. Typically only uncertainty in the luminosity calculation brings a bigger contribution to the total uncertainty of an analysis result.

5. Jet particology

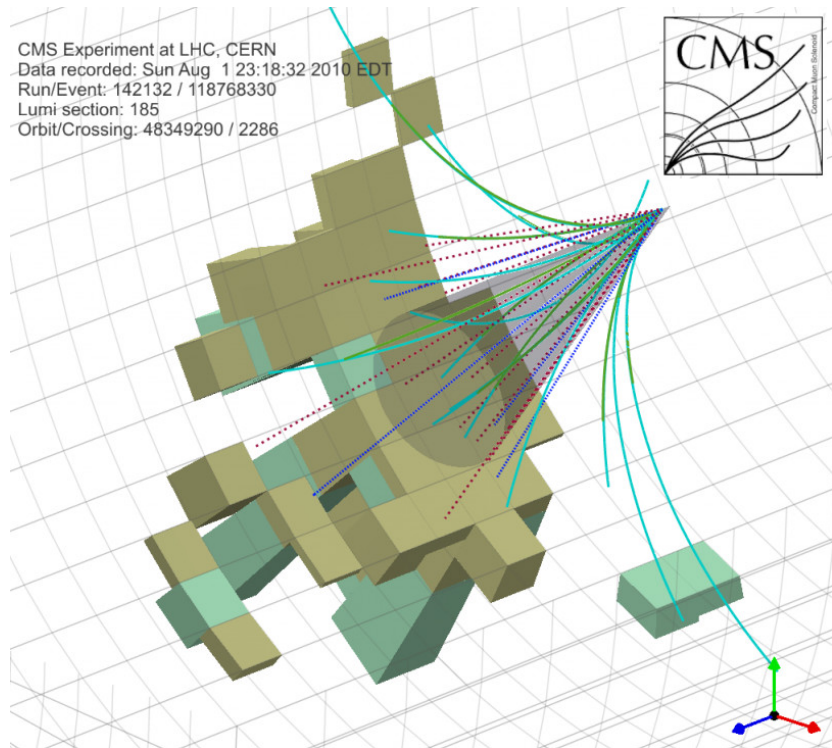


Figure 5.1. An actual jet reconstructed with the particle-flow algorithm. The constituent particles are classified as photons (red dashed line), neutral hadrons (blue dotted line) and charged hadrons (solid green and cyan lines). Electrons and muons are rarely present in jets. The ECAL deposits are drawn as olive green wedges and HCAL deposits in mint green. [65]

Throughout the history of experimental particle physics jets have been considered as a garbled collection of particles that leave their energy to the calorimeters and not much effort has been put for experimentally accessing the actual contents of the hadronic showers, probably mostly because the technology was not yet ready for closer inspection of jets. It is only with the combined advantages of the particle-flow algorithm and

the design of the CMS detector that facilitate what we call jet particology, the detailed study of jets as collections of distinguishable and classifiable individual particles.

In this chapter the access to the particle-level jet information is used in three different studies: for measuring how the energy of jets is distributed to different particle types, for studying how robust the PF-jets are to calorimeter calibration and for measuring the effectiveness of the charged hadron subtraction algorithm in pileup jet reduction.

5.1 Jet energy composition

By exploiting the PF particle classification and the bottom-up jet reconstruction approach we can look inside jets and measure the jet energy composition, *i.e.* see how much of the total jet energy is carried by which kind of particles. We can divide the energy content into the five categories that PF classifies the reconstructed particles to: charged and neutral hadrons, photons, electrons and muons. The latter two are merged under the category ‘Leptons’ here due to their minuscule contributions.

Furthermore, by the virtue of the CHS algorithm, we can separate the pileup component of the charged hadrons into its own category. In the forward region, $|\eta| > 2.4$, where the absence of tracking information reduces the particle classification capabilities of particle-flow, the jet energy is divided into hadronic and electromagnetic components.

5.1.1 Composition in simulation

The full particle composition of jets originating from 8 TeV proton-proton collisions as predicted by the tune Z2* of the PYTHIA 6.4 event generator [95] as a function of jet transverse momentum is shown in Figure 5.2. In this simulation particles are considered stable if for them holds $c\tau > 1$ cm, where c is the speed of light and τ is the lifetime of the particle in its reference frame. Otherwise the simulated particles are made to decay, most notably π^0 :s that promptly decay to photons. This study does not take into account the delay in particle decays due to time dilation in the reference frame of the relativistic particles.

Starting from the bottom of the stacked energy fractions of Figure 5.2, charged pions ($\pi^\pm$) make roughly 40% of the jet energy, the other notable charged meson contribution coming from charged kaons ($K^\pm$) with 10%

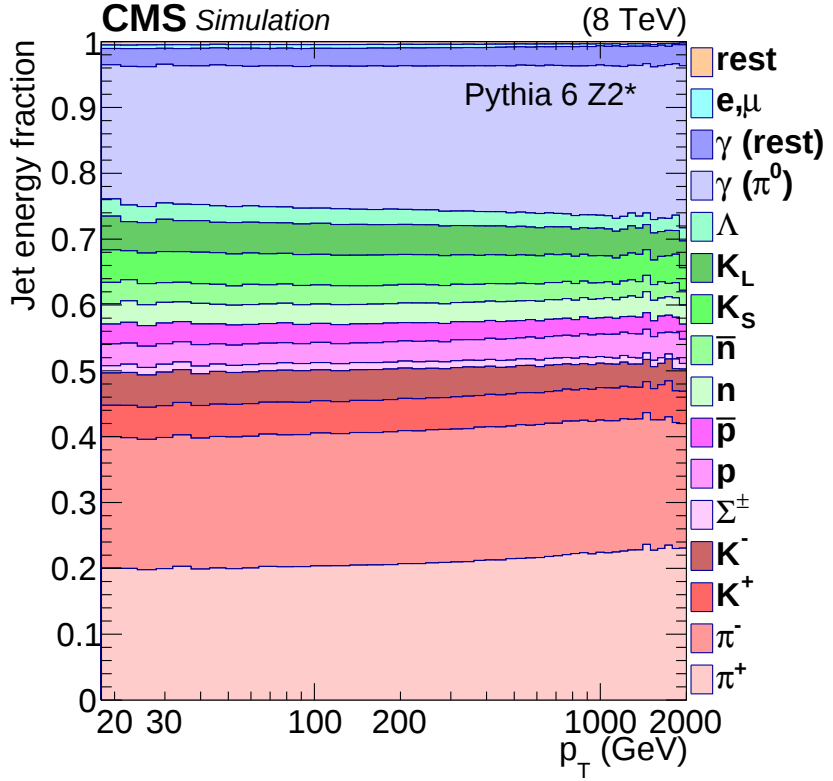


Figure 5.2. Jet energy composition as a function of jet p_T at $|\eta| < 1.3$ as predicted by the PYTHIA 6.4 Z2* event generator. The charged hadron contribution is stacked as the seven bottommost components drawn in different hues of red and violet. Neutral hadron component are shown next as the five green layers before photons, which are drawn in light and dark blue. The minuscule contributions of leptons (e and μ) and all the other particles are drawn to the top of the stack in cyan and peach, respectively. The contributions of different particles are presented in the stack in the same order as in the legend on the right. [57]

energy fraction. Charged baryons (p , $\bar{p}$ and $\Sigma^\pm$) bring another 8%, so the total energy fraction of charged hadrons is 58%, and it increases slightly as a function of jet transverse momentum. The 1% increase in charged hadron fraction originates from the increasing fraction of π^+ -particles in the jets, as can be seen from the bottommost component of the fraction stack in Figure 5.2. While the π^+ -content increases by as much as 3%, the other charged hadron components decrease by 2%. The aforementioned phenomenon can be qualitatively understood to arise from the steep increase in quark jets and decrease in gluon-jets as a function of jet energy as shown in Figure 5.3. Here in the used simulations quark and gluon jets are defined by matching them to the outgoing partons of the $2 \rightarrow 2$ process at the generator level. In data only probabilities can be assigned to jets

by studying jet substructure and kinematics, as explained for example in Reference [16]. As the quark jets produced in proton-proton collisions are more likely to originate from a positively charged up-quark than a down-quark, this converts to a stronger production of positive pions. Gluon-jets naturally do not favor any charge over another. Further studies, however, are needed for a solid comprehension of the phenomena.

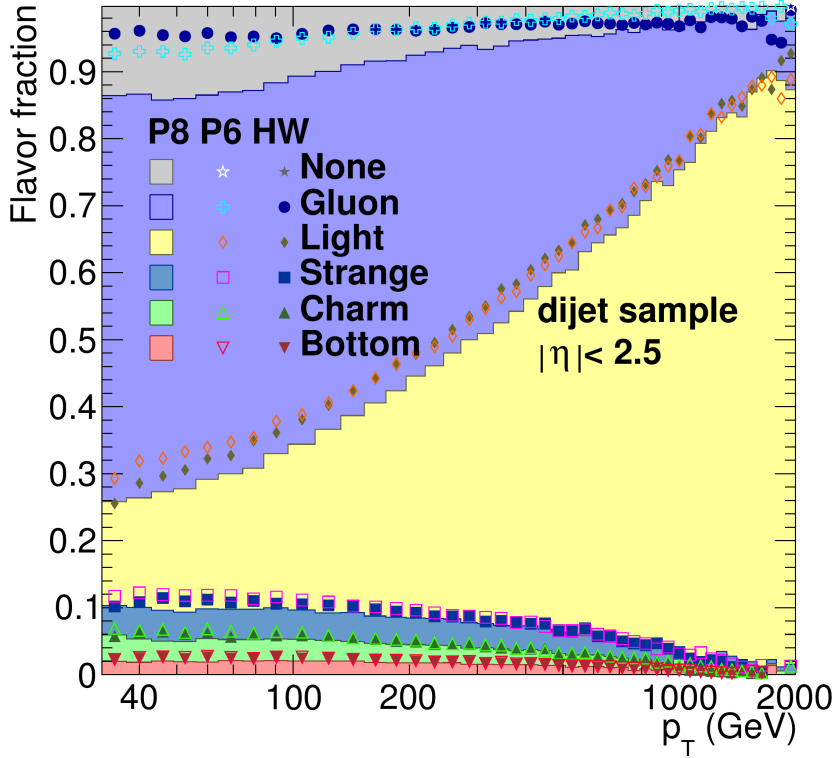


Figure 5.3. The fraction of jets originating from different parton types in dijet events as a function of jet transverse momentum. The filled histogram shows the prediction of PYTHIA8 simulation (P8), empty markers show PYTHIA6 simulation (P6) and full markers represent Herwig++ simulation (HW). The fraction of light quark (*i.e.* u, d) jets is seen to increase steeply with the increasing jet p_T . [94]

Neutral hadrons contribute roughly 17% to the jet energy, out of which 5% is carried by neutrons (n) and anti-neutrons ($\bar{n}$), 8% by long-lived kaons (K_L) and short-lived kaons (K_S), and 2% by lambda-baryons (Λ). The neutral hadron fraction decreases by about 2% from 18% at 20 GeV to 16% at 2 TeV.

Photons from the virtually instantly decaying π^0 mesons make 20% of the jet energy – just like π^+ and π^- mesons that get copiously produced in the hadronization process. The energy fraction from the $\pi^0 \rightarrow \gamma\gamma$ process increases from 20% at 20 GeV to 22% at 2 TeV. A small amount of

isolated photons is also produced and brings a 3% energy contribution to jets. Electrons, muons and all the other particles contribute roughly 1% to the total energy of an average jet.

A number of differences arising from experimental effects will be observed between this purely theoretical prediction and the jet composition measured from data. These differences will be examined in detail in the following sections.

5.1.2 Composition with respect to jet p_T

Jet energy composition with respect to jet transverse momentum in data and simulation in the barrel region $|\eta| < 1.3$ is shown in Figure 5.4. At a first glance on average the jet energy composition follows the same pattern as seen in Figure 5.2; charged hadrons and photons from neutral pions make the bulk of jets with a smaller contribution coming from pseudo-stable neutral hadrons and a minuscule fraction coming from electrons and muons. A closer inspection, however, reveals three significant differences between data and simulation.

Pileup effect

The charged hadrons associated with pileup vertices bring a significant contribution to low- p_T jets with up to 14% share, which decreases to zero for the TeV scale jets where the relative energy contribution from pileup contamination diminishes. Another effect speeding up the decrease of the pileup component is the fact that energetic jets are typically narrow, thus leaving less volume for the uniformly distributed pileup energy. The charged pileup component, drawn in dark red histogram (MC) and open circles (data), is removed from jets by the CHS algorithm, but is preserved here as we want to study its behavior. The smoothly decreasing pileup component also causes slight energy dependence to the jet energy composition up to the energy of 100 GeV, below which modest downward slope is seen in charged hadron and photon contributions. This, however, is only an artefact arising from visualizing the jet energy composition in stacked fractions; the other energy components are shifted upwards at low jet energies due to the significant pileup contribution there, but reach a plateau at higher energies where pileup component becomes small. The excellent agreement between data and simulation shows that the pileup component in jets is well understood and properly modeled by the simulation.

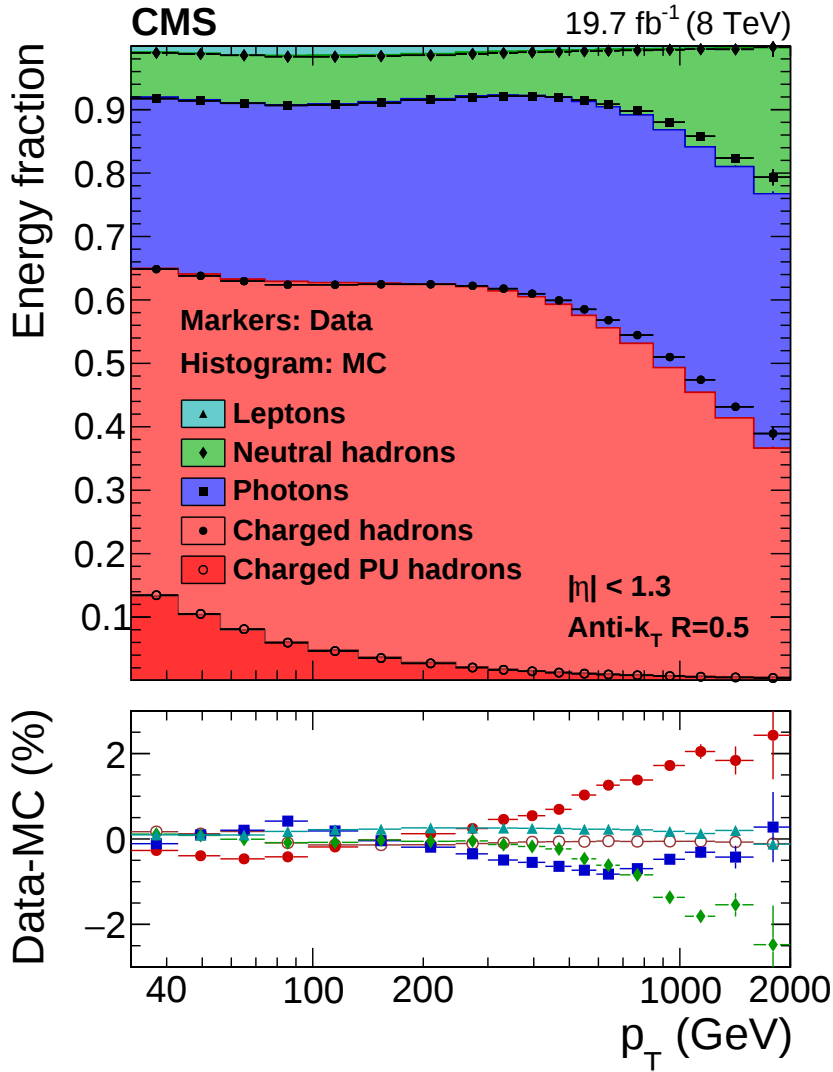


Figure 5.4. Jet energy composition in the barrel as a function of jet p_T with a data sample corresponding to 19.7 fb^{-1} of 8 TeV proton-proton collisions (Data) and a simulation sample of 10 million QCD dijet events (MC). Measured fractions are drawn with markers and the filled histograms show the theory prediction after the detector simulation. Charged hadrons associated to pileup (PU) vertices are drawn separately and the minuscule contribution from electrons and muons are merged together to 'Leptons' category. The bottom panel shows the difference of data and simulation. [58]

Overlapping tracks

Where in simulation of Figure 5.2 the jet composition as a function of jet p_T is nearly flat, in Figure 5.4 there is a clear decrease in charged hadrons that is compensated by increase in neutral hadrons and photons in the region between 400 and 2000 GeV. This is understood to arise from the high density and low curvature of tracks inside the narrow cores of high- p_T jets. When track density increases sufficiently to produce tracks that overlap within the tracking granularity, the capability of particle-flow in identifying charged hadrons diminishes, which results to a higher amount of reconstructed neutral hadrons and photons. More precisely, although PF classifies the measured track energy that is followed by deposits in ECAL and HCAL as charged hadrons, when track overlaps occur, only a part of the true tracks can be identified and thus be linked to the calorimeter deposits. The calorimeter energy that cannot be associated to any tracks is classified as photons (when energy is present in ECAL only) or neutral hadrons (when energy is present in ECAL and HCAL). The charged hadron fraction, including pileup, decreases from the 63-65% below 100 TeV to less than 40% at 2 TeV. The photon fraction correspondingly increases by more than 10% from around 28% at 100 GeV to 40% at 2 TeV. The other 10% that vanishes from the charged hadron fraction is seen in neutral hadrons, which are seen to contribute twice as much to jet energy at 2 TeV as they are at lower energies; the neutral hadron fraction grows from around 10% contribution at the 40 to 500 GeV range to 20% at 2 TeV.

The effect of decreasing charged hadron fraction and the compensating increase in neutral hadrons and photons is fairly well modeled by the simulation, as can be seen from Figure 5.4. The bottom panel of 5.4, however, shows a systematically growing deviance between data and simulation at the 400 to 2000 GeV energy range. In the simulation the charged hadrons decrease more than in the data, which is balanced by an increase in the simulated photon fraction at the TeV-scale. This discrepancy is understood to be caused by insufficient modeling of the particle classification inefficiency, which is caused by the overlapping tracks at the dense cores of energetic jets. Preliminary results with more recent version of the detector simulation indicate that the modeling of the effects of track overlaps has become more realistic in 13 TeV simulations, as such systematic deviance is not observable there.

In-flight decays

The third outstanding difference between the simulation of Figure 5.2 and the measured composition in Figure 5.4 is the significantly smaller neutral hadron fraction in the measurement. Where in the pure PYTHIA 6 simulation pseudo-stable neutral hadrons make 15-18% of the average jet energy, the measurement shows that particle-flow finds only 7-8% of jet energy to be carried by neutral hadrons at the 40 to 500 GeV region where the above discussed effect of overlapping tracks do not cause considerable effects yet. The smaller neutral hadron fraction is balanced by an overall increase in charged hadrons and photons with respect to the pure simulation, which both increase roughly in proportion to their energy fractions. This means that the energy missing from neutral hadrons is roughly evenly distributed to the other two dominant jet energy components, with slightly bigger proportional increase in the photon fraction.

While this effect is well modeled by the detector simulation as shown by the excellent agreement between data and simulation below 500 GeV in Figure 5.4, full understanding of the mechanisms behind the phenomenon calls for some careful detective work. The major part of the phenomenon can be explained to be caused by short-lived neutral hadrons that are considered stable in the study of Figure 5.2, as they fulfill $c\tau > 1$ cm, but in reality decay already inside the beam pipe or inside the tracker before reaching calorimeters. The short-lived particles which carry considerable fraction of jet energy are K_S and Λ particles, with mean flight distance of 2.7 and 7.8 cm, respectively. The short-lived kaons make up 5% of jet energy and lambdas around 2%. K_S decay 70% of the time to a pair of charged pions with the remaining 30% going to π^0 's, which instantly decay to pairs of photons. All in all, then, K_S decays inside the detector can be approximated to cause 3.5% of the energy flow from neutral hadrons to charged hadrons and 1.5% to photons.

Although Λ 's have almost three times longer mean flight distance than short-lived kaons, remembering that the first calorimeter layer of CMS is more than a meter away from the nominal interaction point, we can approximate with negligible bias that all the lambdas decay already before the reaching the ECAL. The dominant decay processes are $\Lambda \rightarrow p\pi^-$ with 64% branching fraction and $\Lambda \rightarrow n\pi^0 (\rightarrow \gamma\gamma)$ with 36% branching fraction. Thus, the effect to jet energy composition is a 1.3% increase in charged hadrons and a 0.35% increase in photons, the other 0.35% from the $\Lambda \rightarrow n\pi^0$ decay staying in the neutral hadron component.

Let us take stock. The total observed difference in neutral hadron fraction between the prediction of Figure 5.2 and the measurement of Figure 5.4 for a jet of 100 GeV is around 9%. With the above reasoning we can understand 7% of this to be caused by the in-flight decays of short-lived neutral hadrons. The 7% distributes to an increase of 4.8% in the charged hadron fraction and of 1.85% in photon fraction. The remaining 0.35% stays in neutral hadrons. At 100 GeV the charged hadron fraction is seen to be 5% higher in measurement than in simulation, which is well explained by the decaying K_S and Λ particles. For photons the increase at the same energy is 4%, out of which approximately 2% can be accounted for the decaying hadrons. The remaining 2% then needs to be explained by some other mechanism. The leading hypothesis is that despite the careful calibration sequence of the CMS event reconstruction, the difficulty of reconstructing neutral hadrons leads to residual miscalibration, which is observed by the 2% deficiency. It is also possible that the capability of particle-flow algorithm in separating charged hadrons from photons systematically underestimates the neutral hadron fraction, as the classification using the rather similar detector signals is very difficult if not impossible to carry out unambiguously. A thorough inspection of the reconstruction process on a step-by-step level would be necessary to see what is the true origin of this 2% difference.

5.1.3 Composition with respect to jet η

Jet energy composition with respect to jet pseudorapidity is shown in Figure 5.5. Not surprisingly, the general distribution of jet energy is in line with what we have seen above. In Figure 5.5 the composition is measured in the energy interval of 56 to 74 GeV, where at $\eta = 0$ the jet energy is distributed to 8% in charged pileup, 56% in charged hadrons, 27% in photons, 8% in neutral hadrons and 1% in leptons *i.e.* electrons and muons.

While in the central region around $|\eta| < 1$ the composition is fairly flat, starting from $|\eta| = 1$ non-trivial features are present. First of all, the neutral hadron fraction increases together with lepton fraction until around $|\eta| = 1.6$. At this pseudorapidity a significant increase in charged hadron fraction is observed. Charged hadrons rise from 64% contribution up to 69% at $|\eta| = 1.8$, after which the fraction starts to decrease. The fast decrease in charged hadrons is compensated by a rapid increase in photon fraction at $1.8 < |\eta| < 2.2$, and in both photon and neutral hadron fractions at $2.2 < |\eta| < 3.0$. This can be explained by the tracker coverage

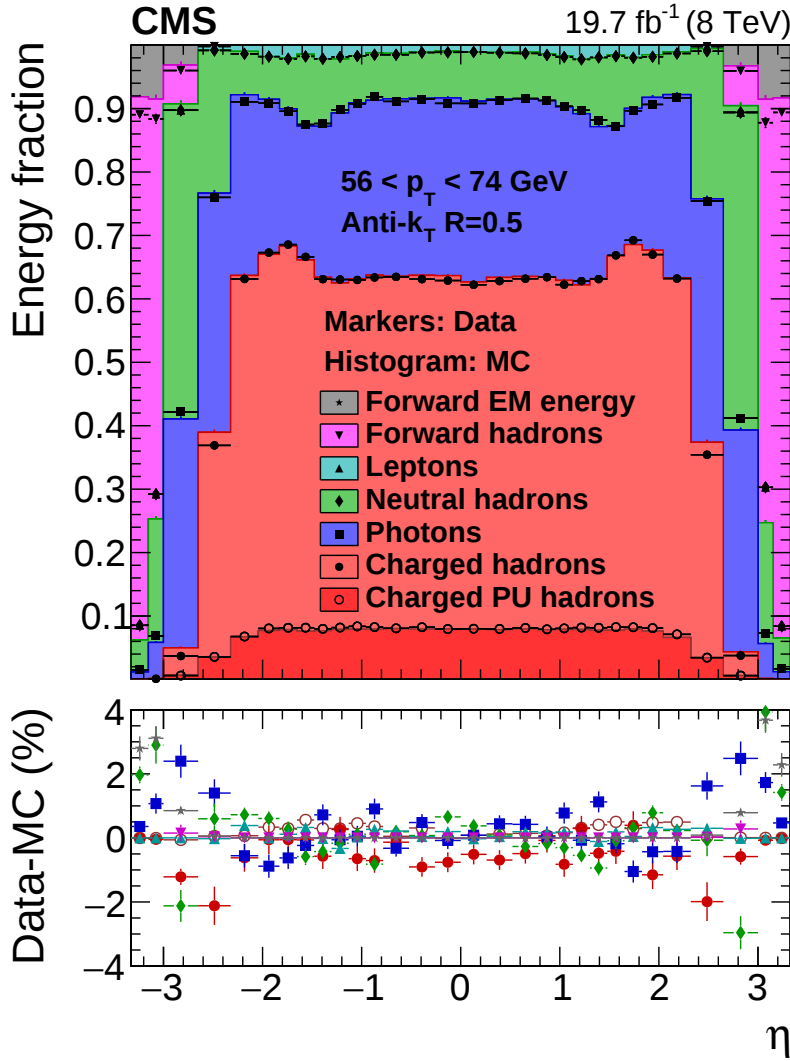


Figure 5.5. Jet energy composition at the energy range of 56 to 74 GeV as a function of jet η with a data sample corresponding to 19.7 fb^{-1} of 8 TeV proton-proton collisions (Data) and a simulation sample of 10 million QCD dijet events (MC). Measured fractions are drawn with markers and the filled histograms show the theory prediction after the detector simulation. Charged hadrons associated to pileup (PU) vertices are drawn separately and the minuscule contribution from electrons and muons are merged together to 'Leptons' category. At the forward region $|\eta| > 3$ the jet energy is categorized to forward hadronic energy and forward electromagnetic energy. The bottom panel shows the difference of data and simulation. [58]

ending at around $|\eta| = 2.5$, after which charged hadrons cannot be identified by particle-flow and energy is reconstructed as neutral hadrons and photons instead. The forward region, $|\eta| > 3$, is outside the coverage of ECAL and HCAL, and thus there particles are only classified as hadrons or electromagnetically interacting particles due to the more simple detector structure.

Although the detector simulation models well the effects described above, the jet energy composition with respect to jet η is not thoroughly understood, and more studies are needed to fully understand the phenomena. However, as the detector simulation reproduces the aforementioned unknown effects, a careful step-by-step inspection of the event reconstruction process would once again be helpful for understanding the full η -dependence of jet energy composition.

5.1.4 Composition with respect to pileup

Jet energy composition as a function of the pileup variable μ is shown in Figure 5.6. The μ -variable is defined as the mean event pileup over 24 seconds of data taking, during which the instantaneous luminosity and thus pileup conditions can be approximated to stay constant. The composition versus pileup is studied in the barrel region, $|\eta| < 1.3$, and is averaged over the jet transverse momentum interval of 56 to 84 GeV.

The pileup dependence of jet energy composition is studied using bins with the width of one interaction, with the bins ranging from 7 to 35 proton-proton interactions per event. As expected, the charged pileup hadron fraction increases steadily with μ and increases from 2% at $\mu = 7$ to 12% at 35 simultaneous collisions per bunch crossing. The increase in pileup energy is balanced by a decrease in charged hadrons from the leading interaction vertex, which goes down from 60% at $\mu = 7$ to 50% at $\mu = 35$. Photon, neutral hadron and lepton fractions are not affected by the number of pileup interactions and stay constant with 27%, 9% and 1% contributions, respectively.

As the exact pileup conditions of LHC collision runs were not known at the time when the Monte Carlo simulation samples were produced, the simulated events are carefully re-weighted so that the effects of pileup to simulated jet composition are similar with the effects to measured events. The re-weighting is implemented by weighting simulated events according to the probability of finding an event with similar pileup conditions in the data. We re-weight simulation based on the variable μ , which in

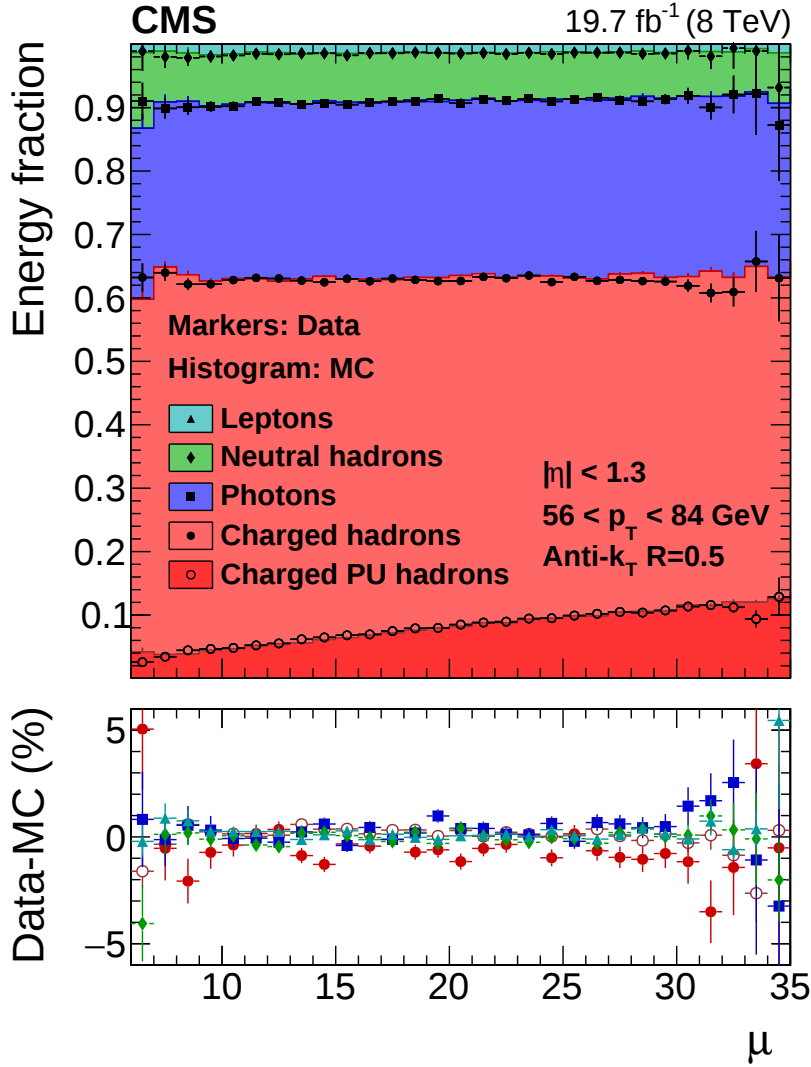


Figure 5.6. Jet energy composition at the energy range of 56 to 84 GeV and at pseudo-rapidity $|\eta| < 1.3$ as a function of average number of pileup interactions per event μ with a data sample corresponding to 19.7 fb⁻¹ of 8 TeV proton-proton collisions (Data) and a simulation sample of 10 million QCD dijet events (MC). Measured fractions are drawn with markers and the filled histograms show the theory prediction after detector simulation. Charged hadrons associated to pileup (PU) vertices are drawn separately and the minuscule contributions from electrons and muons are merged together to 'Leptons' category. The bottom panel shows the difference of data and simulation. [58]

simulation is defined as the Poisson mean of the number of interactions, from which observed in-time and out-of-time pileup are generated. For data μ is the mean number of interactions per bunch crossing, and it is calculated from the instantaneous luminosity and the total proton-proton interaction cross section.

After the re-weighting procedure, which is done for all the different composition studies in Figures 5.4 – 5.6, the measured pileup and also other energy components are in very good agreement with the simulation, as confirmed by the bottom panel of Figure 5.6. The differences are mostly within one per cent, apart from some statistical fluctuation at low and high μ due to modest number of events per bin in the measured data.

5.2 Sensitivity of jet response to single pion response

Jet response, as defined in Equation 4.1, tells us how far off and to which direction our measured jet energy is from the true jet energy, which in simulations is defined by clustering stable particles at the generator level, as described in Section 4.4. It is the most important measurable quantity when we study jet energy scale and derive correction coefficients. By exploiting the detailed detector simulation of the CMS detector and a lightweight version of detector simulation, we can study how the calibration of different detector components affect jet response. As we can also study the response of calorimeter-based calo-jets at the same time, this research is instrumental in measuring quantitatively how much the particle-flow jet reconstruction approach brings advantages to jet measurements over the traditional method of clustering jets from calorimeter deposits only. The latter method is used in the calo-scouting flavor of the dijet resonance search introduced in Chapter 6.

The energy scale and calibration of calorimeters is one of the leading sources of systematic uncertainty in jet energy measurements. With the help of Fast simulation [93], the streamlined parameterization of the full CMS detector simulation, we can modify the properties of different detector elements and study how the modifications affect jet measurements. With this computationally lightweight simulation we can change the response of the CMS calorimetry and measure the resulting change in jet response. We varied the response of ECAL and HCAL to single pions simultaneously by $\pm 3\%$ for measuring the total effect of single pion response to jet response and also varied the response of ECAL and HCAL separately for measuring how miscalibration in either of the calorimeters affects jet reconstruction performance. Although the calorimeters are carefully calibrated with test beams, and their response to charged pions is also checked with proton-proton collision data. The response varies with particle transverse momentum and the uncertainty in the energy measurement of single pions by calorimetry is measured to be within $\pm 3\%$ in the barrel region. [42]

These studies are done with a modest sample of 100 000 simulated QCD dijet events, which is, however, well sufficient for the purpose, when the same generator-level events are used with varying detector conditions. As the events are the same in each scenario, *i.e.* 100% correlated, all other differences than the calorimeter sensitivity are eliminated and all the ob-

served effects are due to the introduced modifications. To further clarify the simulation sample production, a total of seven independent simulation and reconstruction iterations are done in order to get the seven different samples needed for the studies: one with ECAL+HCAL +3% and one with -3% modification, and similarly two each for the samples where only ECAL or HCAL sensitivity is modified. The seventh sample is the control sample in the denominator of the ratio studies, where the default values of the detector calibration are used. The 100% correlation between the samples is ensured by using the same random number generator seed for each simulation+reconstruction iteration.

5.2.1 ECAL+HCAL response variation

The effect of varying the ECAL+HCAL single pion response by $\pm 3\%$ to PF-jets and calorimeter-based calo-jets is shown in Figure 5.7. The effect is studied by measuring the ratio of jet response with (numerator) and without (denominator) the $\pm 3\%$ modification of ECAL and HCAL sensitivity.

For calo-jets the particle response variation largely propagates directly to jet response, and we observe roughly $\pm 2\%$ shift in the calo-jet response with respect to unity, which would mean that the jets were completely unaffected by the sensitivity modification. The big effect of the variation to calo-jet response over the entire energy range of 20 to 2000 GeV is the expected behavior, as calo-jet rely entirely on calorimeter measurements. Only two thirds of the introduced variation propagates to the calo-jet response, because only roughly two thirds of jet energy is carried by hadrons, as was shown in Section 5.1. The energy fraction carried by photons, which are accurately measured with ECAL, is not affected by this modification.

For PF-jets the response ratio varies from less than 0.5% below 40 GeV to 2% at 2 TeV. This is also an expected result, which shows how particle-flow jets are insensitive to calorimeter scale at low energies. This is by virtue of the exploitation of the tracking information in the reconstruction process, where almost no calorimeter information is needed for soft jets. At low jet p_T the PF candidates are reconstructed using also track information, which provides more precise energy measurement than pure calorimeter measurement. The observed behavior of the PF-jet response ratio starting from almost unity and by steady increase approaching the calo-jet behavior is due to the difficulties in tracking increasing with jet

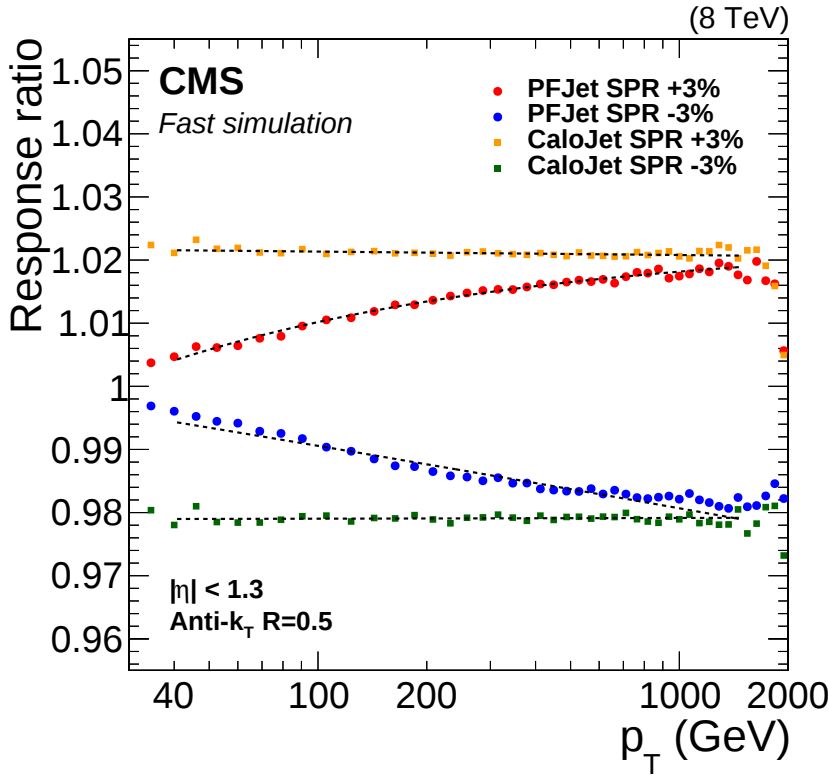


Figure 5.7. Ratio of varied and unmodified response of anti- k_T $\Delta R = 0.5$ jets at $|\eta| < 1.3$ as a function of jet p_T from 30 to 2000 GeV, when the single pion response (SPR) of ECAL and HCAL is varied by $\pm 3\%$ in CMS Fast simulation. The response ratio of particle-flow jets (PFJet) is shown as red (+3%) and blue (-3%) circles, and for calorimeter-based CaloJets as yellow (+3%) and green (-3%) squares. [57]

energy, which is the same effect that was already discussed in Section 5.1.2 and also greatly affects the jet composition energy behavior in Figure 5.4. As the jet energy grows, the track density and track overlaps inside the jet also increases, which greatly deteriorates the tracking efficiency and forces the PF to rely more and more on the calorimeter information. At around 2 TeV the advantages of exploiting tracks in jet reconstruction are mostly gone, and the TeV-scale jets are measured only by calorimetry.

This study is yet another proof of the advantages of the particle-flow jet reconstruction approach, and the results of Figure 5.7 are also directly used in assessing the jet energy scale uncertainties, a major contribution to which comes from the calorimeter single pion response uncertainty.

Unfortunately the modest event sample used for these studies proves insufficient for studying the effects accurately above 1 TeV jet transverse momenta, as the statistical fluctuations start to dominate there. A similar

study with at least an order of magnitude more simulated events would be beneficial for extending the study to the TeV scale. The importance of thorough understanding of jet behavior at very high transverse momenta is pronounced at LHC Run 2, as the 13 TeV collision energy makes possible creation of jets with transverse momenta up to 3-4 TeV, as will be seen in Chapter 6.

5.2.2 ECAL scale variation

In Figure 5.8 the effect of single pion response to jet response is studied separately in ECAL. The sensitivity of ECAL to single pions is varied by $\pm 3\%$ while keeping HCAL at its default sensitivity.

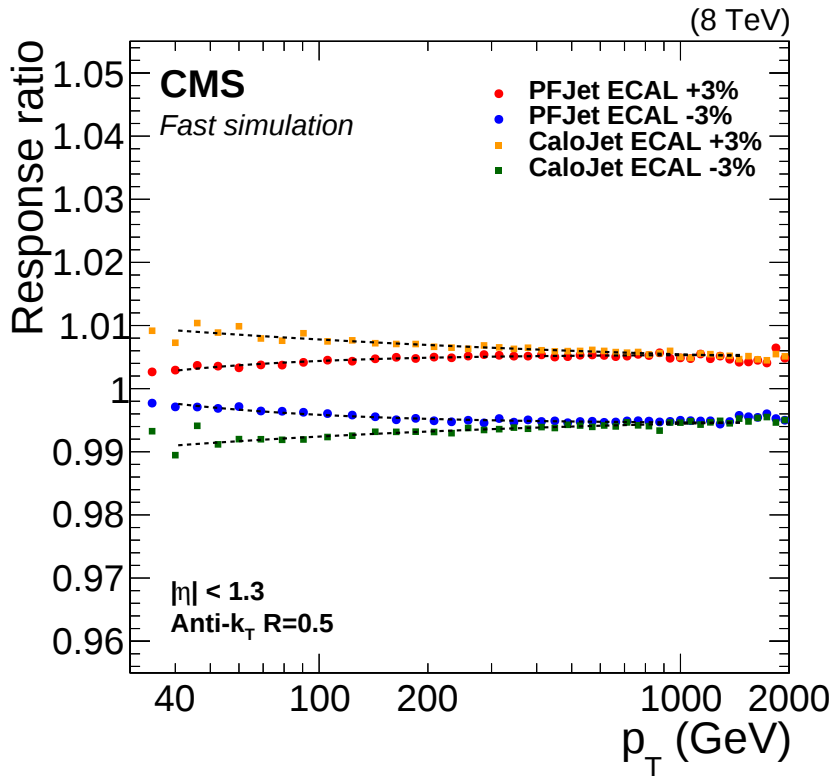


Figure 5.8. Ratio of varied and unmodified response of anti- k_T $\Delta R = 0.5$ jets at $|\eta| < 1.3$ as a function of jet p_T from 30 to 2000 GeV, when single pion response of ECAL is varied by $\pm 3\%$ in CMS Fast simulation. Response ratio of particle-flow jets (PFJet) is shown as red (+3%) and blue (-3%) circles, and for calorimeter-based CaloJets as yellow (+3%) and green (-3%) squares. [57]

The effects of single pion response modification in ECAL are rather modest when compared to the study of Figure 5.7, which is because the hadronic component of jet energy is mostly measured in HCAL and only

the lowest energy jets have an appreciable hadronic contribution from ECAL. At 40 GeV the calo-jet response is affected by 1% and PF-jet response by 0.2%, and both go to 0.5% at 2 TeV. The decreasing sensitivity of calo-jets follows from the fact that the proportion of jet energy from hadronic activity in ECAL decreases with energy. For PF-jets the behavior is similar as in Figure 5.7: at low p_T the track information yields robustness with respect to calorimeter calibration, but with increasing jet energy the result of particle-flow jet reconstruction approaches that of calorimeter-only reconstruction.

5.2.3 HCAL scale variation

The effect of HCAL $\pm 3\%$ single pion response variation to jet response is shown in Figure 5.9, where now ECAL is kept at its default sensitivity. For clarification, the results of Section 5.7 can be reproduced by adding together the results of Figures 5.8 and 5.9.

The effects of the variation are seen to be much more pronounced in the case of HCAL, which is natural as the lion's share of the hadronic showers are contained in the thick and dense HCAL. At $p_T = 40$ GeV the calo-jet response is shifted by 1.2% and slightly grows with energy and reaches 1.6% at 2 TeV. This is due to the same fact why the ECAL ratio in Figure 5.8 goes down; the relative proportion of jet energy in HCAL increases with energy as the showers get longer. PF-jets behave as already seen twice in the previous studies; the effect of the variation is almost negligible for soft jets, for which the effect is only $\mathcal{O}(0.1\%)$ as they are dominantly measured with tracks, but the response ratio grows with increasing jet energy and approaches the calo-jet ratio due to the decreasing tracking efficiency and greater dependence on calorimetry.

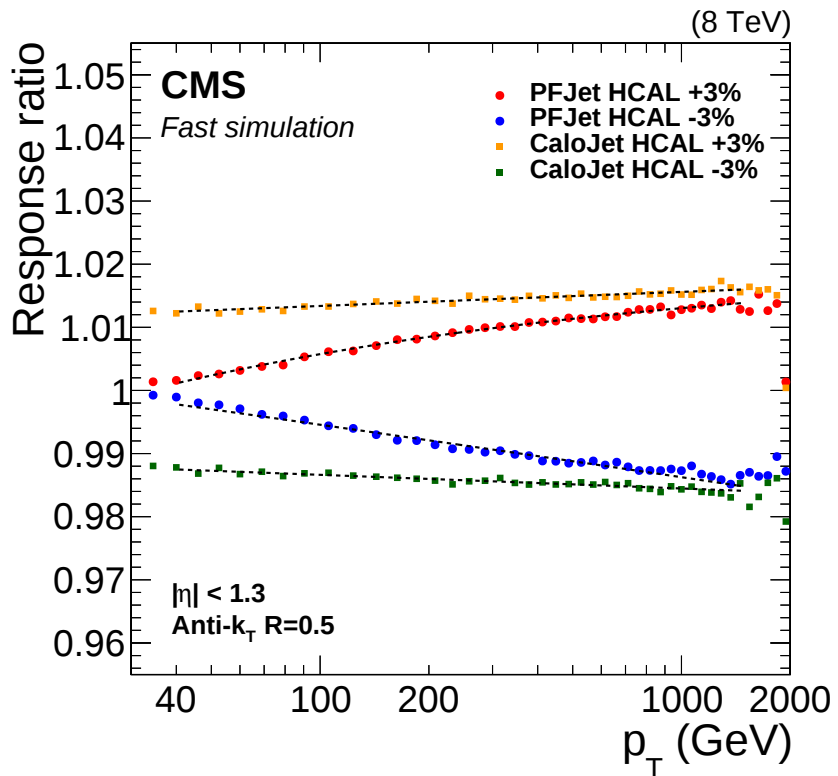


Figure 5.9. Ratio of varied and unmodified response of anti- k_T $\Delta R = 0.5$ jets at $|\eta| < 1.3$ as a function of jet p_T from 30 to 2000 GeV when single pion response of HCAL is varied by $\pm 3\%$ in CMS Fast simulation. Response ratio of particle-flow jets (PFJet) is shown as red (+3%) and blue (-3%) circles, and for calorimeter-based CaloJets as yellow (+3%) and green (-3%) squares. [57]

5.3 Charged hadron subtraction

The charged hadron subtraction algorithm is an invaluable tool in mitigating the difficulties introduced by the harsh pileup conditions, which are unavoidable if we want to maximize the number of proton-proton collisions and thus maximize the potential for new discoveries. CHS also helps directly in measuring jets more accurately, as a considerable part of the extra energy contained in jets from pileup interactions is directly removed already before jet reconstruction. As will be shown in the study of this section, the CHS algorithm removes a good part of pileup energy from the jets originating from the hard interaction process and improves jet energy resolution. It also efficiently removes pileup jets, *i.e.* soft jets with $\mathcal{O}(10)$ GeV transverse momenta. Pileup jets are formed by overlapping low-energy jets from non-leading interaction vertices.

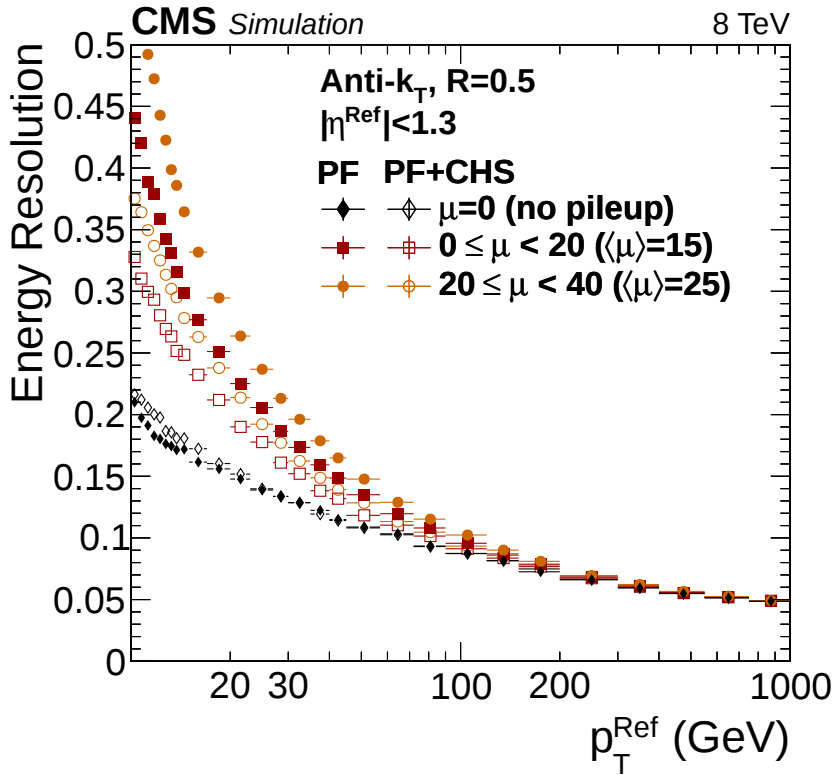


Figure 5.10. Energy resolution of PF-jets (full markers) and PF+CHS jets (open markers) as a function of jet p_T , studied with simulated events. The effect of charged hadron subtraction (CHS) is studied in three scenarios with different number of pileup interactions denoted by μ : without pileup *i.e.* $\mu = 0$ (diamonds), with $0 \leq \mu \leq 20$ with $\langle \mu \rangle = 15$ (squares) and with $20 \leq \mu \leq 40$ with $\langle \mu \rangle = 25$ (circles). The energy resolution is significantly improved *i.e.* decreased by virtue of the CHS procedure. More details are given in the text. [58]

In Figure 5.10 the effect of CHS for jet energy resolution is studied by comparing the resolution of normal PF-jets and PF+CHS jets for which the charged hadrons from pileup vertices are subtracted from the events before jet clustering. The effect is studied in three pileup conditions, out of which the one without pileup is used as a control case showing that it is indeed the CHS procedure that causes the effects observed in the other two cases. CHS does not significantly affect the resolution in the zero pileup scenario. With 0 to 20 pileup interactions per event the energy resolution for 20 GeV jets is decreased by 16% from 25% for PF-jets to 21% for PF+CHS jets. In the high pileup scenario of 20 to 40 pileup vertices the improvement is even bigger, as the energy resolution at 20 GeV decreases by 20% from 30% to 24%. The advantages brought by CHS to jet resolution gradually weaken with increasing jet p_T , as the relative contribution of pileup energy diminishes and other sources of uncertainty in the energy measurement start dominating. Already at 50 GeV the difference in resolution between PF- and PF+CHS-jets is around 10%, and for jets with $p_T = 200$ GeV and higher the improvement is negligible.

In Figure 5.11 we have studied how the application of the CHS algorithm affects the number of jets as a function of jet pseudorapidity. Jets in the simulated events are categorized to hard, soft and pileup jets. Hard and soft jets are jets which are matched to the generator level jets originating from the lead-vertex with the criterion

$$\Delta R_{gen} \equiv \sqrt{(\phi_{gen} - \phi_{reco})^2 + (\eta_{gen} - \eta_{reco})^2} < 0.25,$$

which ensures that the reconstructed jet is the same as the generated one, save for some vanishingly rare event configurations which do not cause appreciable bias to the studies. If a matched jet is one of the two *leading jets*, *i.e.* has the highest or second-highest p_T of all the jets in the event, it is classified as a hard jet. Otherwise a matched jet is called a soft jet. This categorization is justified by leading-order QCD predictions, where jets are typically created in pairs with balanced transverse momenta, with any additional jets usually much lower in p_T . Jets that are not matched to any generator level jet from the lead vertex are classified as pileup jets.

Events are chosen so that the average transverse momentum of the two leading jets is between 100 and 150 GeV and only jets with reconstructed $p_T > 25$ GeV are considered due to the limitations coming from the jet energy cut in the used simulation sample. The uncertainties are included in the study by summing the statistical uncertainties and systematic uncertainties from jet energy corrections in quadrature.

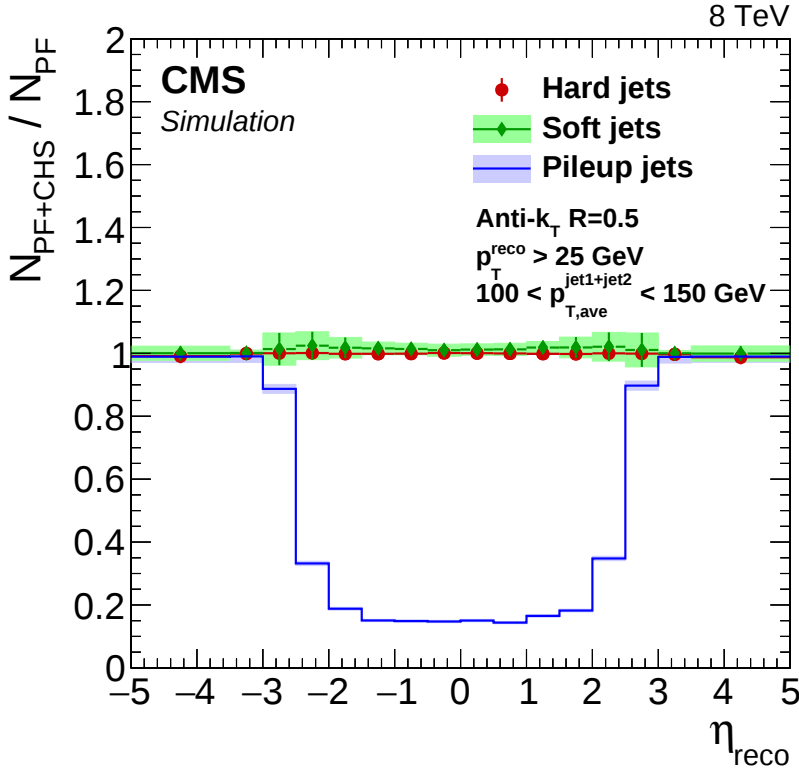


Figure 5.11. The ratio of the number of hard jets (red circles), soft jets (green diamonds) and pileup jets (blue line) as a function of jet η with (numerator) and without (denominator) charged hadron subtraction applied before jet clustering. The uncertainties drawn as lighter hue contain both statistical and systematic uncertainties. More details and definitions of the categories are given in the text. [58]

From the ratio of the number of PF+CHS and PF jets in events, shown in Figure 5.11, it is evident that the charged hadron subtraction process drastically reduces the number of unwanted pileup jets, as can be seen in the deep dip of the blue line showing the effect of CHS for pileup jets. For soft jets the numbers above $p_T = 25 \text{ GeV}$ are effectively the same on average, despite the fact that the amount of offset energy subtracted from jets by CHS and area-based offset subtraction are different from jet to jet. The small deviations from unity are within the relative jet energy scale uncertainties between PF and PF+CHS. At the pseudorapidity region $|\eta| < 2.4$, that is covered by the tracker, the reduction in pileup jets is 65-85%. At the forward region the number of pileup jets in the events is identical before and after the application of CHS, as there charged hadrons are not identified due to the lack of tracking information.

As important as the efficient reduction in pileup jets is the fact that

both hard and soft jets from the leading proton-proton interaction stay untouched by the CHS algorithm, as can be seen as the uniform behavior of the green diamonds (soft jets) and red circles (hard jets) in Figure 5.11.

In conclusion, the charged hadron subtraction process, that is now used in the majority of CMS physics analyses, yields significant improvements for jet measurements, here seen as the decrease in jet energy resolution and the efficient rejection of pileup jets. Once again, the CHS procedure is possible thanks to the particle-flow event reconstruction approach, which enables the identification and removal of charged particles originating from non-leading interaction vertices.

6. Dijet resonance search

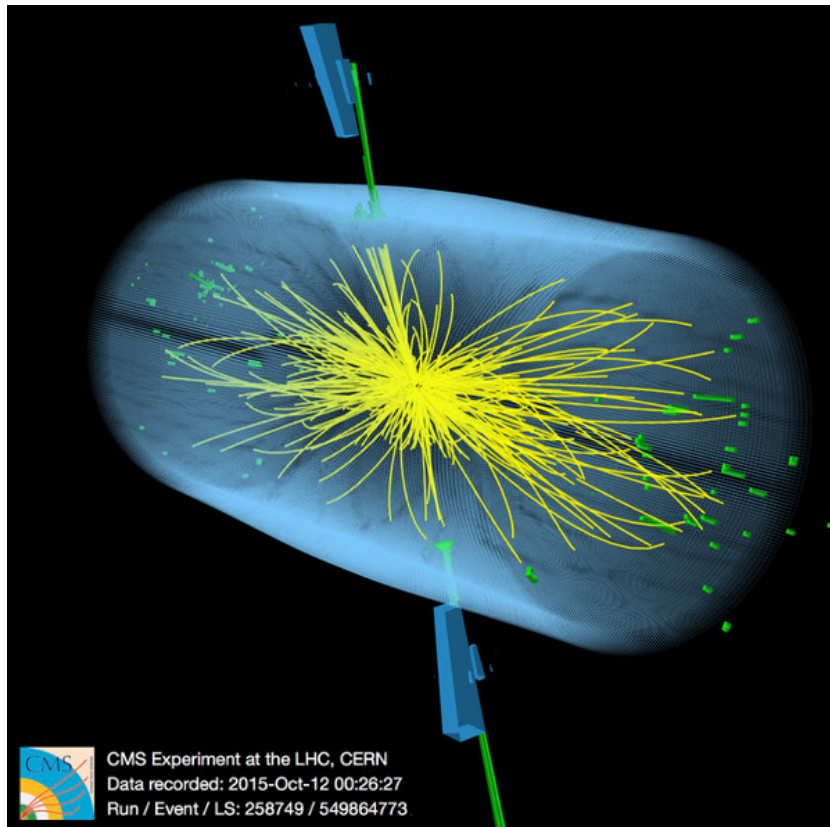


Figure 6.1. A reconstructed dijet event which was recorded with the CMS detector during a 13 TeV proton-proton collision run in October 2015. The yellow lines are reconstructed charged particles, ECAL energy deposits are shown in green and HCAL deposits as blue wedges. The event has a typical dijet topology with two high-energy jets measured at opposite directions.

Photo courtesy of the CMS Collaboration

As became evident in Section 2.4, despite its unprecedented success, the Standard Model leaves numerous questions unanswered and something beyond it must exist in nature. In a hadron collider the new particles that

we may produce are almost exclusively particles or *resonances*, which feel the strong force. This is because the dominant production mechanism, called the *s-channel production*, is the fusion or annihilation of two partons to an intermediary particle, which in most scenarios instantaneously decays to more stable particles. As the intermediary particle is produced from colored particles, it must also decay to colored particles, and thus the vast majority of reactions measured from the LHC proton-proton collisions involve strongly interacting particles. As we learned from QCD in Chapter 2, whatever is colored, will yield jets as the experimental footprint.

The dijet resonance search is an inclusive new physics search with exceptional discovery potential and unprecedented energy reach. The results of the analysis are among the most anticipated in the field of experimental particle physics, and provide invaluable input to theorists working with BSM physics models. The analysis produces the most stringent mass exclusion limits to various new physics models and it can also be used to search for new particles that could help in solving the mystery of dark matter.

The results presented in this chapter are either presented in or are supporting material for References [54, 59, 60].

6.1 Previous searches

New physics searches using the dijet event topology before the LHC era have been performed with proton-anti-proton collisions of the CERN $S\bar{p}pS$ and the Fermilab Tevatron colliders with center-of-mass energies of up to 0.63 and 1.96 TeV, respectively. The accessible dijet mass range for the UA1 and UA2 experiments at $S\bar{p}pS$ was roughly 50 to 300 GeV, and that of the CDF and D0 experiments at Tevatron around 0.2 to 1.3 TeV. [76]

Since the successful start of the LHC in 2010, many dijet searches with proton-proton collisions have been performed and published by both the ATLAS and CMS collaborations. The results with the $\sqrt{s} = 7$ TeV 2010–2011 LHC Run 1 data are published in four ATLAS papers [28, 29, 30, 31] and in three CMS papers [41, 44, 49], and the analyses are updated with the 2012–2013 $\sqrt{s} = 8$ TeV collision data in one publication by ATLAS [33] and two publications by CMS [50, 53]. With the 8 TeV data ATLAS and CMS were able to probe the dijet mass spectrum up to resonance masses of 5 TeV, with CMS catching a record event with invariant mass

Table 6.1. Summary of the observed 95% confidence level mass exclusion intervals or limits set to various beyond Standard Model (BSM) physics theories with the 8 TeV Run 1 data by the ATLAS and CMS collaborations. Both experiments do not set limits to all the same BSM models. [33, 53]

Model	Final state	ATLAS mass limit	CMS exclusion
String resonance	$q\bar{q}$	-	1.2 – 5.0 TeV
Excited quark	$q\bar{q}$	4.1 TeV	1.2 – 3.5 TeV
E_6 diquark	$q\bar{q}$	-	1.2 – 4.7 TeV
W' boson	$q\bar{q}$	2.5 TeV	1.2 – 1.9 TeV, 2.0 – 2.2 TeV
Chiral W^* boson:			
- leptophobic	$q\bar{q}$	1.75 TeV	-
- leptophilic	$q\bar{q}$	1.65 TeV	-
Z' boson	$q\bar{q}$	-	1.2 – 1.7 TeV
RS graviton	$q\bar{q}, g\bar{g}$	-	1.2 – 1.6 TeV
Axigluon/Coloron	$q\bar{q}$	-	1.3 – 3.6 TeV
Color-octet scalar	$g\bar{g}$	2.8 TeV	1.3 – 2.5 TeV
Quantum black holes	$q\bar{q}, g\bar{g}$	5.7 TeV	-

above 5 TeV.

Contemporaneously with the first $\sqrt{s} = 13$ TeV LHC Run 2 CMS dijet results, which are presented in this thesis, ATLAS published a paper with similar results and comparable exclusion limits [34].

The previous dijet resonance searches have not found evidence of resonant particle production in the mass ranges within the reach of the analyses, and the results are interpreted as mass limits for new particles predicted by BSM theories. The most stringent exclusion limits to various new physics models from 8 TeV LHC Run 1 data are summarized in Table 6.1.

6.2 Analysis strategy

Despite its extraordinary discovery potential, the basic analysis strategy of the dijet resonance search is remarkably simple. In short, we look for a bump in a smoothly falling spectrum, which gives the analysis strategy the nick name *bump hunt*. This is illustrated in Figure 6.2, and the Feynman diagram of the dijet resonance signal X is shown in Figure 6.3. Feynman graph for the dominant background process, QCD t-channel di-

jet production, is shown in Figure 6.4.

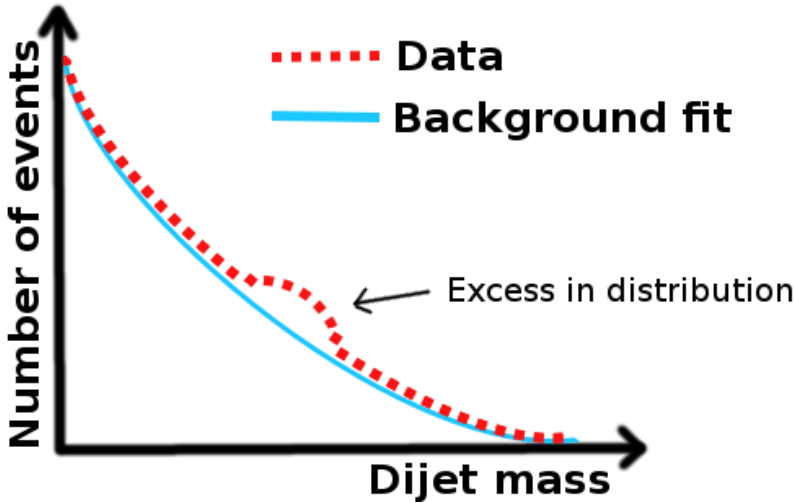


Figure 6.2. A sketch of the 'bump hunt' analysis strategy. If an excess is observed in the distribution and it has high enough statistical significance, it can be interpreted as a signal caused by a new massive resonance at that mass value. [87]

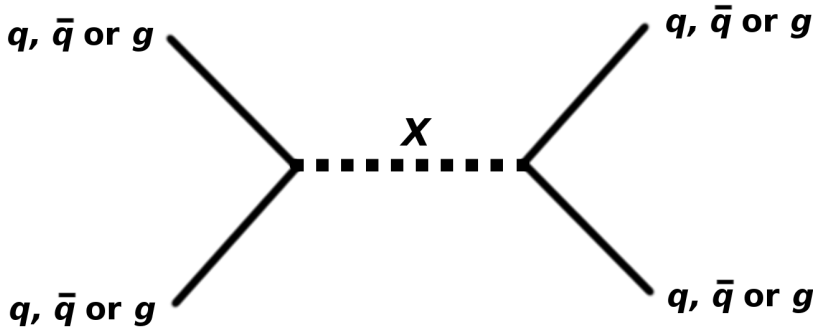


Figure 6.3. Feynman diagram of the s-channel dijet production where two partons (a quark and an anti-quark or two gluons) merge into a resonance state X , which then decays back to partons that yield back-to-back jets in the experiment. [87]

We are looking for any exotic *narrow* resonance X that decays hadronically, and we do it by looking for deviations from the SM prediction in the dijet mass spectrum. By narrow we mean that the intrinsic decay width of the resonance – a quantum mechanical property that every non-stable particle has – is negligible compared to the mass resolution that we are able to achieve.

According to the Standard Model and QCD, the spectrum of the invariant mass of two back-to-back jets falls smoothly with increasing dijet

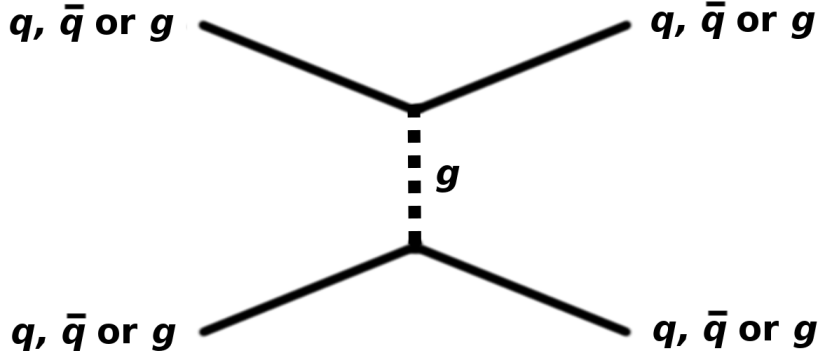


Figure 6.4. Feynman diagram of the t-channel QCD dijet production, where two partons scatter from each other producing more or less back-to-back jets in the experiment. [87]

mass. This prediction is also experimentally verified by *e.g.* ATLAS and CMS collaborations with their SM precision measurements of the dijet cross section in proton-proton collisions [32, 43]. As detector effects make it hard to reproduce the exact QCD spectrum that we should see in the detector, and also the uncertainties in QCD theory predictions are large, the QCD background is measured directly from the data by a background fit. The fit is a parameterization inspired by the QCD prediction for the dijet cross section. The background fit function is extended with one or two additional parameters for allowing enough flexibility for the detector effects, but it must be simple enough not to absorb possible excesses in the spectrum. We study the differences between the measured mass spectrum and the smooth background fit, and if there is a statistically significant excess, it can be interpreted as a signal of production of a new particle. If no excesses are observed, we set exclusion limits for models predicting heavy resonances in the dijet mass spectrum.

We probe as large an energy interval as possible for maximizing the possibility of finding a signal of new physics. At the low mass end the search is limited by restrictions related to triggering, and at the high mass end by the center-of-mass collision energy and the amount of energy carried by the initial partons, which is modeled by parton distribution functions.

6.2.1 Low-mass calo-scouting analysis

For extending the access to the sub-TeV mass range, alongside the classic high-mass analysis using full CMS particle-flow event reconstruction, we perform a *calo-scouting* analysis where we can probe resonance masses as

low as 600 GeV. This analysis exploits the novel CMS technique of *data scouting* [47], with which we can bypass the problem of heavily pre-scaled jet triggers by only storing an extremely compact version of each event. In practice, calo-scouting triggers record events which are reconstructed by the HLT, but would potentially be discarded in order to reduce event recording rates.

Where a normal fully reconstructed collision event takes roughly one megabyte of disk space, the events recorded by the scouting triggers require only 1-10 kilobytes per event, meaning a reduction with a factor of up to one thousand. To the extremely compressed events used in the calo-scouting dijet search, we store only anti- k_T $R = 0.4$ calo-jets, locations of the interaction vertices, event energy density ρ and event missing transverse energy, MET. MET is used mainly for data quality monitoring. Due to the reduced event size we can record events with up to five times higher rates, and thus are able to record data from un-prescaled jet triggers down to the 100-200 GeV range. Trigger pre-scaling is introduced in Section 3.2.8.

This low-mass extension was greatly motivated by the excess at 750 GeV observed by both ATLAS and CMS analyses in the diphoton mass spectrum in the 13 TeV LHC data recorded in 2015. As more data was collected and analyzed in 2016, the suspected effect vanished and turned out to be a mere statistical fluctuation, which happened to tantalizingly coincide in both experiments. [35, 55, 36, 56]

6.3 Event selection

Events are selected with the aim of minimizing the contribution of the overwhelming QCD dijet background and for maximizing the prominence of an excess. An example of an event with the dijet topology is shown in Figure 6.5.

Except for the calo-scouting analysis, the events are required to have at least one reconstructed proton-proton interaction vertex. The vertex with the highest sum of p_T^2 of the associated tracks is selected as the lead vertex and other vertices are classified as pileup interactions. At least two jets are required to originate from the lead vertex and the two with the highest transverse momenta are defined as leading jets. For ensuring a back-to-back topology, the angle between the leading jets must fulfill the condition $|\Delta\eta| \equiv |\Delta\eta_{jet1} - \Delta\eta_{jet2}| < 1.3$. This requirement is imposed for

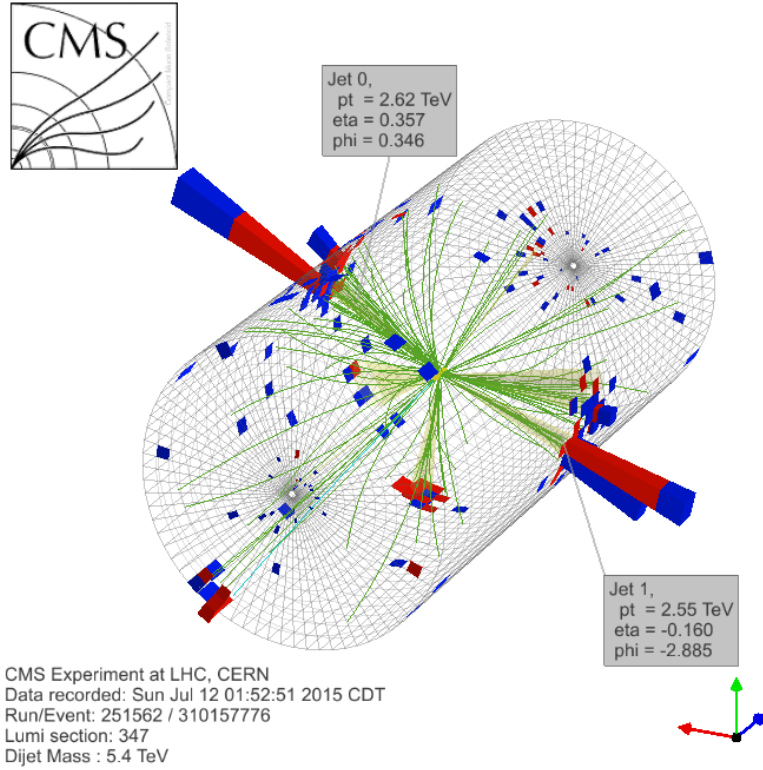


Figure 6.5. A collision event with a clear dijet event topology from the first LHC 13 TeV collision runs in 2015. [52]

boosting the signal-to-background ratio – a resonance X would decay producing a back-to-back topology in its rest frame, and thus nearly back-to-back topology in the detector, where a QCD background event has no such a preferential decay topology. The limit value of the $\Delta\eta$ -cut is carefully optimized in simulation studies to yield the best background reduction. An event display illustrating the 'delta-eta' cut is shown in Figure 6.6.

The leading jets are also required to be within $|\eta| < 2.5$, because at $|\eta| = 2.5$ the structure of the CMS detector changes significantly and jet measurements become less accurate. The pseudorapidity of the jet is measured from its centroid defined as the direction of the vector sum of its constituents, and thus parts of jets can slightly violate the $|\eta| < 2.5$ requirement, especially as we re-cluster the jets to $\Delta R = 1.1$ wide-jets, as explained in detail in Section 6.3.2.

A lower dijet mass threshold is also imposed on the events by the trigger turn on, as discussed in Section 6.4. We use only events from the mass

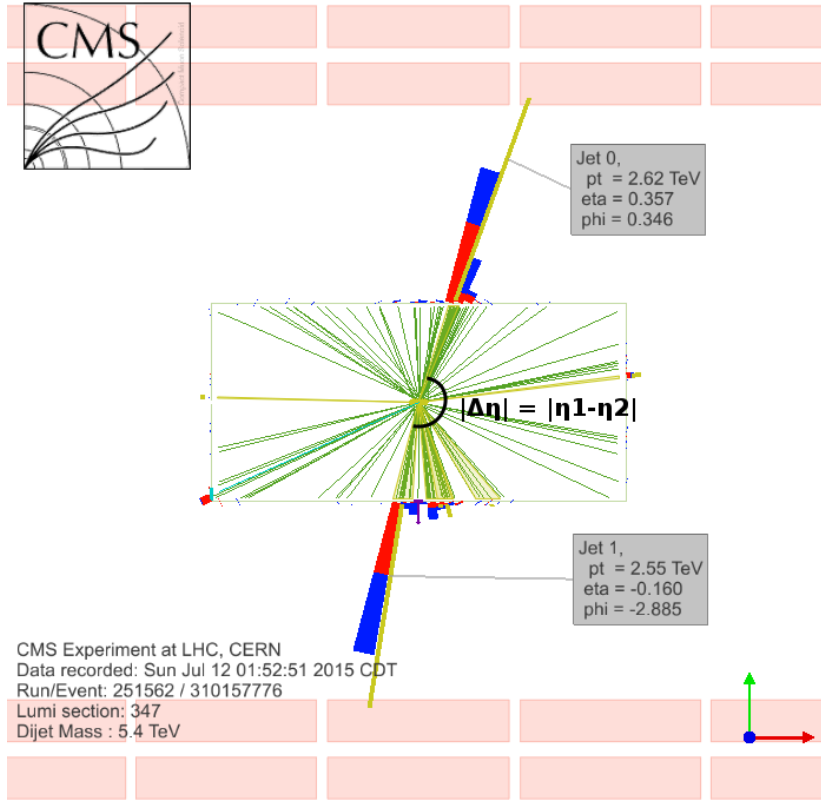


Figure 6.6. A collision event with the required dijet event topology from the first LHC 13 TeV collision runs in 2015. Both of the jets are well within the $|\eta| < 2.5$ pseudorapidity requirement and $|\Delta\eta| = |0.357 - (-0.160)| = 0.517 < 1.3$. [52]

region where the analysis triggers are measured to be fully efficient.

6.3.1 Jet selection

The jets are clustered from particle-flow candidates using the anti- k_T jet clustering algorithm with the distance parameter $\Delta R = 0.4$ after the event is cleaned from pileup with the charged hadron subtraction procedure, as described in Section 4.3. In the low-mass calo-scouting analysis the jets are reconstructed using calorimeter deposits at the HLT level, and CHS cannot be applied due to the absence of tracking information and PF candidates.

The leading jets are also subjected to jet identification (ID) criteria for filtering out spurious "fake" jets, which originate mostly from calorimeter or read-out electronics noise. The particle-flow jet ID is based on the jet energy composition described in Chapter 5. For a PF jet to be classified as

a good quality jet, it must have non-zero charged hadron fraction if within the fiducial tracker coverage $|\eta| < 2.4$, and neutral hadron, electron and photon fractions all less than 100%. These requirements are shown to efficiently filter out jets of spurious origin. [40]

In the case of calo-scouting analysis where the jet energy fractions cannot be reconstructed, the calo-jets are required to have at least 5% but no more than 95% of their energy deposited in the electromagnetic calorimeter. This ensures that jets originating entirely from noise in ECAL or HCAL are rejected.

If either of the leading jets do not fulfill the above criteria, the event is considered to be of low quality and it is discarded from the analysis.

Jet energy corrections are applied to the jets as described in Section 4.4. The corrections are derived using test beam measurements, simulations and 13 TeV collision data using *in-situ* techniques. Calo-jet corrections are derived separately using the available calorimeter information. A residual bias correction function is calculated by comparing the calo-jets to the well calibrated PF jets. After the bias correction the calo-jets have similar response with the particle-flow jets. The JEC uncertainties are also taken to be the same, because we have explicitly matched the average p_T of calo-jets and PF jets.

6.3.2 Wide-jet clustering

Following the event and jet selection the anti- k_t $\Delta R = 0.4$ PF jets are clustered into wide-jets of $\Delta R = 1.1$. This further increases the sensitivity of the analysis for isotropic decays of dijet resonances in the presence of background events from normal QCD dijet production. The main reason for this procedure is to include the jets originating from final state radiation emitted by energetic partons to the dijet mass calculation.

The wide-jet clustering is done by starting from the centroid of each of the leading jets and adding the four-momentum of any jet within $\Delta R = 1.1$ to the leading jets. The wide-jets do not require any extra corrections, because they are formed from already well calibrated PF- or calo-jets.

The wide-jet clustering technique is illustrated in Figure 6.7, where a gluon-jet from final state radiation (FSR) is captured by the wide-jet, but would be missed by the normal narrow PF- or calo-jets. As the a parton from a QCD background event is just as likely to emit FSR than a parton from a resonance decay, the wide-jet technique does not directly increase signal-to-background ratio, but the procedure enhances the mass reso-

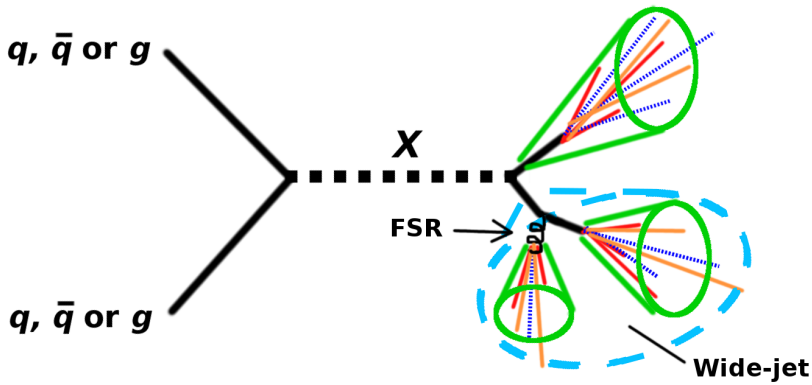


Figure 6.7. Illustration of the wide-jet clustering method. In this event sketch a resonance X decays to a pair of quarks, one of which emits a gluon in a final state radiation (FSR) process, before the partons hadronize into jets. The jet from the splitted gluon is captured by the wide-jet, but not by the normal narrow $\Delta R = 0.4$ jets, which are drawn in green. [87]

lution of the search, because now also the energy of the radiated gluon is included into the invariant dijet mass. Better mass resolution makes a possible resonance peak to spread less, which in turn increases signal separability. This can be seen in Figure 6.15 in Section 6.7, where the resonance shapes of gluon-gluon final states spread much more than the ones with quarks present.

6.4 Triggering

We trigger the events using the two staged L1+HLT trigger system, that is introduced in Section 3.2.8. At Level 1 we use triggers with loose jet requirements which then feed the selected potential jet events to the high-level trigger. At HLT we use both single-jet p_T triggers with a minimum p_T requirement, and jet H_T triggers that select events based on a threshold in H_T , the scalar sum of transverse momenta of all jets in the event. The jet H_T triggers only consider jets inside $|\eta| < 3$. More details of triggering is given later in this section.

6.4.1 Efficiency calculation

Trigger efficiencies are determined using triggers that are *orthogonal*, which means that they fire based on physics objects that are completely independent and uncorrelated to the jet-based triggers that we use in the

search. The orthogonal trigger of our choice for the high-mass analysis is a single muon trigger that capture events with at least one muon detected in the muon chambers.

By deriving the trigger *turn-on* curve, we can determine the mass value where the trigger becomes close to 100% efficient in firing on the type of collision event that it is designed for. The turn-on curve is computed by using the orthogonal trigger as a reference by measuring how often an analysis trigger and the reference trigger fire for the same event. In practice the used logical relation is:

$$\text{Jet trigger efficiency} = \frac{(\text{Jet trig. fired}) \text{ AND } (\text{Muon trig. fired})}{\text{Muon trigger fired}} \quad (6.1)$$

The logic of the measurement is that with the reference muon trigger we ensure that the event is recorded regardless of the jet activity in it, and by measuring at which dijet mass values the jet trigger and the non-related reference trigger fire simultaneously we obtain the jet trigger turn-on curve and can determine the efficiency threshold.

For determining the analysis lower mass threshold we determine the mass point where the measured trigger inefficiency is less than the statistical uncertainty of the number of events in the bin corresponding to the same dijet mass. As an example, let us consider a situation where we wish to measure whether the trigger is efficient at a bin which contains one million dijet events. This bin has a statistical uncertainty of $\frac{\sqrt{1000000}}{1000000} = 0.1\%$, so we would require the trigger efficiency to exceed 99.9% for the inefficiency to be covered by the statistical uncertainty.

In the high-mass analysis with 36 fb^{-1} of data, for example, the statistical uncertainty of the lowest included mass bin is 0.08%, and the trigger inefficiency in this bin is 0.05%.

6.4.2 High-mass analysis

The principal triggers used in the high-mass analysis are 'PFHT800' and 'PFHT900' -triggers, which are based on the H_T calculation performed with particle-flow. These triggers fire if the event jet activity H_T exceed 800 or 900 GeV, respectively. However, due to issues concerning the Level 1 triggers that feed events to the high-level PFHT triggers, a significant amount of inefficiency is present in part of the 2016 dataset. For this, three additional triggers unaffected by the PFHT problem are included for restoring a sharp trigger turn-on and keeping the analysis lower mass limit as low as possible. The three 'back-up' triggers are all single-jet

triggers with different jet p_T requirements. The first looks for PF jets with transverse momenta above 500 GeV, second fires for $p_T > 500$ calo-jets and the third reconstructs $\Delta R = 0.8$ anti- k_T PF jets and captures events with such jets with the energy cut at $p_T > 450$ GeV.

By combining events from all the five triggers we are able to restore trigger efficiency, and determine the turn-on to happen at the mass bin with the lower edge at 1246 GeV, as shown in Figure 6.8.

In the two previous searches which were performed with the 2015 and 2016 13 TeV data and are presented in References [54, 59] the high-mass analysis triggers were measured to reach full efficiency at 1.1 and 1.2 TeV, respectively.

6.4.3 Calo-scouting analysis

The low-mass calo-scouting analysis uses a trigger that calculates the event H_T from calo-jets with p_T 40 GeV and fires for $H_T > 250$ GeV. Unfortunately, single-jet triggers that would recover events lost for the faulty Level 1 H_T trigger do not exist for the calo-scouting analysis. Thus, we have to discard one fourth of the integrated luminosity of 2016, which leaves us 27 fb^{-1} of data.

The turn-on of the low-mass 'HT250' trigger is measured to happen at the dijet mass value of 490 GeV, as shown in Figure 6.9. In the earlier 13 TeV dijet analysis that included the low-mass search, the trigger efficiency and analysis lower mass threshold was at 450 GeV. Trigger efficiency thresholds vary between collision runs due to updates in trigger algorithms and changes trigger energy thresholds. The analyses thus cover well the interesting energy region of 750 TeV. [59]

6.5 Background fit

The background fit for the high-mass analysis is done with the parameterization

$$\frac{d\sigma}{dm_{jj}} = \frac{P_0(1-x)^{P_1}}{x^{P_2+P_3\ln(x)}}, \quad (6.2)$$

where the invariant dijet mass m_{jj} is scaled with the center-of-mass energy $\sqrt{s}$ as $x = m_{jj}/\sqrt{s}$ and $P_0 - P_3$ are four free parameters. Parameterization 6.2 is used for all three 13 TeV high-mass analyses and the first 13 TeV calo-scouting analysis that is published in Reference [59].

For the full 2016 calo-scouting analysis we use an extended fit function

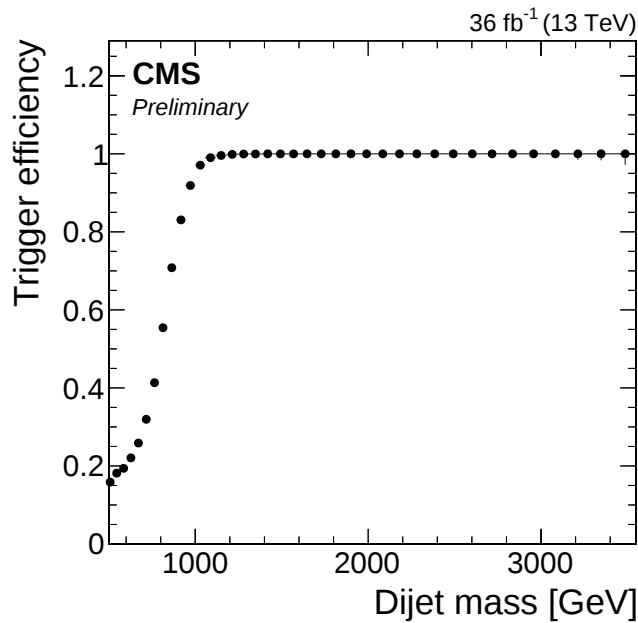


Figure 6.8. Trigger turn-on of the high-mass analysis measured with respect to a single-muon trigger with 45 GeV transverse momentum requirement. Full trigger efficiency is achieved at the dijet mass bin with the lower edge at 1246 GeV. [87]

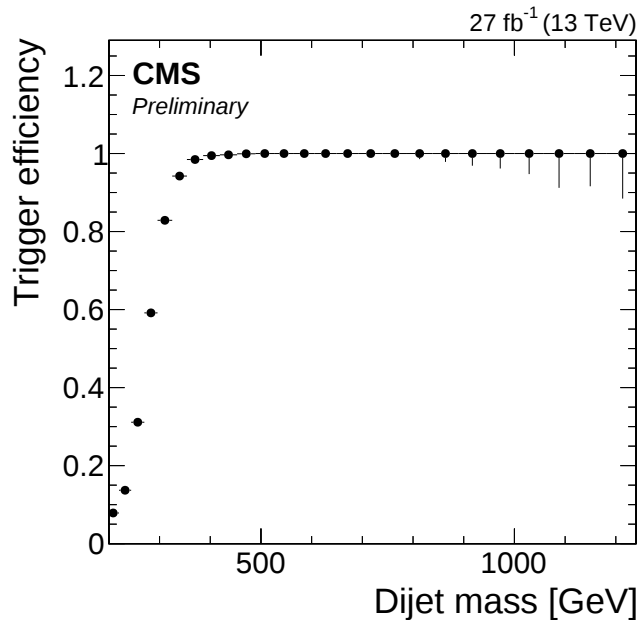


Figure 6.9. Trigger turn-on of the low-mass calo-scouting analysis measured with respect to a trigger that requires only that an event passes a jet H_T requirement at Level 1 trigger. Full trigger efficiency is achieved at the dijet mass bin with the lower edge at 490 GeV. [87]

which takes the form

$$\frac{d\sigma}{dm_{jj}} = \frac{P_0(1-x)^{P_1}}{x^{P_2+P_3\ln(x)+P_4\ln(x)^2}}, \quad (6.3)$$

where P_4 is a fifth free parameter that is needed in order to achieve a fit that models the background well. The added parameter is justified by the Fisher F-test [82], and also by extensive trials of alternative parameterizations, none of which produced better fit results.

For fitting the background parameterization to the observed dijet mass spectrum we use a maximum likelihood fit on a dijet mass histogram with narrow 1 GeV bins for approximating unbinned fit, for which experimental resolution does not cause problems.

6.6 Measured dijet mass spectra

The results of 13 TeV CMS dijet resonance searches are published in the three References [54, 59, 60] with integrated luminosities of 2.4, 12.9 and 36 fb⁻¹, respectively. The mass distributions are presented in the following sections for all three searches, but the interpretation of the results as exclusion limits is done only for the latest publication with highest integrated luminosity. In the first 13 TeV search only a high-mass analysis was performed. The results published in Reference [52] with the first 40 pb⁻¹ of 13 TeV data are not presented here due to the modest amount of integrated luminosity and thus limited physics impact of the search.

6.6.1 Results with 2.4 fb⁻¹ of integrated luminosity

The dijet mass spectrum of the high-mass analysis with 2.4 fb⁻¹ of integrated luminosity is shown in Figure 6.10. The spectrum is well modeled by the background parameterization and we observe no evidence of resonances. The dijet event with the highest invariant mass is observed at 6.1 TeV. In this search a QCD prediction is included in the figure, but it is shown for reference only and it is not used in the analysis. [54]

6.6.2 Results with 12.9 fb⁻¹ of integrated luminosity

The dijet mass spectrum of the low-mass analysis with 12.9 fb⁻¹ of integrated luminosity is shown in Figure 6.11 and the high-mass spectrum in Figure 6.12. The spectra are well modeled by the background parameterizations and we observe no evidence of resonant particle production in the low- or high-mass result, now covering as well the mass region from 0.5 to 8 TeV. The dijet event with the highest invariant mass is observed at 7.7 TeV. [59]

6.6.3 Results with 36 fb⁻¹ of integrated luminosity

The dijet mass spectrum of the low-mass analysis with 27 fb⁻¹ of integrated luminosity is shown in Figure 6.13 and the high-mass spectrum with the full 2016 36 fb⁻¹ dataset is presented in Figure 6.14. The spectra are well modeled by the background parameterizations and we observe no evidence of resonant particle production in the low- or high-mass analyses. The dijet event with the highest invariant mass is observed at 7.5 TeV, which is lower than the highest mass event in the 12.9 fb⁻¹ data (7.7

TeV), which is also a subset of the full 2016 data. This apparent inconsistency is possible due to updates in jet energy corrections, which can have up to 2–3% effects on the multi-TeV jets of the most energetic events. [59]

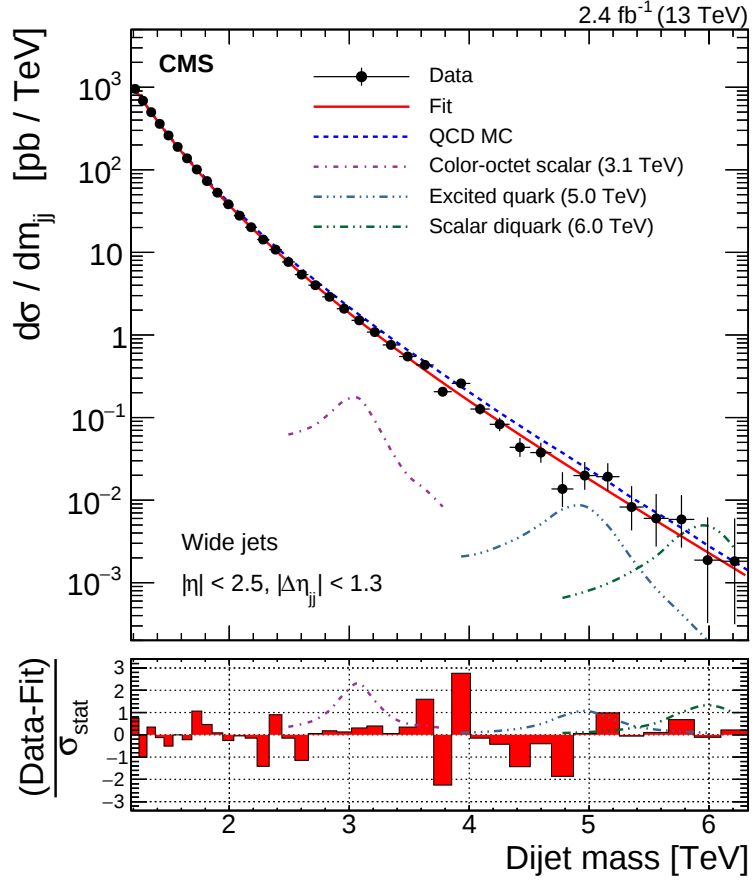


Figure 6.10. Measured dijet mass distribution (black points) compared to a fitted parameterization (red solid line) and a QCD prediction (bright blue dashed line). Dijet mass is on the x -axis and differential dijet cross section on the y -axis. The bottom panel shows the difference between the data and the fit, divided by the statistical uncertainty of the data. Three theory models, that are excluded with 95% confidence level, are drawn to the figure in violet, dark blue and green dash-dotted lines. [54]

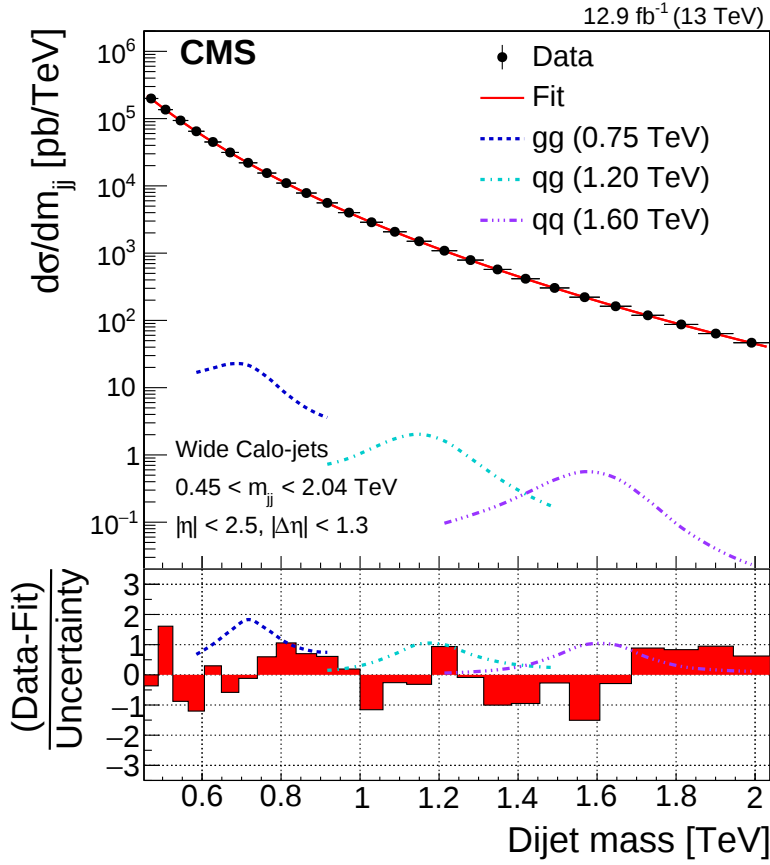


Figure 6.11. Measured dijet mass distribution (black points) of the calo-scouting analysis compared to a fitted parameterization (red solid line). Dijet mass is on the x -axis and differential dijet cross section on the y -axis. The bottom panel shows the difference between the data and the fit, divided by the statistical uncertainty of the data. Gluon-gluon (gg), quark-gluon (qg) and quark-quark resonances, that are excluded with 95% confidence level, are drawn to the figure in blue, cyan and violet dashed or dash-dotted lines. [59]

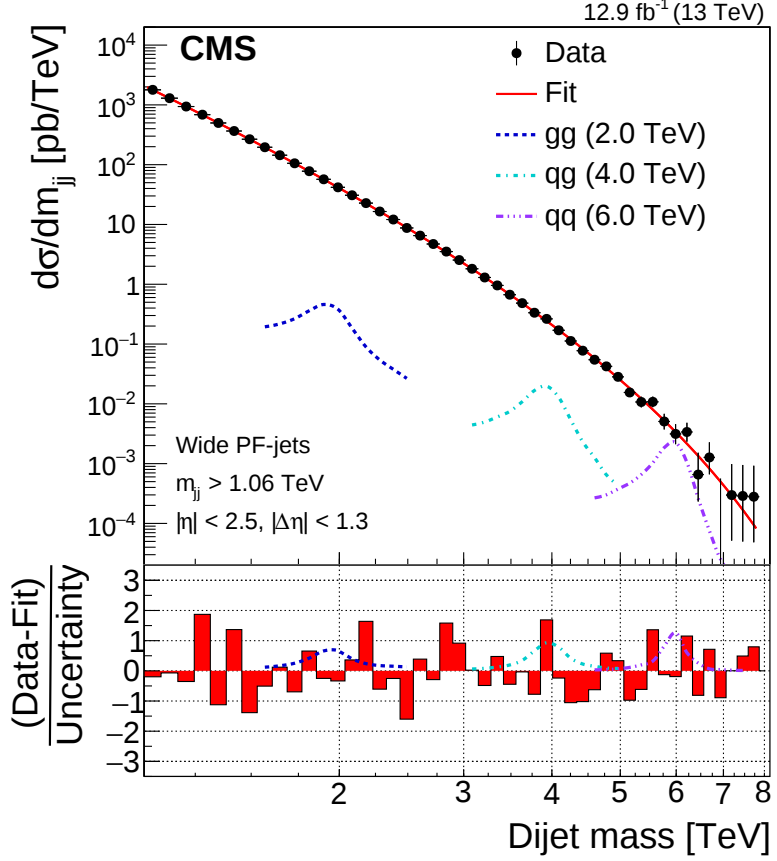


Figure 6.12. Measured dijet mass distribution (black points) of the high-mass analysis compared to a fitted parameterization (red solid line). Dijet mass is on the x -axis and differential dijet cross section on the y -axis. The bottom panel shows the difference between the data and the fit, divided by the statistical uncertainty of the data. Gluon-gluon (gg), quark-gluon (qg) and quark-quark resonances that are excluded with 95% confidence level are drawn to the figure in blue, cyan and violet dashed or dash-dotted lines. [59]

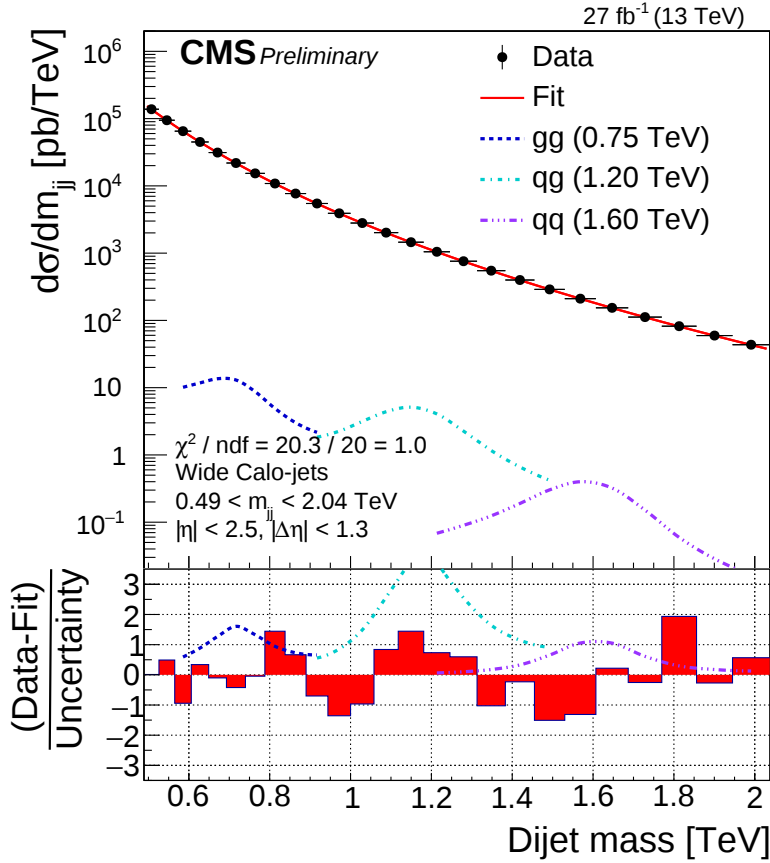


Figure 6.13. Measured dijet mass distribution (black points) of the low-mass analysis compared to a fitted parameterization (red solid line). Dijet mass is on the x -axis and differential dijet cross section on the y -axis. The bottom panel shows the difference between the data and the fit, divided by the statistical uncertainty of the data. Gluon-gluon (gg), quark-gluon (qg) and quark-quark resonances, that are excluded with 95% confidence level, are drawn to the figure in blue, cyan and violet dashed or dash-dotted lines. [60]

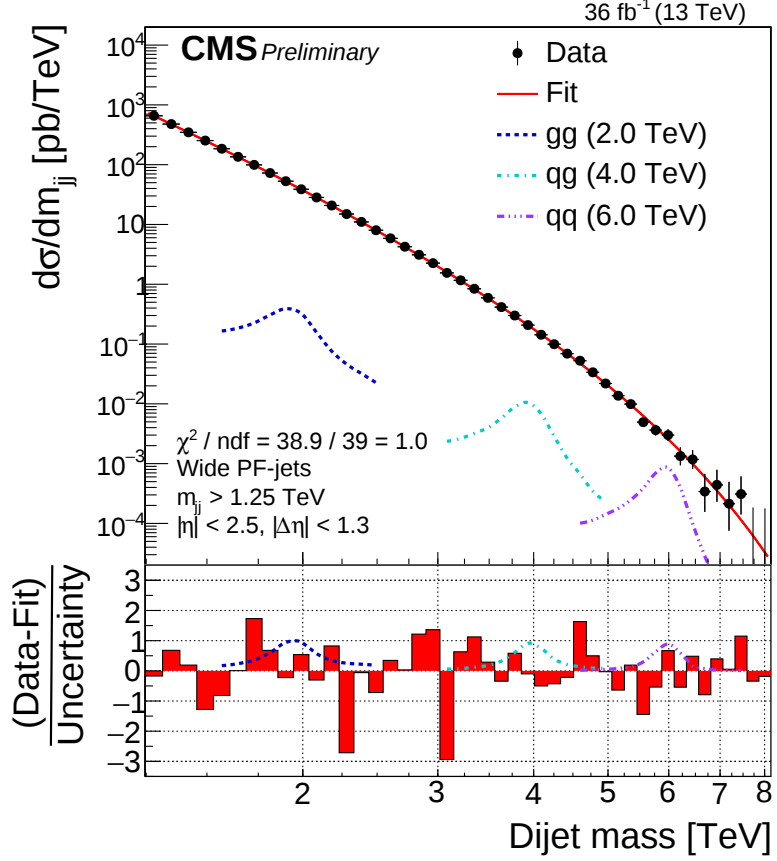


Figure 6.14. Measured dijet mass distribution (black points) of the high-mass analysis compared to a fitted parameterization (red solid line). Dijet mass is on the x -axis and differential dijet cross section on the y -axis. The bottom panel shows the difference between the data and the fit, divided by the statistical uncertainty of the data. Gluon-gluon (gg), quark-gluon (qg) and quark-quark resonances, that are excluded with 95% confidence level, are drawn to the figure in blue, cyan and violet dashed or dash-dotted lines. [60]

6.7 Signal models

For calculating exclusion limits we create simulation samples with general qq , qg and gg final states with various resonance masses. These samples are produced using PYTHIA8 event generator followed by the GEANT4 based detector simulation and full CMS event reconstruction. This ensures that the simulated signal shapes include all the experimental effects that cause energy and final state dependent spread and asymmetry, and thus represent what we would actually see in the spectrum in case of an exotic resonance present.

The signal shapes generated for the high-mass analysis are shown in Figure 6.15 for quark-quark, quark-gluon and gluon-gluon final states with resonance masses of 1, 3, 5 and 7 TeV. In the limit setting procedure signal shapes in between these mass values can be interpolated.

In Figure 6.15 we see that despite the underlying assumption of narrow resonances, the signal distribution widths vary from a few hundred GeV up to close to 1 TeV for the heaviest resonances due to experimental effects. Also, the spread of signal shapes depends heavily on the final state with quark-quark resonances producing the most narrow and gluon-gluon resonances the widest resonances, quark-gluon shapes falling in between. This is due to the differences in quark and gluon jets; quark jets are typically narrow cones while gluons radiate much more in the showering process, and thus produce wider jets with higher particle multiplicities.

6.8 Exclusion limits

As the background fit models the measured distributions well and significant pulls are not seen in the bottom panels of Figures 6.10-6.14, we interpret the result as observing no evidence of production of massive particles in the dijet mass distributions in Section 6.6. Thus, we proceed to calculating mass exclusion limits for new physics models. We calculate cross section limits for the production of resonance decays to qq , qg and gg final states and also interpret the results as mass exclusion limits for specific BSM models, including a dark matter interpretation.

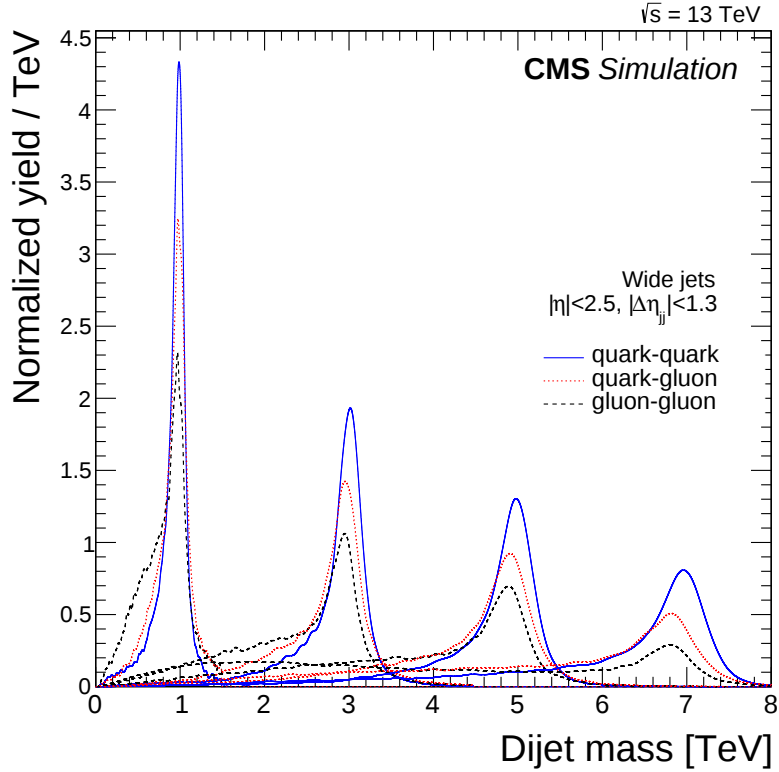


Figure 6.15. Signal shapes for the high-mass analysis from 1 to 7 TeV resonance masses. Quark-quark resonance shapes are drawn in blue solid line, quark-gluon shapes in red dotted line and gluon-gluon shapes in black dashed line. In event generation the quark-quark and gluon-gluon processes model decays of Randall-Sundrum gravitons and quark-gluon processes are modeled as decays of excited quarks. [54]

6.8.1 Limit calculation

We determine 95% confidence level exclusion limits for production cross sections using the modified frequentist method. [89, 78]

The limit calculation for a given signal is based on evaluation of the likelihood that the observed spectrum represents better a distribution with both signal and background events than a background only distribution. In other words, we quantify the level of incompatibility of the data with the hypothesis that the data contains a signal.

The first step is defining the test statistic q_μ with the signal strength μ and background and signal event yields b and s , respectively. The test statistic in the modified frequentist approach is a logarithmic ratio of signal+background and background only likelihoods:

$$q_\mu = -2 \ln \left(\frac{\mathcal{L}(\text{data}|\mu s + b)}{\mathcal{L}(\text{data}|\hat{\mu} s + b)} \right), \hat{\mu} \geq 0, \quad (6.4)$$

where $\hat{\mu}$ is the signal strength that maximizes the numerator likelihood $\mathcal{L}(\text{data}|\mu s + b)$.

In the modified frequentist (CLs) method the confidence of the background+signal hypothesis is defined with the observed data q_μ^{data} as

$$\text{CL}_{s+b} = P(q_\mu \geq q_\mu^{\text{data}} | \mu s + b), \quad (6.5)$$

which is computed from the observed distribution. Additionally, for evaluating the confidence of the background only hypothesis, defined as

$$\text{CL}_b = P(q_\mu \geq q_\mu^{\text{data}} | b), \quad (6.6)$$

a large number of pseudo-experiments are generated by random sampling of the background fit function. Finally, we calculate the modified frequentist confidence level variable CL_s as the ratio of the signal+background and background only probabilities:

$$\text{CL}_s = \frac{\text{CL}_{s+b}}{\text{CL}_b}. \quad (6.7)$$

The signal models with $\text{CL}_s \leq 0.05$ can be excluded with 95% confidence level. [9]

6.8.2 Limits for qq, qg and gg resonances

We calculate model-independent 95% confidence level CL_s limits for the product of cross section σ , branching fraction B and search acceptance A for qq, qg and gg final states. The observed and expected limits are shown for quark-quark resonances in Figure 6.16, for quark-gluon resonances in Figure 6.17 and for gluon-gluon resonances in Figure 6.18. Also one and two standard deviation uncertainty bands are calculated and shown for the expected cross section limits, and the fact that the observed limits stay roughly within the two standard deviation uncertainty indicates that no significant signals of new physics are present.

The qq and gg limits are weighted for corresponding the decays of Randall-Sundrum gravitons in Figure 6.19, and all the exclusion limits are summarized in the busy Figure 6.20. The datasets of the low- and high-mass analyses overlap in the mass range of 1.25 to 2.1 TeV, where the events of the 27 fb⁻¹ calo-scouting dataset are also found in the 36 fb⁻¹ high-mass analysis. While the low-mass analysis has worse mass resolution, both

datasets produce consistent limits in the overlapping range. We show calo-scouting limits up to 1.6 TeV to avoid edge effects in the high-mass limit calculations, and then switch sharply to high-mass analysis, whose expected limits are more stringent above 1.6 TeV.

The observed and expected lower mass limits which can be extracted from Figures 6.16–6.20 are listed in Table 6.2 together with limits from the two earlier 13 TeV CMS dijet searches and an 8 TeV CMS search for comparison.

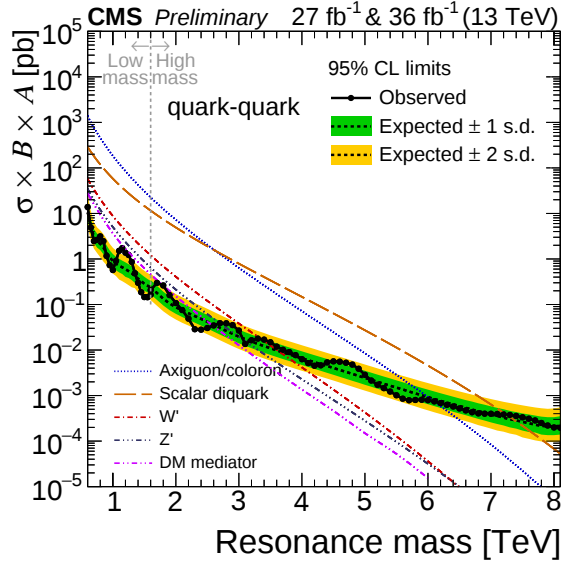


Figure 6.16. 95% confidence level observed (black dots) and expected (dashed black lines) limits as a function of mass for resonances decaying to two quarks. Uncertainty bands of one and two standard deviations (s.d) are shown for the expected limits in green and yellow, respectively. Five BSM models that the result constrains are shown in colorful dashed, dotted and dash-dotted lines. [60]

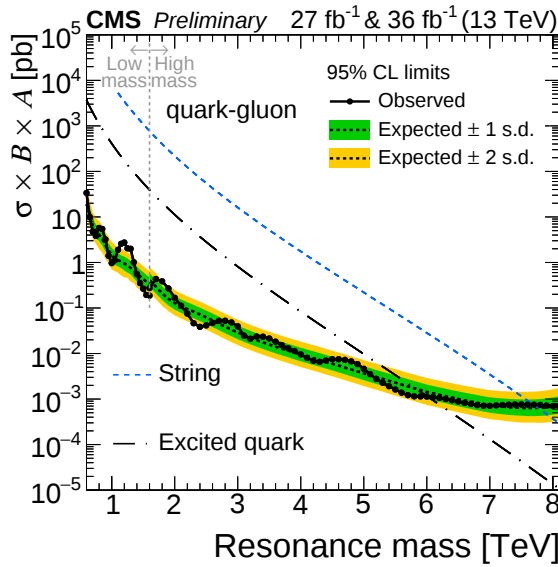


Figure 6.17. 95% confidence level observed (black dots) and expected (dashed black lines) limits as a function of mass for quark-gluon resonances. Uncertainty bands of one and two standard deviations (s.d) are shown for the expected limits in green and yellow, respectively. String and excited quark resonance models that the result restricts are shown in blue dashed and black dash-dotted lines, respectively. [60]

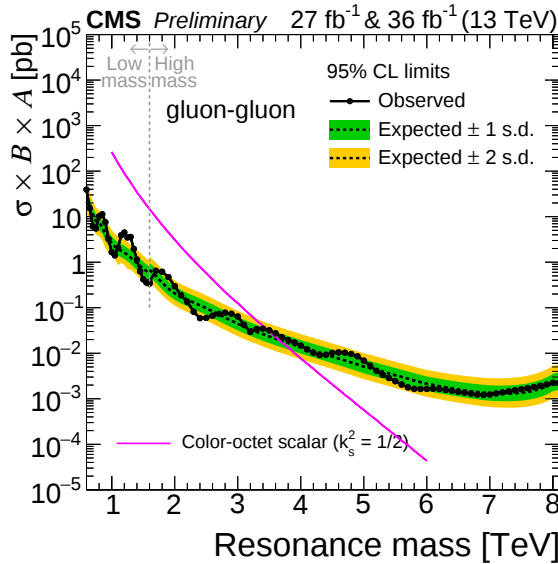


Figure 6.18. 95% confidence level observed (black dots) and expected (dashed black lines) limits as a function of mass for gluon-gluon resonances. Uncertainty bands of one and two standard deviations (s.d) are shown for the expected limits in green and yellow, respectively. The color-octet scalar resonance model that the result restricts is shown in magenta line. [60]

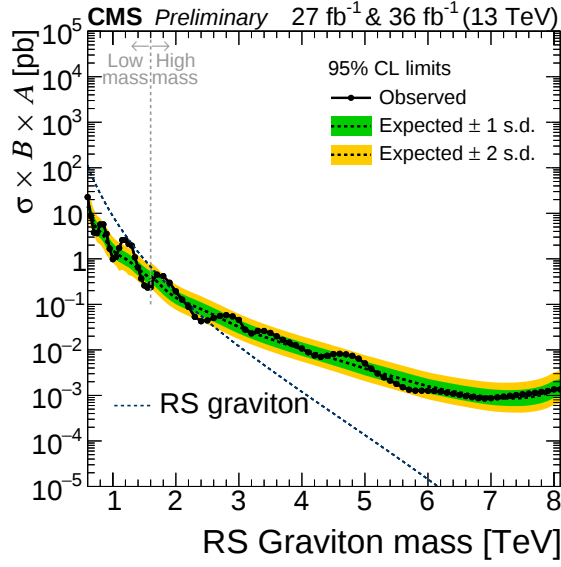


Figure 6.19. 95% confidence level observed (black dots) and expected (dashed black lines) limits as a function of mass for Randall-Sundrum (RS) gravitons. Uncertainty bands of one and two standard deviations (s.d) are shown for the expected limits in green and yellow, respectively. [60]

Table 6.2. Observed and expected mass upper limits at 95% confidence level from the 36 fb^{-1} search at $\sqrt{s} = 13 \text{ TeV}$, compared to limits of the 12.9 fb^{-1} and 2.4 fb^{-1} $\sqrt{s} = 13 \text{ TeV}$ analyses, and also to the LHC Run 1 analysis with 20 fb^{-1} at $\sqrt{s} = 8 \text{ TeV}$ [53]. When combined with the limits from the 27 fb^{-1} caloscouting analysis, the listed models are excluded between 0.6 TeV and the indicated mass limit by the full 2016 high-mass analysis, which gives the most stringent limits of this comparison. The limits assume parameter $k_s^2 = 1/2$ for color-octet scalars, $k/M_{\text{PL}} = 0.1$ for Randall-Sundrum (RS) gravitons, and $m_{\text{DM}} = 1 \text{ GeV}$ for dark matter (DM) mediators.

Model		Observed (expected) mass limit [TeV]				
		Final	36 fb^{-1}	12.9 fb^{-1}	2.4 fb^{-1}	20 fb^{-1}
		State	13 TeV	13 TeV	13 TeV	8 TeV
String	$q\bar{q}$		7.7 (7.7)	7.4 (7.4)	7.0 (6.9)	5.0 (4.9)
Scalar diquark	qq		7.2 (7.4)	6.9 (6.8)	6.0 (6.1)	4.7 (4.4)
Axigluon/coloron	$q\bar{q}$		6.1 (6.0)	5.5 (5.6)	5.1 (5.1)	3.7 (3.9)
Excited quark	$q\bar{q}$		6.0 (5.8)	5.4 (5.4)	5.0 (4.8)	3.5 (3.7)
Color-octet scalar	gg		3.4 (3.6)	3.0 (3.3)	—	—
W'	$q\bar{q}$		3.3 (3.6)	2.7 (3.1)	2.6 (2.3)	2.2 (2.2)
Z'	$q\bar{q}$		2.7 (2.9)	2.1 (2.3)	—	1.7 (1.8)
RS graviton	$q\bar{q}, gg$		1.7 (2.1)	1.9 (1.8)	—	1.6 (1.3)
DM mediator	$q\bar{q}$		2.6 (2.5)	2.0 (2.0)	—	—

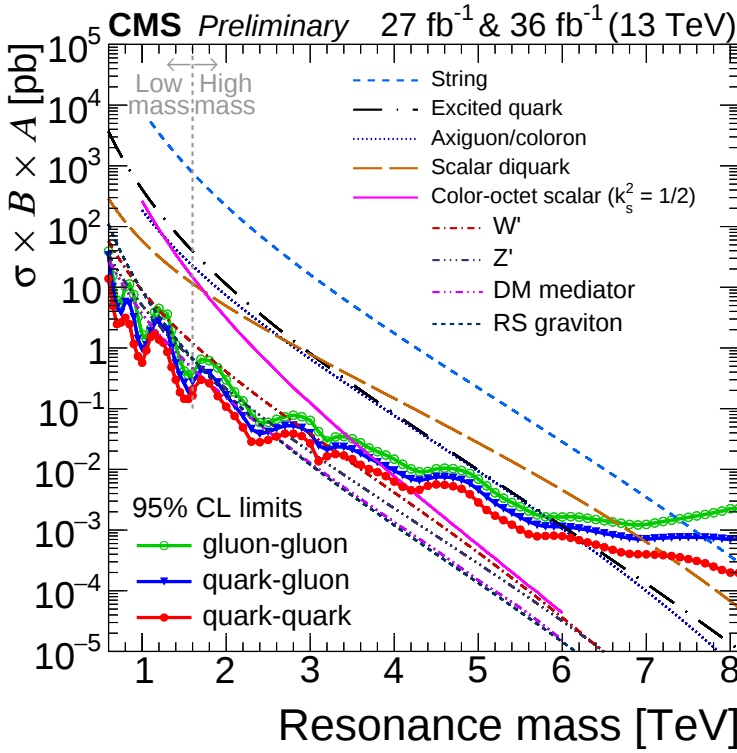


Figure 6.20. 95% confidence level exclusion limits for qq, qg and gg resonances combined with nine BSM theory models that the results constrain. [60]

6.8.3 Dark matter interpretation

In the dark matter interpretation we assume that dark matter consists of Dirac fermions with mass m_{DM} , and that there is also a dark matter mediator particle with mass M_{med} that couples to both DM and Standard Model particles. Furthermore, we assume the mediator to couple to quarks with the strength $g_q = 0.25$ and to dark matter with the strength $g_{DM} = 1.0$ for both axial-vector and vector mediators. More details about the possibilities and limitations of dark matter searches at the LHC can be found in Reference [1].

The observed and expected 95% confidence level exclusion regions in the m_{DM} – M_{med} plane for axial-vector DM mediator is shown in Figure 6.21 and for vector DM mediator in Figure 6.22. For comparison we show also constraints from the cosmological relic DM density based on astrophysical measurements from WMAP [96] and Planck [3], and MADDM 2.0.6 calculations [10, 11].

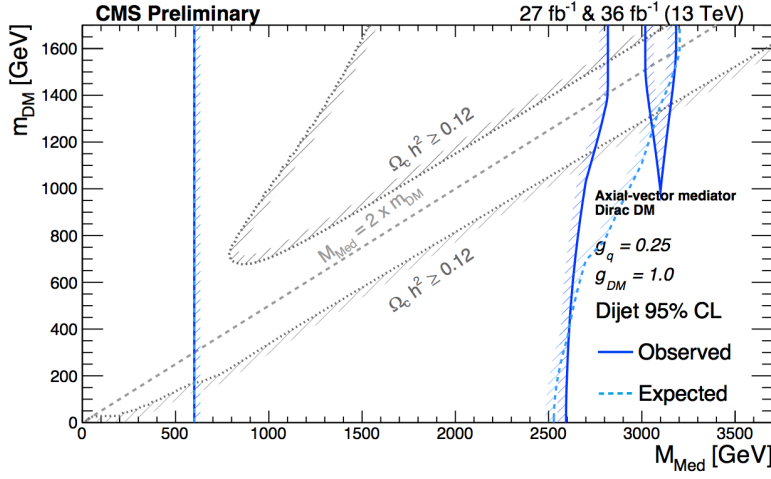


Figure 6.21. Observed (solid blue) and expected (dashed cyan) 95% confidence level exclusion regions for axial-vector dark matter mediator in the m_{DM} - M_{med} plane. The hatched regions are excluded. Dark matter constraints from cosmological observations are drawn in grey for comparison. [60]

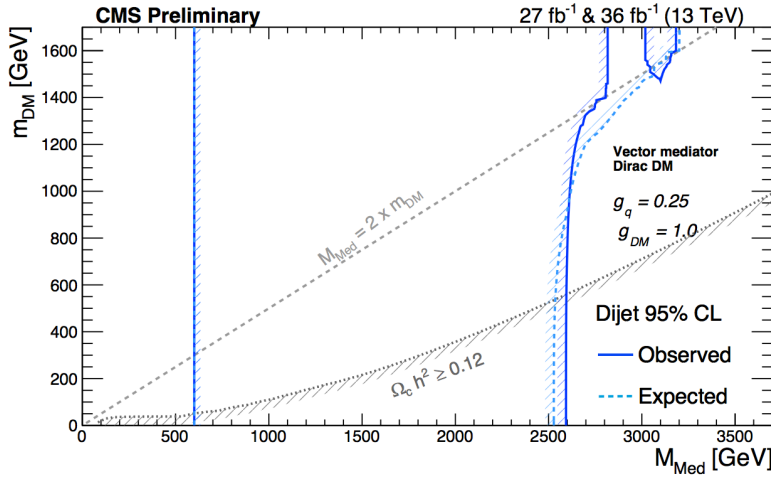


Figure 6.22. Observed (solid blue) and expected (dashed cyan) 95% confidence level exclusion regions for vector dark matter mediator in the m_{DM} - M_{med} plane. The hatched regions are excluded. Dark matter constraints from cosmological observations are drawn in grey for comparison. [60]

7. Summary

The central findings of this thesis are summarized in this chapter, followed by an outlook to the future operation of the LHC and prospects of particle physics in general.

7.1 Jet particology

The jet particology studies presented in Chapter 5 show the feasibility of in-depth jet particle substructure studies using the bottom-up jet reconstruction that the particle-flow algorithm allows us to perform. Measurements of jet energy composition are generally in per cent level agreement with the Monte Carlo simulations and GEANT4 detector modeling. But as we wish to push the jet measurements to highest possible precision, the small differences between data and simulation are important indications on where further developments are needed.

By comparing the jet energy composition with respect to jet p_T between the pure simulation in Figure 5.2 and measurement in Figure 5.4, we noticed that after accounting for the in-flight decays of short-lived neutral hadrons a residual difference remains. For understanding the origin of this roughly 2% discrepancy we would need to examine the particle-flow classification of neutral hadrons and photons closer, although this can also be an artefact of the fragmentation model used for predicting the jet particle spectrum. For testing the latter hypothesis, a similar study using simulations with other fragmentation models would be needed. Also the systematically increasing deviation between data and simulation in Figure 5.4 is an interesting phenomenon. This effect possibly arises from residual imperfections in the detector modeling. This assumption is supported by the preliminary and unpublished jet composition studies that are performed with 13 TeV data.

Jet energy composition with respect to jet pseudorapidity η and average event pileup μ in Figures 5.5 and 5.6 are within 1-2% agreement with simulation, with the differences mostly at the same level as statistical fluctuations. The non-trivial η -dependence of jet composition, however, is not a well understood effect and could be studied with careful examination of the reconstruction process and by repeating the pure simulation study of Figure 5.2 with respect to jet pseudorapidity.

The jet single pion response simulation study in Section 5.2 quantifies the robustness of particle-flow jets to detector miscalibrations and, shows how detector effects can be efficiently studied even with simulation samples with very modest event counts. The main limitation and place for improvement is the fact that the CMS fast simulation does not contain all the detector effects, as has been seen for example when studying jet energy composition with fast simulation in Reference [85]. The modest number of simulated events used in this study limits studying phenomena above 1 TeV jet transverse momenta, and a repetition of the study with a simulation sample of higher statistical power would be helpful for understanding jet behavior at the TeV scale, which is now experimentally accessible.

In Section 5.3 the charged hadron subtraction algorithm was shown to drastically reduce the number of pileup jets inside the tracker coverage. The advantages of charged hadron subtraction are already in use in the majority of CMS analyses. Even better pileup mitigation is in the works though, as the more advanced Pileup Per Particle Identification (PUPPI) algorithm is fast making its way as a trusted pileup reduction technique. Where CHS removes the charged particles associated with pileup vertices, PUPPI calculates for each particle a probability that it originates from a pileup vertex. [15]

7.2 Dijet resonance search

Despite using the highest collision energy combined to the highest integrated luminosity ever achieved in a hadron collider, we observe no evidence for the production of new massive particles in the dijet mass spectrum between 0.5 and 8 TeV. The overwhelming and irreducible QCD dijet background that the search is subject to makes it hard to find massive resonances from the dijet mass spectrum, especially if the resonance production cross section is small. Also the low mass resolution at high resonance

masses, as shown in Figure 6.15, brings more challenges for this search, as a possible signal can spread out so much that it is indistinguishable from the background distribution.

As summarized in Table 6.2, we set limits to nine massive resonances predicted by new physics theories. We exclude string resonances with masses below 7.7 TeV, scalar diquarks below 7.2 TeV, axigluons and colorons below 6.1 TeV, excited quarks below 6.0 TeV, color-octet scalars below 3.4 TeV, W' bosons below 3.3 TeV, Z' bosons below 2.7 TeV, Randall-Sundrum gravitons below 1.9 TeV and DM mediators below 2.6 TeV. These are the most stringent mass exclusion limits for these models to date.

Although the most exciting time of collecting the first inverse femtobarns at 13 TeV and doubling the dataset in every analysis iteration is now behind us, the dijet search will greatly benefit from the larger integrated luminosities that the LHC delivers in 2017–2018. Many of the probed BSM models are ruled out only at the 2–3 TeV resonance masses, and thus are still potential models for explaining the issues with SM. The dijet resonance search is being extended to also look for wide resonances and b-quark tagged jets. Both of these techniques significantly extend the discovery potential of the search.

7.3 Future of particle physics

As shown in Figure 7.1, the LHC Run II will continue until the end of 2018 and it is estimated to collect a dataset corresponding to up to 120 fb^{-1} , more than three times the integrated luminosity of the the 13 TeV results presented in this thesis. After the Long Shutdown 2 (LS2) from late 2018 until the end of 2019 the LHC Run III starts with the increased collision energy of $\sqrt{s} = 14 \text{ TeV}$ and collects data for three years, until the end of 2022. After Run III we are expected to have 300 fb^{-1} worth proton-proton collisions. The next step is the High-Luminosity LHC (HL-LHC) that is due to begin collecting data in 2025, after substantial upgrades are made to the accelerator and experiments for producing and surviving the much higher radiation rates. The HL-LHC phase continues for 10 years until 2035 and is expected to collect 3000 fb^{-1} .

While LHC keeps on delivering data, a suite of proposals for future colliders are being worked on. As the LHC is a discovery machine and is not well suited for high-precision measurements, for example of the Higgs properties, a linear electron-positron collider is proposed to be the next

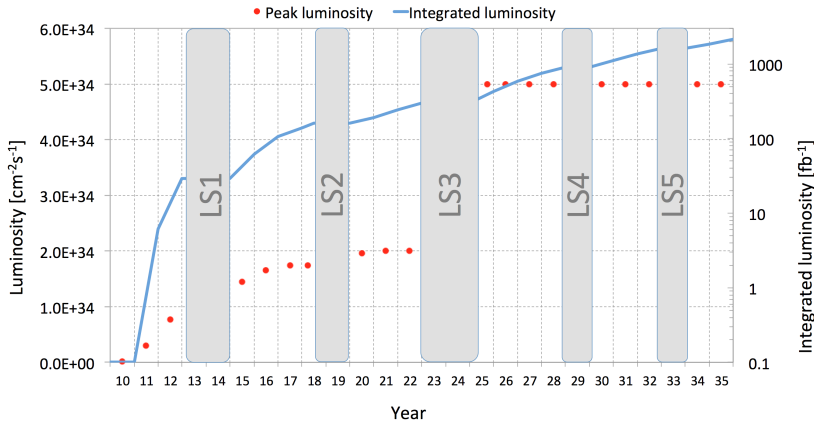


Figure 7.1. Operation schedule of the LHC from 2010 until 2035. [80]

big machine to be built. Two candidates exist: the International Linear Collider (ILC) [13] and Compact Linear Collider (CLIC) [4], both of which would bring our understanding of the Higgs boson and electro-weak symmetry breaking to a much higher level by for example measuring the Higgs couplings to SM particles at many times higher precision. This is the ultimate test for the SM, as many of the proposed new physics phenomena would modify Higgs couplings and we would see evidence of new particles even without directly discovering them.

Also building of new circular colliders is being planned. CERN studies the possibility to upgrade the LHC energy to 28 TeV in the High-Energy LHC (HE-LHC) scenario [83] and also has a courageous project called the Future Circular Collider (FCC) which would potentially produce e^+e^- , e^-p and pp collisions, the latter up to 100 TeV center-of-mass energies. The FCC is planned to have a circumference of 80 to 100 kilometers and it could follow the tradition and utilize the existing CERN accelerator chain for injection. [14] A similar project is being studied by Chinese physicists for a giant 80 kilometer e^+e^- synchrotron that goes with the name Circular Electron Positron Collider (CEPC) and would be later upgraded to SppC, Super proton proton Collider. [74] This project is thus nearly identical with CERN's FCC plans.

It goes without saying that in the turbulent state of affairs in world politics, realization of any of these proposals would be a great victory for particle physics. For the LHC the physics case was strong; it was clear from irrefutable theoretical arguments related to WW-scattering that either the Higgs or something new breaking the SM would be found. We don't have such an argument to give to the funding agencies for the fu-

ture colliders, at least not yet. What we do have is an excellently performing hadron collider and excellently performing detectors, which are run by strong collaborations. We are only in the beginning of exploring the vast energy scales that the 13–14 TeV collisions and the unprecedented integrated luminosities of the LHC enable us to study. Pessimists will say that new physics is only at the Planck scale, but I am an optimist and believe that the next great discovery is right behind the corner.

References

- [1] Daniel Abercrombie et al. Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum. 2015. arXiv:1507.00966.
- [2] P. A. R. Ade et al. Planck 2013 results. I. Overview of products and scientific results. *Astron. Astrophys.*, 571:A1, 2014.
- [3] P. A. R. Ade et al. Planck 2013 results. XVI. Cosmological parameters. *Astron. Astrophys.*, 571:A16, 2014.
- [4] M Aicheler, P Burrows, M Draper, T Garvey, P Lebrun, K Peach, N Phinney, H Schmickler, D Schulte, and N Toge. A Multi-TeV Linear Collider Based on CLIC Technology. 2012. <http://cds.cern.ch/record/1500095>.
- [5] Guido Altarelli. The Standard model of particle physics. 2005. arXiv:hep-ph/0510281v1.
- [6] Luis A. Anchordoqui, Haim Goldberg, Dieter Lust, Satoshi Nawata, Stephan Stieberger, and Tomasz R. Taylor. Dijet signals for low mass strings at the LHC. *Phys. Rev. Lett.*, 101:241803, 2008.
- [7] Ryan Atkin. Review of jet reconstruction algorithms. *J. Phys. Conf. Ser.*, 645(1):012008, 2015.
- [8] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012.
- [9] ATLAS Collaboration, CMS Collaboration and the LHC Higgs Combination Group. Procedure for the LHC Higgs boson search combination in Summer 2011. Technical Report CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, Aug 2011. <https://cds.cern.ch/record/1379837>.
- [10] Mihailo Backovic, Kyoungchul Kong, and Mathew McCaskey. MadDM v.1.0: Computation of Dark Matter Relic Abundance Using MadGraph5. *Physics of the Dark Universe*, 5-6:18–28, 2014.
- [11] Mihailo Backović, Antony Martini, Olivier Mattelaer, Kyoungchul Kong, and Gopolang Mohlabeng. Direct Detection of Dark Matter with MadDM v.2.0. *Phys. Dark Univ.*, 9-10:37–50, 2015.
- [12] U. Baur, I. Hinchliffe, and D. Zeppenfeld. Excited Quark Production at Hadron Colliders. *Int. J. Mod. Phys.*, A2:1285, 1987.

- [13] Ties Behnke, James E. Brau, Brian Foster, Juan Fuster, Mike Harrison, James McEwan Paterson, Michael Peskin, Marcel Stanitzki, Nicholas Walker, and Hitoshi Yamamoto. The International Linear Collider Technical Design Report - Volume 1: Executive Summary. 2013.
- [14] Michael Benedikt, B. Goddard, Daniel Schulte, F. Zimmermann, and M. J. Syphers. FCC-hh Hadron Collider - Parameter Scenarios and Staging Options. In *Proceedings, 6th International Particle Accelerator Conference (IPAC 2015): Richmond, Virginia, USA, May 3-8, 2015*, page TUPTY062, 2015.
- [15] Daniele Bertolini, Philip Harris, Matthew Low, and Nhan Tran. Pileup Per Particle Identification. *JHEP*, 10:059, 2014.
- [16] Biplob Bhattacharjee, Satyanarayan Mukhopadhyay, Mihoko M. Nojiri, Yasuhito Sakaki, and Bryan R. Webber. Quark-gluon discrimination in the search for gluino pair production at the LHC. *JHEP*, 01:044, 2017.
- [17] Giorgio Busoni et al. Recommendations on presenting LHC searches for missing transverse energy signals using simplified s -channel models of dark matter. 2016.
- [18] C. Buttar et al. Standard Model Handles and Candles Working Group: Tools and Jets Summary Report. In *Physics at TeV colliders, La physique du TeV aux collisionneurs, Les Houches 2007 : 11-29 June 2007*, pages 121–214, 2008.
- [19] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti- k_t jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063, 2008.
- [20] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet User Manual. *Eur. Phys. J.*, C72:1896, 2012.
- [21] F. Capozzi, E. Lisi, A. Marrone, D. Montanino, and A. Palazzo. Neutrino masses and mixings: Status of known and unknown 3ν parameters. *Nucl. Phys.*, B908:218–234, 2016.
- [22] Mikael Chala, Felix Kahlhoefer, Matthew McCullough, Germano Nardini, and Kai Schmidt-Hoberg. Constraining dark sectors with monojets and dijets. *Journal of High Energy Physics*, 2015(7):89, Jul 2015.
- [23] CMS Collaboration. "The CMS experiment at the CERN LHC". *JINST*, 03:S08004, 2008. <https://cds.cern.ch/record/1129810>.
- [24] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012.
- [25] ALEPH Collaboration. Performance of the ALEPH detector at LEP. *Nucl. Instrum. Meth.*, A360:481–506, 1995.
- [26] ALICE Collaboration. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008.
- [27] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

- [28] ATLAS Collaboration. Search for New Particles in Two-Jet Final States in 7 TeV Proton-Proton Collisions with the ATLAS Detector at the LHC. *Phys. Rev. Lett.*, 105:161801, 2010.
- [29] ATLAS Collaboration. Search for New Physics in Dijet Mass and Angular Distributions in pp Collisions at $\sqrt{s} = 7$ TeV Measured with the ATLAS Detector. *New J. Phys.*, 13:053044, 2011.
- [30] ATLAS Collaboration. Search for New Physics in the Dijet Mass Distribution using 1 fb^{-1} of pp Collision Data at $\sqrt{s} = 7$ TeV collected by the ATLAS Detector. *Phys. Lett.*, B708:37–54, 2012.
- [31] ATLAS Collaboration. ATLAS search for new phenomena in dijet mass and angular distributions using pp collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 01:029, 2013.
- [32] ATLAS Collaboration. Measurement of dijet cross sections in pp collisions at 7 TeV centre-of-mass energy using the ATLAS detector. *JHEP*, 05:059, 2014.
- [33] ATLAS Collaboration. Search for new phenomena in the dijet mass distribution using $p - p$ collision data at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Phys. Rev.*, D91(5):052007, 2015.
- [34] ATLAS Collaboration. Search for new phenomena in dijet mass and angular distributions from pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Lett.*, B754:302–322, 2016.
- [35] ATLAS Collaboration. Search for resonances in diphoton events at $\sqrt{s}=13$ TeV with the ATLAS detector. *JHEP*, 09:001, 2016.
- [36] ATLAS collaboration. Search for scalar diphoton resonances with 15.4 fb^{-1} of data collected at $\sqrt{s}=13$ TeV in 2015 and 2016 with the ATLAS detector. 2016.
- [37] ATLAS Collaboration. Jet reconstruction and performance using particle flow with the ATLAS Detector. 2017.
- [38] CMS Collaboration. Energy resolution of the barrel of the CMS electromagnetic calorimeter. *JINST*, 2:P04004, 2007.
- [39] CMS Collaboration. CMS Tracking Performance Results from early LHC Operation. *Eur. Phys. J.*, C70:1165–1192, 2010.
- [40] CMS Collaboration. Jet Performance in pp Collisions at 7 TeV. 2010.
- [41] CMS Collaboration. Search for Dijet Resonances in 7 TeV pp Collisions at CMS. *Phys. Rev. Lett.*, 105:211801, 2010.
- [42] CMS Collaboration. Single-Particle Response in the CMS Calorimeters. Technical Report CMS-PAS-JME-10-008, CERN, Geneva, 2010. <http://cds.cern.ch/record/1279141>.
- [43] CMS Collaboration. Measurement of the differential dijet production cross section in proton-proton collisions at $\sqrt{s} = 7$ TeV. *Phys. Lett.*, B700:187–206, 2011.

- [44] CMS Collaboration. Search for Resonances in the Dijet Mass Spectrum from 7 TeV pp Collisions at CMS. *Phys. Lett.*, B704:123–142, 2011.
- [45] CMS Collaboration. Technical proposal for the upgrade of the CMS detector through 2020. Technical Report CERN-LHCC-2011-006. LHCC-P-004, Jun 2011. <https://cds.cern.ch/record/1355706>.
- [46] CMS Collaboration. Data Parking and Data Scouting at the CMS Experiment. Sep 2012. <https://cds.cern.ch/record/1480607>.
- [47] CMS Collaboration. Data Parking and Data Scouting at the CMS Experiment. 2012. <http://cds.cern.ch/record/1480607>.
- [48] CMS Collaboration. Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV. *JINST*, 7:P10002, 2012.
- [49] CMS Collaboration. Search for narrow resonances and quantum black holes in inclusive and b -tagged dijet mass spectra from pp collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 01:013, 2013.
- [50] CMS Collaboration. Search for narrow resonances using the dijet mass spectrum in pp collisions at $\sqrt{s} = 8$ TeV. *Phys. Rev.*, D87(11):114015, 2013.
- [51] CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014.
- [52] CMS Collaboration. Search for Narrow Resonances using the Dijet Mass Spectrum with 40 pb^{-1} of pp Collisions at $\sqrt{s} = 13$ TeV. 2015.
- [53] CMS Collaboration. Search for resonances and quantum black holes using dijet mass spectra in proton-proton collisions at $\sqrt{s} = 8$ TeV. *Phys. Rev.*, D91(5):052009, 2015.
- [54] CMS Collaboration. Search for narrow resonances decaying to dijets in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 116(7):071801, 2016.
- [55] CMS Collaboration. Search for Resonant Production of High-Mass Photon Pairs in Proton-Proton Collisions at $\sqrt{s} = 8$ and 13 TeV. *Phys. Rev. Lett.*, 117(5):051802, 2016.
- [56] CMS Collaboration. Search for resonant production of high mass photon pairs using 12.9 fb^{-1} of proton-proton collisions at $\sqrt{s} = 13$ TeV and combined interpretation of searches at 8 and 13 TeV. 2016.
- [57] CMS Collaboration. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *JINST*, 12(02):P02014, 2017.
- [58] CMS Collaboration. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10):P10003, 2017.
- [59] CMS Collaboration. Search for dijet resonances in proton–proton collisions at $\sqrt{s} = 13$ TeV and constraints on dark matter and other models. *Phys. Lett.*, B769:520–542, 2017.
- [60] CMS Collaboration. Searches for dijet resonances in pp collisions at $\sqrt{s} = 13$ TeV using data collected in 2016. 2017. <http://cds.cern.ch/record/2256873>.

- [61] CMS ECAL/HCAL Collaboration. The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c. *Eur. Phys. J.*, C60:359–373, 2009. [Erratum: *Eur. Phys. J.*C61,353(2009)].
- [62] LHCb Collaboration. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008.
- [63] Edvige Corbelli and Paolo Salucci. The Extended Rotation Curve and the Dark Matter Halo of M33. *Mon. Not. Roy. Astron. Soc.*, 311:441–447, 2000.
- [64] Schuyler Cullen, Maxim Perelstein, and Michael E. Peskin. TeV strings and collider probes of large extra dimensions. *Phys. Rev.*, D62:055012, 2000.
- [65] Brian Dorney. Anatomy of a jet in cms. <http://www.quantumdiaries.org/2011/06/01/anatomy-of-a-jet-in-cms/>, Accessed 4.7.2017.
- [66] E. Eichten, I. Hinchliffe, Kenneth D. Lane, and C. Quigg. Super Collider Physics. *Rev. Mod. Phys.*, 56:579–707, 1984. [Addendum: *Rev. Mod. Phys.*58,1065(1986)].
- [67] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Physical Review Letters*, 13:321–323, August 1964.
- [68] J. Abdallah et al. Simplified models for dark matter searches at the LHC. *Physics of the Dark Universe*, 9–10:8 – 23, 2015.
- [69] Lyndon Evans and Philip Bryant. "LHC Machine". *JINST*, 03:S08001, 2008. <https://cds.cern.ch/record/2103617?ln=en>.
- [70] European Organisation for Nuclear Research. CERN Document Server. <http://cdsweb.cern.ch/>.
- [71] G.F. Giudice. *A Zeptospace Odyssey: A Journey Into the Physics of the LHC*. Oxford University Press, 2009.
- [72] S. L. Glashow. Quarks with Color and Flavor. *Sci. Am.*, 233N4:38–50, 1975. [*Sci. Am.*233,38(1975)].
- [73] V. Gribov and B. Pontecorvo. Neutrino astronomy and lepton charge. *Physics Letters B*, 28(7):493 – 496, 1969.
- [74] CEPC-SPPC Study Group. CEPC-SPPC Preliminary Conceptual Design Report. 1. Physics and Detector. 2015.
- [75] G. S. Guralnik, C. R. Hagen, and T. W. Kibble. Global Conservation Laws and Massless Particles. *Physical Review Letters*, 13:585–587, November 1964.
- [76] Robert M. Harris and Konstantinos Kousouris. Searches for Dijet Resonances at Hadron Colliders. *Int. J. Mod. Phys.*, A26:5005–5055, 2011.
- [77] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys.Rev.Lett.*, 13:508–509, 1964.
- [78] Thomas Junk. Confidence level computation for combining searches with small statistics. *Nucl. Instrum. Meth.*, A434:435–443, 1999.

- [79] Jan Kretzschmar. Searches for extra dimensions with the atlas and cms detectors. *Nuclear and Particle Physics Proceedings*, 273-275(Supplement C):541 – 545, 2016. 37th International Conference on High Energy Physics (ICHEP).
- [80] Mike Lamont. LHC 10 year schedule. <http://lhc-commissioning.web.cern.ch/lhc-commissioning/schedule/LHC-schedule-update.pdf>, Accessed 7.9.2017.
- [81] Jianglai Liu, Xun Chen, and Xiangdong Ji. Current status of direct dark matter detection experiments. *Nature Phys.*, 13(3):212–216, 2017.
- [82] R. G. Lomax and D. L. Hahs-Vaughn. *Statistical concepts: A second course*. Routledge Academic, London, 2007.
- [83] M. Myers. CERN Accelerator Strategy. In *Proceedings, EuCARD-AccNet-EuroLumi Workshop: The High-Energy Large Hadron Collider (HE-LHC10): Villa Bighi, Malta, Republic of Malta, October 14-16, 2010*, 2011.
- [84] K. A. Olive et al. Review of Particle Physics. *Chin. Phys.*, C38:090001, 2014.
- [85] Juska Pekkanen. Jets in CMS Experiment. 2013. Diploma Thesis, Aalto University School of Science.
- [86] Juska Pekkanen. Photo courtesy of the author, 2017.
- [87] Juska Pekkanen. Private production for illustration purposes only, 2017.
- [88] Lisa Randall and Raman Sundrum. A Large mass hierarchy from a small extra dimension. *Phys. Rev. Lett.*, 83:3370–3373, 1999.
- [89] A L Read. Presentation of search results: the cls technique. *Journal of Physics G: Nuclear and Particle Physics*, 28(10):2693, 2002.
- [90] 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):250 – 303, 2003.
- [91] A. D. Sakharov. Violation of CP Invariance, c Asymmetry, and Baryon Asymmetry of the Universe. *Pisma Zh. Eksp. Teor. Fiz.*, 5:32–35, 1967. [Usp. Fiz. Nauk161,61(1991)].
- [92] Philipp Schieferdecker. Presentation slides, Vivian’s Meeting 17.4.2009. https://twiki.cern.ch/twiki/bin/viewfile/Sandbox/Lecture?rev=1;filename=Philipp_Schieferdeckers_Lecture.pdf, Accessed 10.1.2017.
- [93] Sezen Sekmen. Recent Developments in CMS Fast Simulation. *PoS, ICHEP2016*:181, 2016.
- [94] Hannu Siikonen. Jet flavors: From the standard candles to the top quark mass. 2016. Diploma Thesis, Aalto University School of Science.
- [95] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.
- [96] D. N. Spergel et al. Wilkinson Microwave Anisotropy Probe (WMAP) three year results: implications for cosmology. *Astrophys. J. Suppl.*, 170:377, 2007.

- [97] The ATLAS and CMS Collaborations. Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV. *JHEP*, 08:045, 2016.
- [98] Wikipedia. Standard model — wikipedia, the free encyclopedia, 2016. https://en.wikipedia.org/w/index.php?title=Standard_Model&oldid=756155093 accessed 22-December-2016.
- [99] Kenneth G. Wilson. Confinement of quarks. *Phys. Rev. D*, 10:2445–2459, Oct 1974.
- [100] Dieter Zeppenfeld. Lecture notes from prospects in theoretical physics, 2005. <http://www.sns.ias.edu/pitp2/2005files/schedule2005.html> accessed 29-December-2016.

Jets are collimated sprays of hadronic particles and the experimental footprint of quarks and gluons, the constituents of all matter. They are present in virtually every proton-proton (pp) collision at the Large Hadron Collider (LHC), and thus a thorough understanding of jets is essential for experimental particle physics. We study jets at the particle level using the CMS detector and the sophisticated particle-flow event reconstruction algorithm. We call these pioneering studies *jet particology*.

We also use jets as tools for new discoveries in the *dijet resonance search*, where we search for new massive particles that decay to quarks or gluons producing two energetic jets back-to-back. This search has an exceptional discovery potential and an unprecedented energy reach, as we probe new physics phenomena up to 8 TeV.

The studies are performed using simulations and 8 and 13 TeV pp collisions from the CERN LHC. The results are published also in peer-reviewed particle physics journals.



ISBN 978-952-60-7663-8 (printed)
ISBN 978-952-60-7662-1 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Science
Applied physics
www.aalto.fi

**BUSINESS +
ECONOMY**

**ART +
DESIGN +
ARCHITECTURE**

**SCIENCE +
TECHNOLOGY**

CROSSOVER

**DOCTORAL
DISSERTATIONS**